Chromosomal rearrangements in evolution and in modern mammals: Different analyses and novel techniques for identifying potential fertility problems

A thesis submitted to the University of Kent for the degree of

DOCTOR OF PHILOSOPHY

In the Faculty of Sciences

March 2023

Lisa Marie Bosman

The School of Biosciences

Declaration

No part of this thesis has been submitted in support of an application for any degree or qualification of the University of Kent or any other University or institute of learning.

Lisa Marie Bosman

Acknowledgements

First, I would like to say a big thank you Darren and Peter for all your help, support, and guidance since I started my master's and followed on to my PhD. The lab and ERGO have been fantastic to work with, especially the incredible people from all over the world and research that is being done. It has been a very special couple of years. Thank you!

A huge thank you to Lucas, who mentored and guided me through my first year. You have been so helpful, humble, kind and an incredible friend over the years. Thank you to Izzy for the email jokes, fancy coffees, and fantastic friendship. Thank you to Carla for the best hugs, motivation, and care that someone could offer. Marie Claire, I can't thank you enough for your help with the thesis work, dinner parties, and just being the best "Italian" around.

Giuseppe and Claudia, and the rest of the Griffin and Ellis labs, THANK YOU! Thank you for always supporting me and being there for me when I needed you. You are all so special to me in your individual ways and are such kind people. It has been an honour working with you.

To Sheryl, a massive thank you for giving me the opportunities you have with Andrology Solutions and TDL Andrology. It was a great experience working with, and for you, and I very much appreciate your support and understanding.

Last, but not least, thank you to my parents, Mike, and Sabine; sister, Aimee; and close friends, for all your support over the last 4 years. I wouldn't have been able to do this without you and all your love. You are all such phenomenal people who have shown nothing but patience, care, compassion, enthusiasm, and stability when it sometimes felt like my world was collapsing in on itself. I wouldn't be where I am today without you. I love you more than you know!

Contents

List of Figures

517L16 (Texas Red) probes were used for image B. Normal chromosomes are circled in both (A) and (B). Scale bar is 10 µm (Image source: O'Connor *et al*., 2017).

Figure 1-11 The role of oxidative stress (OS) in DNA, mitochondria, and the sperm membrane (Image source: Wagner, Cheng & Ko, 2017).

Figure 2-1 A standard microscope slide layout, including two cover slips indicated by the two squares.

Figure 2-2 The phylogenetic tree of the 11 species tested (far right boxes) in specific aim 1 (section 3) with the 48 cattle BACs.

Figure 2-3 Octochrome device layout. Each box (top left of each square) indicates a different chromosome (middle of each square). Boxes 1-8 were used for equine chromosomes 25-X. Proximal (p) to the centromere probes are labelled with FITC and probes distal (d) to the centromere were labelled with Texas Red. The p-arm was labelled in FITC, and the q-arm was labelled in Texas Red for equine chromosome X.

Figure 2-4 Multiprobe device layout with each box (number in the top left of each square) indicating a different chromosome (middle of each square). Boxes 1-24 were used for equine chromosomes 1-24. Metacentric equine chromosome 1-13 p-arm probes were labelled with FITC, and q-arm probes were labelled with Texas Red. Acrocentric equine chromosome 14-24 were labelled as proximal (FITC, p) and distal (Texas Red, d) to the centromere for each chromosome.

Figure 2-5 Examples on the same chromosome of how different FLPter measurements were made. The top left picture indicates how a full chromosome measurement was completed and the other images are related to measurements from the centromere to the probe (yellow line).

Figure 2-6 Abcam lipid peroxidation (LPO) assay graph of red/green ratio (x axis) to counts (y axis). The positive control using 250 µM hydrogen peroxide (H_2O_2) is indicated on the left curve (blue), and sample control on the right curve (orange). (Image source: Abcam, 2019).

Figure 2-7 Graphs from Aitken *et al*. (2007) demonstrating the analysis of the impact of time and ferrous ion promotion on lipid peroxidation (LPO) in sperm with BIODIPYC_11, also a lipid peroxidation sensor (LPS). From left to right: Flow cytometry analysis of control cells (left); and flow cytometry analysis demonstrating the progressive increase in BIODIPYC_11 green fluorescence in human sperm samples treated with 80 µM Fe(II) for 15 minutes (middle) or 60

minutes (right). FL1-H (x axis) is fluorescence intensity and FL2-H (y axis) are events. (Image source: Aitken *et al*., 2007).

Figure 2-8 An example of the methods used to determine the sperm population and lipid peroxidation (LPO) ratio. (A) A forward scatter (FSC) vs side scatter (SSCa) plot as a quality control (QC) for sperm in the sample (Q1-UR); (B) FITC vs FSC graph used to indicate damaged cells in the sample whereby E2 shows the sperm population; and (C) PE vs FSC graph used to indicate damaged cells in the sample whereby E3 shows the sperm population.

Figure 2-9 Fragmented DNA or single stranded DNA (ssDNA), red fluorescence) (x axis) versus the native DNA stainability or double stranded DNA (dsDNA) (green fluorescence) (y axis) (18A). The change from green to red is used to calculated Sperm Chromatin Structure Assay® (SCSA®) DNA fragmentation index (DFI) (%). Gating of the SCSA® sperm data (18B) used from Evenson (2011). Norm: Normal sperm; HDS: High DNA Stainability; Mod DFI: Moderate DFI. (Image source: (A) Evenson, 2016; (B) Evenson, 2011).

Figure 2-10 The formula for DNA fragmentation index (DFI), with the main variables calculated from this formula, below it, including the mean of the DFI in the population, standard deviation of the DFI population, and the percentage DFI. (Image source: Evenson *et al*., 1985)

Figure 3-1 All the species tested (far right) in this specific aim, are shown in a phylogenetic tree indicating their various nodes of similarity and divergence from one another. Mammals in the same colour boxes, form part of the same order.

Figure 3-2 Cattle chromosome 3 probes CH240- 474H7 (FITC) and CH240-465O11 (Texas Red) hybridising to cattle chromosome 3 (left) and sheep chromosome 1 (right) (Magnification x1000) (Image source: Own images).

Figure 3-3 Scatter plots of the (A) Divergence time (millions of years apart, MYA) of each of the species from cattle (x axis) versus their respective hybridisation success rates (HSR) (%) (y axis); and (B) the species' diploid numbers (no.) (2n) (x axis) versus their respective hybridisation success rates (HSR) (%) (y axis). The line of best fit indicated the correlation between the two variables. Each species is represented by a point on the graph.

Figure 3-4 Cattle BACs CH240-380F23 (FITC) and CH240-344K23 (Texas Red) hybridising to human chromosomes; and cattle BAC clone CH240-226K16 (Texas Red) hybridising to possum chromosomes (Magnification x1000) (Image source: Own images).

Figure 3-5 Box plots of the different groups of species studied (x axis) versus their cattle BAC clone hybridisation success rates (y axis). Mean (+), minimum, maximum, median, interquartile ranges, and data points such as outliers, are all shown for each box plot. For specific values, refer to Table 3-7.

Figure 3-6 Ideogram of the most highly conserved cattle BAC, CH240-256G3, on all 11 different species with chromosomal location identified with FLPter values, and their respective National Center for Biotechnology Information (NCBI) BLASTn CH240-256G3 sequence identity (%) match, underneath each species acronym. The chromosomal sizes are not relative to one another, unless there are multiple such as in the case of the possum.

Figure 3-7 Cattle chromosomes 2, 11 and X are displayed with their relative probe FLPter values and locations. Chromosomes are in relative size to one another, and clones are shown by their different colours.

Figure 4-1 CH241-205O16 (equine chromosome 5) indicating a single pair of chromosomal subtelomeric q-arm signals (left); and CH241-470F5 (equine chromosome 9) demonstrating two different pairs of Texas Red subtelomeric signals (right) (Magnification x1000) (Image source: Own images).

Figure 4-2 The FITC (blue) and Texas Red (orange) probe multiprobe (equine chromosomes 1-24) and octochrome (equine chromosomes 25-X) hybridisation success rate based on the probes per chromosome in Table 4-5. Chromosomes are listed on the x axis and the average percent hybridisation success rate, is on the y axis. Chromosomes 14, 21, 24 and 29 did not have appropriate proximal probes (FITC) and thus were excluded, while the distal probes (Texas Red) for those chromosomes were included.

Figure 4-3 Horse A, with an X chromosome translocation (A); and a normal X chromosome arrangement (B). The red circles indicate normal X chromosomes, while the yellow circles show the translocations. Thus, this horse was characterised as having a mosaic X chromosome translocation (Magnification x1000) (Image source: Own images).

Figure 4-4 Images of horse H with the equine chromosome X FISH probes. A normal chromosome is indicated with white arrows pointing to the Xp (FITC) and Xq (Texas Red) probes (left). Based on the FISH results in the left image, chromosome Xp (green circle) has swapped with 10q, which has attached itself to the centromere of chromosome X. Part of Xq has attached itself to the end of 10q (red arrow), while the other part of Xq has rearranged to the opposite end of the centromere (red circle). The overall sex chromosomal translocation

originally identified with karyotyping prior to use of BAC clones, is demonstrated on the right. In this image, “X” indicates part of Xq, “Y” shows 10q and “Z” is the other part of Xq. (Magnification x1000) (Image source: Own images)

Figure 4-5 Two photos from the 2023 Stallion AI Ltd. *The Directory: A Breeders Guide*; including a written piece regarding karyotyping as an option for breeders. (Image source: Own images).

Figure 5-1 Increasing hydrogen peroxide (H_2O_2) concentration on human sperm. (A) No H_2O_2/Untreated sample; (B) 250 µM H_2O_2; (C) 500 µM H_2O_2; (D) 1 mM H_2O_2; (E) 10 mM H_2O_2; (F) 100 mM H_2O_2; (G) 1 M H_2O_2; and (H) 9.8 M H_2O_2. Settings for capturing images: TRITC exposure 2.00 seconds, Gain 6.612, no BGR auto or enhance; FITC exposure 2.00 seconds, Gain 3.718, no BGR or enhance. DAPI exposure 1.825 seconds, Gain 1.00, no BGR and enhance on. Auto was off for all images. (BGR – blue green, red) (Magnification x1000) (Image source: Own images).

Figure 5-2 Scatter plot of the untreated human sample lipid peroxidation (LPO) ratios (red/green) (x axis) versus the LPO ratio of the same samples treated with various concentrations of hydrogen peroxide (H_2O_2) (y axis). Trend lines for each of the concentrations are shown, relative to the colour of each H_2O_2 concentration. The black dotted line is a threshold, below which, treated sperm should have a reduced LPO ratio value than their respective untreated sperm, and thus, is doing what we expect, and damage is being induced. uM: µM.

Figure 5-3 The effect of the incubation time of the hydrogen peroxide (H_2O_2) (bars in chart) on patient sperm sample (x axis) lipid peroxidation (LPO) ratios (y axis).

Figure 5-4 The effect of time length between washing the lipid peroxidation sensor (LPS) (bars in chart) on patient sperm sample (x axis) lipid peroxidation (LPO) ratios (y axis).

Figure 5-5 Box and whisker plot demonstrating the distribution of the results when testing immediately after washing of the lipid peroxidation sensor (LPS) (blue) versus testing 1-hour post-washing (red). N = 2 outliers are not demonstrated on the plot for 1-hour post-washing testing. Lipid peroxidation ratio (red to green) values are stated on the x axis.

Figure 5-6 The distribution of the patient sperm concentration data (million sperm/mL) is demonstrated in the left box plot; and the progressive (%), and total motility (%), data is shown in the right box plots, respectively. Minimum, maximum, median, interquartile range, mean (+) and outlier data points are shown for each box plot.

Health Organization (WHO) guidelines for morphology assessment. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation.

List of Tables

Table 2-6 The standardised variables for each test described in Table 2-5. M/ml: Million sperm/mL. IT: Incubation time. Testing: Time to testing after LPS washing. No time: Immediate testing. 300 cells/sec: 300 cells/second flow rate on the flow cytometer.

Table 2-7 Different tests completed with the stallion semen and the variables in each test. V/O/E is dependent on the test. V: Variable; O: Outcome; or E: Effect.

Table 2-8 The survey responses which were automatically classified into initial categories by the Jisc Online Survey Tool, and then were manually set into new categories based on the specific answers the respondents gave. Response numbers (left), initial category (middle) and new category (right) are shown.

Table 3-1 Species are depicted with their common name, order they belong to, and the divergence time from Bos taurus in millions of years apart (millions of years apart, MYA). BTA: *Bos taurus*. BBI: *Bison bison*. OAR: *Ovis aries.* RDU: *Rucervus duvaucelii*. SSC: *Sus scrofa*. CFA: *Canis familiaris*. ECA: *Equus caballus*. HSA: *Homo sapiens*. MMU: *Mus musculus*. RRA: *Rattus rattus*. TVU: *Trichosurus vulpecula*. Divergence times were calculated using TimeTree 5: An expanded resource for species divergence times (Kumar et al., 2022).

Table 3-2 Species divergence (millions of years apart, MYA) from *Bos taurus*, the diploid chromosome number (2n) and hybridisation success rate (%) from the 48 bovine BACs, is stipulated along with the different within order (Artiodactyl) species tested.

Table 3-3 Successful hybridisations using highly conserved bovine BACs on Artiodactyl species. The table contains the bovine BAC ID's, assumed chromosomal location (pre-quality control), location of proximal (p) (FITC) or distal (d) (Texas Red) sequence to the centromere; and conservation score (CS). Species include: the quality control (QC) on Cattle; and normal within-order (Artiodactyl) FISH investigation on American bison; domestic sheep; barasingha; and domestic pig.

Table 3-4 Species divergence (millions of years apart, MYA) from *Bos taurus*, diploid chromosome number (2n) and their respective overall hybridisation success rates (%) from the 48 bovine BACs tested, along with different out-group order animals (non-Artiodactyl).

Table 3-5 Successful hybridisations using highly conserved bovine BACs on non-Artiodactyl order, Perissodactyla (horse), Carnivora (dog), Rodentia (mouse and rat), Primates (human) and Diprotodontia (possum) species. The table contains the bovine BAC ID's, assumed chromosomal location (pre-quality control), location of the proximal (p) (FITC) or distal (d) (Texas Red) BAC sequence; and conservation score (CS).

Table 3-6 The calculated Pearson correlation coefficients (R) for each of the variables (x and y) are indicated, as well as groups tested against one another. *Significant at a $p < 0.05$.

Table 3-7 Statistics relating to Figure 3-4 are indicated in this table, including the maximum (upper whisker), 3rd and 1st quartiles, median, minimum (lower whisker), number of data points and mean for Artiodactyl, Laurasiatheria, Non-Artiodactyl and overall species hybridisation success rates (%).

Table 3-8 FLPter values calculated for each of the probes on cattle, bison and barasingha. Blank cells are due to either there being poor quality images or no signals, and thus no values were calculated for those probe-chromosome combinations. Probes removed from the table were due to there being no FLPter values for those probes or on the bison/barasingha chromosomes. *Statistically significantly different to cattle FLPter mean value at $p < 0.05$. **Statistically significantly different to cattle FLPter mean value at $p < 0.01$

Table 3-9 Orthologous genes found in various species based on specific BAC clone sequence hybridisation. Chromosomal gene placement is based on the location given by the National Center for Biotechnology Information (NCBI) search results and not the BLASTn results. The first 6 genes were found in all species, and the last part of the table are genes which were significantly different from the cattle results based on FLPter values, or due to the BLAST sequence chromosomal output not matching the NCBI gene chromosomal result for a specific species.

Table 3-10 Orthologous genes found in the cattle BACs were investigated in various species. The number of provisional, model, inferred, reviewed, and validated orthologous genes per species are shown.

Table 4-1 The horses karyotyped are listed, along with their breed and karyotyping results. t: translocation; inv: inversion. Equine chromosome 20p is the chromosomal area proximal to the centromere as there is no p-arm in acrocentric chromosomes. TB: Thoroughbred. Suffolk: Suffolk Punch.

Table 4-2 The results of each “FITC Probe” and “Texas Red Probe”. Chr: The chromosome the probes were selected to be on, based on BAC choice on the National Center for Biotechnology Information (NCBI) prior to labelling and testing. Location: the probe chromosomal location based on imaging (Subtelomeric, Multiple signals, No signal, Universal subtelomeric/centromeric). Signal: the signal strength of the probe (out of 5, 5 being the highest and 0 being the lowest). Chr C/I: if the probe was found on the correct (C) or incorrect

(I) chromosome based on karyotyping. If multiple signals were seen, no signal strength value was given and “Multiple signals” was stated under location, as well as “Multiple” in the Chr C/I. If the probe was karyotyped and found to be on the incorrect chromosome, no signal strength value was given, even if the “Location” was subtelomeric.

Table 4-3 The first combination of p-arm/proximal and q-arm/distal probes tested in pairs. Chr: The chromosome they are assumed to be on based on National Center for Biotechnology Information (NCBI) selection. Correct chr: The outcome of the experiment. Signal: The quality of the signals seen. No: The probes did not work, even with metaphase spreads; TxR: Texas Red probe. Signals were classified as either “Good”, “Average” or “Poor”.

Table 4-4 The second combination of p-arm/proximal and q-arm/distal probes tested in pairs. Chr: The chromosome they are assumed to be on based on National Center for Biotechnology Information (NCBI) selection. Correct chr: The outcome of the experiment. Signal: The quality of the signals seen. No: The probes did not work, even with metaphase spreads; TxR: Texas Red probe. Signals were classified as either “Good”, “Average” or “Poor”.

Table 4-5 The final equine BAC combinations for each of the chromosomes which were used to test the 19 different horses. The table contains the p-arm/proximal and q-arm/distal probes and the chromosome they are assumed to be on, based on karyotyping. Chr: Chromosome.

Table 4-6 The results of the 8 horses tested with the X chromosome probes. The overall results based on karyotyping and FISH probe results available can be seen in the last column. Karyotypes confirmed with FISH are stated as “confirmed with FISH”. Karyotyping mistakes found with FISH are stated as “incorrect karyotype”.

Table 4-7 The results of the 11 horses tested with the multiprobe and octochrome devices. The hybridisation success rate (HSR) is based on the number of probes which hybridised successfully to the chromosomes available. HSR: Hybridisation success rate (%) is considered poor: 1-20; Average: 21-25; Good: 26-32 based on the number of probe or probe combinations which were successful. The overall results based on karyotyping and FISH probe results available can be seen in the last column.

Table 4-8 Overall karyotype and FISH results for the 19 horses. Horse F fell under both a novel translocation and unclear/need more probes categories, hence there being a sample number of 20.

Table 5-1 The untreated human optimisation samples and relative positive control statistics are shown in the table indicating the mean (μ), standard error (SE), standard deviation (s.d., σ)

and interquartile range (IQR). The correlation (R) between the untreated sample and the hydrogen peroxide treatment on those samples, is also shown. Where no value is indicated, there were not enough samples to form a reliable correlation. The number of samples tested for a specific positive control (or untreated) is found in the last column (N).

Table 5-2 The lipid peroxidation (LPO) sensor concentration test statistics are shown in the table indicating the mean, standard deviation (s.d.) and interquartile range (IQR). The correlation (R) between the 1x lipid peroxidation sensor (LPS) test and the other concentrations (0.5x, 0.2x and 0.1x), is also shown. Where no value is indicated, there were not enough samples to form a reliable correlation. The number of samples tested for a specific test is found in the last column (N).

Table 5-3 The mean (μ), standard deviation (σ), interquartile range (IQR) and number (N) of the human samples are shown for each of the parameters tested. IH: In-house; TDL: The Doctor's Laboratory; C: Concentration; M/mL: Million sperm/mL; PR: Progressive motility; TM: Total motility; NF: Normal forms; DF: Defects head; DM: Defects midpiece; DT: Defects tail; ERC: Excess residual cytoplasm**.** ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm. DFI (%): DNA fragmentation index as measured by the Sperm Chromatin Structure Assay® (SCSA®). HDS (%): High DNA stainability as measured by SCSA®. LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio.

Table 5-4 Correlations between the in-house human sperm testing results and TDL testing results. Values in brackets are the number of samples which were tested for both parameters. ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm. DFI (%): DNA fragmentation index as measured by the Sperm Chromatin Structure Assay® (SCSA®). HDS (%): High DNA stainability as measured by SCSA®. LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Significant correlation at $p < 0.05$.

Table 5-5 Correlations between standard semen parameters and the in-house testing results for the human patient samples for the human samples. Values in brackets are the number of samples which were tested for both parameters to create the respective correlations. ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm. DFI (%): DNA fragmentation index as measured by the Sperm Chromatin Structure Assay® (SCSA®). HDS (%): High DNA stainability as measured by SCSA®. LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Significant correlation at $p < 0.05$.

Table 5-6 Correlations between the in-house testing parameter results. Values in brackets are the number of samples which were tested for both parameters. ORP (mV/mil): oxidation

reduction potential (mV) as measured by MiOXSYS per million sperm. DFI (%): DNA fragmentation index as measured by the Sperm Chromatin Structure Assay® (SCSA®). HDS (%): High DNA stainability as measured by SCSA®. LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Significant correlation at $p < 0.05$.

Table 5-7 Correlations between the stallion sperm testing results. Values in brackets are the number of samples which were tested for both parameters to create the respective correlations. Conc: Concentration; PR: Progressive motility; TM: Total motility; NP: Non-progressive motility; ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm; DFI (%): DNA fragmentation index; HDS (%): High DNA stainability as measured by Sperm Chromatin Structure Assay® (SCSA®); LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Significant correlation at $p < 0.05$.

Table 5-8 Correlations between the stallion sperm testing results. ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm; DFI (%): DNA fragmentation index; HDS (%): High DNA stainability as measured by the Sperm Chromatin Structure Assay® (SCSA®); LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Significant correlation at $p < 0.05$.

Table 5-9 Stallion pregnancy results and their respective mean semen parameters are indicated above. Conc: Concentration; PR: Progressive motility; TM: Total motility; ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm; DFI (%): DNA fragmentation index; HDS (%): High DNA stainability as measured by the Sperm Chromatin Structure Assay® (SCSA®)

Table 5-10 Lipid peroxidation (LPO) values for the frozen and chilled samples. Both groups have means and standard deviations (s.d.) for both all the samples in the group, and for stallions A, O and P only. No significant differences were found.

Table 5-11 Sperm Chromatin Structure Assay® (SCSA®) and Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) DNA fragmentation index (DFI) values for the frozen and chilled samples. Both groups have means and standard deviations (s.d.) for both all the samples in the group, and for stallions A, O and P only.

Table 5-12 Egg- and milk-based cryoprotectant results. Conc: Concentration in million sperm/mL; PR: Progressive motility (%); TM: Total motility (%); ORP: Oxidation reduction potential (mV); DFI (%): DNA fragmentation index; HDS (%): High DNA stainability as measured by Sperm Chromatin Structure Assay® (SCSA®).

Table 5-13 Breed comparisons for various semen parameters and tests. N = number of stallions in that category. Conc: Concentration; PR: Progressive motility; TM: Total motility; ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm; DFI (%): DNA fragmentation index; HDS (%): High DNA stainability as measured by Sperm Chromatin Structure Assay® (SCSA®); LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Indicates mean is statistically different from Warmblood horses using an unpaired t-test at a p-value < 0.05. **Indicates mean is statistically different from Racing/Endurance horses using an unpaired t-test at a p-values < 0.05.

Table 6-1 Distribution of laboratories performing semen analysis in the UK. Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents are shown in each cell, with the number of respondents in brackets. The different answers the respondents could select from is shown in the left-hand column. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation.

Table 6-2 The purpose of laboratory semen analysis. Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents are shown in each cell, with the number of respondents in brackets. The different answers the respondents could select from is shown in the left-hand column. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation. *Significantly different from Group 1.

Table 6-3 Laboratory compliance questions - 5, 6 and 15. Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents answering “Yes” are shown in each cell, with the number of respondents in brackets. The different questions are shown in the left-hand column. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation. *Significantly different to Group 1.

Table 6-4 The proportion of laboratories reporting additional seminal fluid parameters. Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents answering “Yes” are shown in each cell, with the number of respondents in brackets. The different parameters which could be selected are shown in the left-hand column. Dark green (100-80%), light green (79-60%), yellow (59-40%), orange (39-20%), red (19-0%). Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation. *Significant when compared to Group 1. **Significant when compared to Group 3.

Table 6-5 Quality control in the laboratory. Questions 16, 16a and 17, are found in the left-hand column. Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents answering “Yes” to each question are shown in the cells, with the

number of respondents for each in brackets. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation. *Significantly different from Group 1. **Significantly different from Group 2. ***Significantly different from Group 3.

Table 6-6 Patient report comments included, if a value is found outside of the normal range. Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents answering “Yes” are shown in each cell, with the number of respondents in brackets. The different answers respondents could select from are shown in the left-hand column. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation. *Significant when compared to Group 1. **Significant when compared to Group 2.

Abbreviations

σ	Standard deviation
μ	Mean
2n	Diploid chromosome number
4HNE	4-hydroxynonenal
AI	Artificial insemination
ANOVA	Analysis of variance
AO	Acridine orange
ARCS	Association of Reproductive and Clinical Scientists
ART	Assisted reproduction technologies
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BBI	*Bison bison,* American bison
BLAST	Basic Local Alignment Search Tool
BMP	Bone morphogenetic protein
BNC2	Zinc finger protein basonuclin-2
bp	Base pairs
BrdUTP	5'-bromo-2'-deoxyuridine 5'-triphosphate nucleotide
BTA	*Bos taurus*, Cattle
C-banding	Constitutive heterochromatin banding
CAMK2G	Calcium/calmodulin-dependent protein kinase type II gamma chain
CAMKMT	Calmodulin-lysine N-methyltransferase
CASA	Computer Assisted Semen Analysis
CCD	Charge-coupled device
cDNA	Complementary DNA
CFA	*Canis lupus familiaris*, Domestic dog
CH240	CHORI-240 BAC library
CH241	CHORI-241 BAC library
CHCHD1	Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 1
CNV	Copy number variation
CO_2	Carbon dioxide
CoGe	Comparative Genomics platform
COP9	Constitutive photomorphogenesis 9

COPS7A	Constitutive photomorphogenesis 9 Signalosome Subunit 7A
CS	Conservation score
CV	Coefficient of variation
DAPI	4,6-Diamino-2-Phenol-Indole
ddH_2O	Deionized water
ddNTP	Dideoxynucleotide triphosphate
DFI	DNA fragmentation index (%)
DHA	Decohexaenoic acid
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DPA	Docosapentaenoic acid
dsDNA	Double-stranded DNA
DSB	Double stranded breaks
DTT	Dithiothreitol
dUTP	Deoxyuridine triphosphate
E. coli	Escherichia coli bacteria
EBR	Evolutionary breakpoint regions
ECA	*Equus caballus*, Domestic horse
EDTA	Ethylenediaminetetraacetic acid
EHBP1 protein 1	Epidermal growth factor receptor substrate 15 homology domain binding
EQA	External quality assurance
EQC	External quality control
ET	Embryo transfer
ETC	Electron transport chain
F-plasmid	Fertility plasmid
FA	Fatty acid
FAO	Food and Agriculture Organization
FC	Flow cytometer/cytometry
FSC	Forward scatter
FISH	Fluorescence *in situ* hybridisation
FITC	Fluorescein isothiocyanate
FLPter	Fraction length relative to the p-terminus

FUT11	Fucosyltransferase 11
G-banding	Giemsa banding
Gb	Gigabases
GDF	Growth differentiation factor
GP	General practitioner
GPX	Glutathione peroxidase
H_2O_2	Hydrogen peroxide
HCl	Hydrochloric acid
HDS	High DNA stainability (%)
HFEA	Human Fertilisation and Embryology Authority
HGP	Horse Genome Project
HHBS	Hank's balanced salt solution with Hepes
HSA	*Homo sapiens*, Human
HSB	Homologous synteny blocks
HSP	High-scoring segment pair
HT-NGS	High-throughput next generation sequencing
HYPP	Hyperkalemic periodic paralysis
ICSI	Intracytoplasmic sperm injection
IQC	Internal quality control
IQR	Interquartile range
ISH	*In situ* hybridisation
ISO	International Organization for Standardisation
IVF	*In vitro* fertilisation
Kb	Kilobase
KCL	Potassium chloride
LB	Luria-Bertani agar
LPO	Lipid peroxidation
LPS	Lipid peroxidation sensor from Abcam
LS	Loin strength
Mb	Mega base
MDA	Malondialdehyde
MiOXSYS	Male Infertility Oxidative System
MMU	*Mus musculus*, House mouse
MYA	Million years apart

MYOZ1	Myozenin-1
NaCl	Sodium chloride
NADPH	Nicotinamide adenine dinucleotide phosphate
NaOH	Sodium hydroxide
NCBI	National Centre for Biotechnology Information
NDST2	Bifunctional heparan sulphate N-deacetylase/N-sulfotransferase 2
NEQAS	National External Quality Assessment Service
NGS	Next generation sequencing
NHS	National Health Service
NICE	National Institute of Clinical Excellence
NOX	Nicotinamide adenine dinucleotide phosphate oxidase
NPAS3	Neuronal Per-Arnt-Sim domain protein 3
NR2F2	Nuclear Receptor Subfamily 2 Group F Member 2
O_2^-	Superoxide anion
OAR	*Ovis aries*, Domestic sheep
OH^-	Hydroxyl radical
OLD	Oldenburger Warmblood (Horse breed)
OLWS	Overo lethal white foal syndrome
ORP	Oxidation-reduction potential
OS	Oxidative stress
OXPHOS	Oxidative phosphorylation
PAX2	Paired box 2
PBS	Phosphate buffered saline
PCD	Primary ciliary dyskinesia
PCDH9	Protocadherin 9
PCF	Predicted chromosome fragments
PCR	Polymerase chain reaction
PE	Phycoerythrin
PSMD14	26S proteasome non-ATPase regulatory subunit 14
PUFA	Polyunsaturated fatty acid
Q-banding	Quinacrine banding
QC	Quality control
QMC	Quality management system
QTL	Quantitative trait loci

R	Correlation coefficient
R-banding	Reverse banding
RACA	Reference assisted chromosome assembly
RDU	*Rucervus duvaucelii*, Barasingha
ROS	Reactive oxygen species
RPM	Revolutions per minute
RRA	*Rattus rattus*, Black Rat
SALL1	Spalt like transcription factor 1
SCD	Sperm Chromatin Dispersion test (Also known as the HALO test)
SCSA®	Sperm Chromatin Structure Assay
s.d.	Standard deviation (σ)
SE	Standard error
SEC24C	Secretory protein 24 Homolog C
SH2D1A	Non-receptor tyrosine kinase homology 2 domain containing 1A
SOD	Superoxide dismutase
SSB	Single-stranded breaks
SSCa	Side scatter
SSC	Saline-sodium citrate
SSC	*Sus scrofa*, Domestic pig
ssDNA	Single-stranded DNA
TB	Thoroughbred (Horse breed)
TDL	The Doctors Laboratory
TdT	Terminal deoxynucleotidyl transferase
TENM1	Teneurin transmembrane protein 1
TNE	Tris-HCl, NaCl and EDTA buffer
TUNEL labelling	Terminal deoxynucleotidyl transferase-mediated fluorescein-dUTP nick end
TVU	*Trichosurus vulpecula*, Common brush-tail possum
UGS	Unplaced genomic scaffold
UK	United Kingdom
UKAS	United Kingdom Accreditation Service
UL	Upper left
UR	Upper right
USA	United States of America

USP54	Ubiquitin specific peptidase 54
WHO	World Health Organization
YAC	Yeast artificial chromosome
ZEB2	Zinc finger E-box-binding homeobox 2
ZNF521	Zinc finger protein 521

Abstract

Reproductive isolation between species and reproductive problems in individual animals have certain parallels. When a chromosomal rearrangement, (e.g., translocation) occurs, it can cause sub-fertility by impeding meiosis, recombination reduction and chromosomally unbalanced gamete production. In extremely rare occasions, during evolution, such an occurrence can begin the process of reproductive isolation (speciation) if the translocation becomes fixed in the population in homozygous form. Analogies include the common (1:1000) 13:14 Robertsonian chromosome fusion (a cause of infertility) and human chromosome 2, a fusion of two ancestral great ape chromosomes. Chromosomal rearrangements occur more in sperm than eggs and are mediated through Deoxyribonucleic acid (DNA) breakage, another correlate of infertility. There is sufficient concern about sperm DNA damage due to reactive oxygen species (ROS) and lipid peroxidation (LPO) (and their connections to male fertility), that some laboratories are establishing routine DNA damage screening services alongside translocation screening. For any mammal, if fertility is indicated, a routine chromosome analysis and a screen for sperm DNA damage or ROS/LPO would be prudent. Ensuring these services follow accreditation body and best practice guidelines is imperative. Animal breeders and physicians make decisions based on these results and are aware that they could have major impacts on the andrology sector and/or the genetic improvement of a species. To this end, this thesis considers aspects of chromosome rearrangement during evolution and as a cause of infertility, causes of sperm DNA damage and the monitoring and regulation testing in the andrology sector. Specifically, this thesis:

- Assessed a panel of 48 sequence- and conservation score-based homologous cattle fluorescence *in-situ* hybridisation (FISH) probes in both phylogenetically similar and more distantly related mammals for cross-species hybridization patterns. This was achieved successfully, with a general pattern that the greater the evolutionary distance from cattle, the less successful the probes. For instance, μ = 76.3% of the probes successfully hybridised to the closely related Artiodactyls such as bison, barasingha and sheep, but only μ = 49.3% for the more distantly related non-Artiodactyl species. The potential for these cattle probes to be used to assess chromosome abnormalities in individual species is discussed.
- Assessed equine fertility using cytogenetic analysis and a novel technique for accurate detection of chromosome rearrangements. This was partly achieved based on previous strategies developed in pigs and cattle using sub-telomeric FISH probes. A total of 64.1% Texas Red and 56.3% Fluorescein isothiocyanate (FITC) probes (60.2% total) successfully hybridised to the required locus, leading to the development of a device that was capable of confirming 10 of 19 horse karyotypes, while identifying three novel rearrangements that were previously untraceable by karyotyping alone. The strategy has been employed in the horse breeding community, as has an adapted approach for assessing DNA damage in equine sperm.
- Developed a flow cytometric assay for membrane lipid oxidation in human and equine sperm thereby testing the hypothesis that there are fundamental differences between the species and DNA damage as well as oxidation reduction potential (ORP) and LPO screening could be valid tools for identifying fertility potential of a male. This was successfully achieved, with differences including horses (μ = 11.7%) having a much lower Sperm Chromatin Structure Assay (SCSA®) DNA fragmentation index (DFI) than humans (μ = 21.6%), while the correlation between LPO and DFI for the frozen-thawed human samples (R = -0.380) was much lower than for the frozen-thawed stallion sperm (R = -0.970). A regulatory framework suitable for introduction in the clinic (humans) or specialist testing laboratories (horse) was suggested for DNA damage and ROS screening, yet further work is needed on the LPO test to provide accurate and repeatable results.
- Tested the hypothesis that there are differences in semen analysis protocols between laboratories and clinics based on the regulatory body with which they are accredited and the type of entity they claim to be. Here it was established that United Kingdom Accreditation Service (UKAS) accredited laboratories are more likely to adhere to International Organization for Standardization (ISO) 15189 standards and World Health Organization (WHO) guidelines, than Human Fertilisation and Embryology Authority (HFEA) licensed clinical laboratories. Many laboratories are not demonstrating any internal quality controls or adhering to external quality assurance programs, and there is a clear lack of standardisation within the United Kingdom (UK) andrology sector.

Overall, this thesis was successful in the achievement of its aims with several parallels across the superficially quite diverse, individual chapters drawn. In particular, results appear to substantiate the correlation between stallion sperm's "live fast, die young" analogy of energy production which has recently been associated with their fertility in the literature. This work has made significant contributions towards understanding gross genomic rearrangement (chromosomal and DNA damage) in both reproduction and reproductive isolation. It has developed novel strategies for fertility diagnosis and provided a framework and insight into how this might be regulated.

1. General Introduction

Mammalian cells undergo multiple processes to develop into their various specialised components. To produce sex cells such as the egg and sperm, which are the two distinctive haploid gametes from both female and male mammals, respectively, the process of cellular division, also known as meiosis, is completed. Within the first stage of this cycle, homologous chromosomes pair, and crossing over occurs, thereby exchanging genetic material between one another. It is during this first meiotic phase where translocations, or unusual breakages and rearrangements of chromosomes segments, can potentially occur, causing either balanced or unbalanced gametes (section 1.2). The prophase I step is followed by a series of events including metaphase I, anaphase I and telophase I, which lead to the complete segregation and production of two haploid (n = 2) cells. This process repeats itself in meiosis II, except this time, chromosomes condense, and sister chromatids separate, resulting 2n = 4 haploid gametes are formed, whereby each chromosome has only one chromatid.

These individual chromatids, one from a male and one from a female, will combine with one another during fertilisation, producing a zygote in diploid state (2n). If there is a translocation which has occurred during either the crossing over in meiosis I of haploid gametes or in the formation of a diploid zygote, it can be due to two chromosomes swapping their genetic material (Reciprocal translocation) or the fusion of acrocentric chromosomes together, thereby forming a metacentric chromosome (Robertsonian translocation). Other forms of chromosomal abnormalities and changes are further discussed in section 1.2.

When a chromosomal rearrangement occurs, there are two possible consequences. Firstly, it could cause the individual to be sub-fertile, provided there is no net gain nor loss of genomic material. For example, a chromosome translocation in heterozygous form (e.g., in a human, boar or bull) results in impediment of meiosis through the formation of a pairing cross, reduction in recombination in the pairing regions and the production of chromosomally unbalanced gametes (Figure 1-1). Unbalanced gametes can often cause mammals to abort or

miscarry, while balanced gametes can compromise the individual animal's fertility potential. In extremely rare occasions, such an occurrence could have a small chance of being the start of speciation. In evolutionary time however, vanishingly small chances become inevitable eventually, and part of the evolution of most eukaryotic organisms is reproductive isolation through chromosome rearrangement.

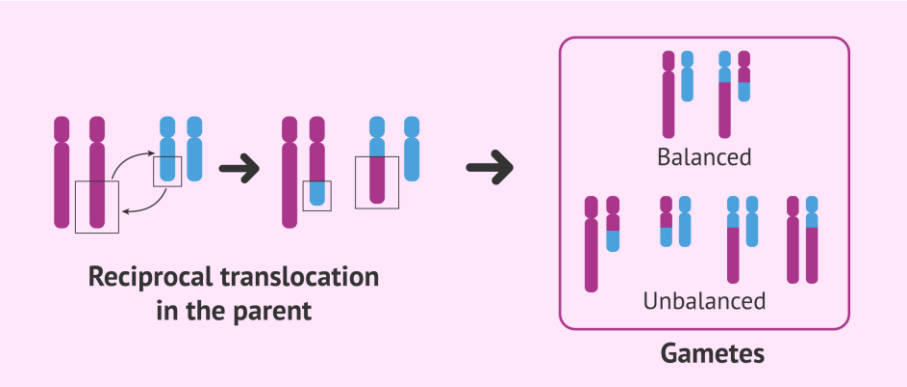


Figure 1-1 Reciprocal translocations and the potential balanced or unbalanced gametes which could be produced (Image source: Garcia & Gutiérrez, 2021).

Speciation or the reproductive isolation of a group of mammals which cannot crossbreed successfully to produce fertile offspring, is mediated by chromosome rearrangement and can only occur if the newly formed rearrangement becomes fixed in homozygous form. To represent how Robertsonian chromosome fusion can lead to both infertility in an individual and throughout evolution, consider a human example (Figure 1-2) (section 1.2). The human karyotype (Figure 1-2A) is thought to have derived from its great ape forebears with a similar karyotype evident in chimpanzees (Figure 1-2C), gorillas and orangutans. The principal change is the human chromosome 2 is thought to have arisen as a result of Robertsonian translocation (fusion) of two acrocentric great ape chromosomes (Figure 1-2D) (Fan *et al.*, 2002). Robertsonian translocations are normal in humans, with a fusion of chromosomes 13 and 14 (Figure 1-2B) the most common, present in around 1 in 1000 individuals (Scriven 2001).

Given the prevalence of the 13:14 translocation (around 1 in 500 of the human population) (Figure 1-2B), it is perhaps not unreasonable to speculate that the evolution of the next human species may have 22 pairs of chromosomes instead of 23, with e.g. the current 13 and 14 (or

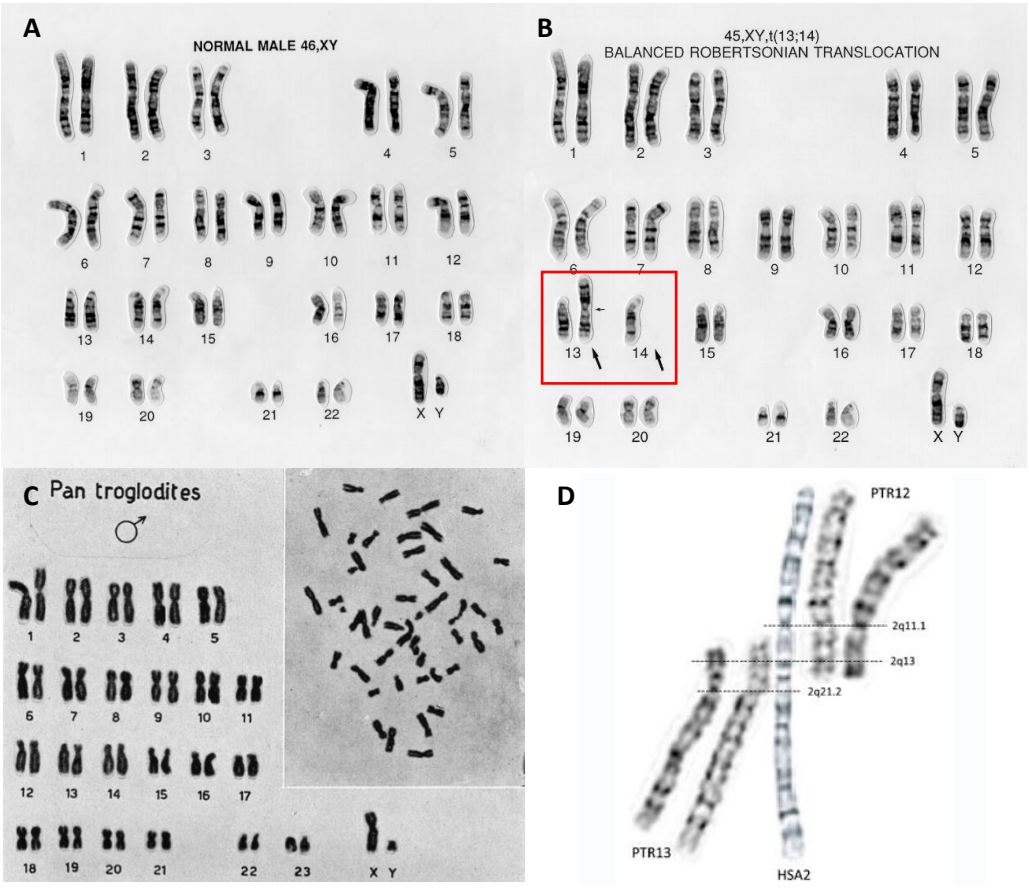


Figure 1-2 The similarity between chromosome rearrangements in evolution and this causing infertility in individuals. (A) represents the normal male karyotype (46,XY). (B) demonstrates the results of a chromosome fusion of human chromosomes 13 and 14 (45,XY,t(13:14)). A chimpanzee karyotype (*Pan troglodytes*, PTR) (C), in which the major difference is a fusion of two human chromosomes, represented as chromosome 2 in humans (HSA2) (D). (Image sources: (A and B) Wessex Reg. Genetics Centre, no date; (C) modified from Chiarelli, 1962; (D) Rejón, 2019).

other acrocentric chromosomes) fused to form a single chromosome about the size of chromosome 3 (Ayala & Coluzzi, 2005). As large-scale karyotype analysis hasn't been investigated across every conceivable human population, this may already exist in some, un-karyotyped, remote pocket of civilization. While some may dismiss this as unreasonable speculation, there is precedent for this speculative hypothesis in the most common translocation in cattle (1:29 Robertsonian) which has been identified in a homozygous form in certain individuals (Jennings, Griffin & O'Connor, 2020). If these individuals were inter-bred to a reasonable population size, they would theoretically represent a new species, and eventually become reproductively isolated from the founder cattle species (section 1.3).

Chromosome rearrangement occurs more in spermatozoa than the egg and is mediated through deoxyribonucleic acid (DNA) breakage. Indeed, DNA breakage in the sperm is another common factor leading to infertility and is one of the most studied areas of reproductive medicine currently (section 1.4). There is sufficient concern about sperm DNA damage due to reactive oxygen species (ROS) and lipid peroxidation (LPO) impairment and their connections with male fertility, that some laboratories (including the one in which this PhD was performed) are now proposing setting up a routine DNA damage screening service alongside its routine screening service for chromosome translocations in pigs and cattle. For both humans and other animals, if fertility is a concern, then a routine karyotype and a screen for sperm DNA damage or ROS would be wise measures. In humans, this would be to assist a couple in having a family; and in (e.g., domestic) animals, this may mean that certain males are not included in a breeding programme.

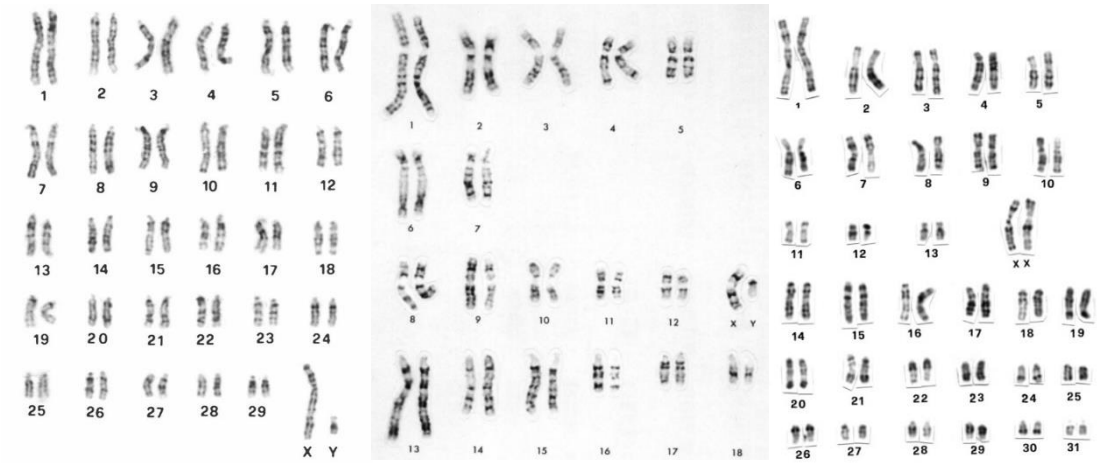
Additionally, ensuring these types of practices are following accreditation body guidelines, as well as best practice guidelines, is imperative. Regulation of a standard semen analysis as well as any testing in a laboratory or clinical setting should be highly monitored, especially when results are being relayed to clinicians, breeders and physicians who are making decisions which could have major impacts on the andrology sector and/or the genetic improvement of a species (section 1.5).

This thesis considers aspects of evolutionary chromosome rearrangement, chromosome rearrangement as a cause of infertility, causes of sperm DNA damage in humans and other animals, as well as the monitoring and regulation of this type of testing in the andrology sector in the United Kingdom (UK).

1.1 The karyotypes of different animals and how they define the genome

1.1.1 Examples of different mammalian karyotypes

Classical cytogenetics is the study of chromosomes and their abnormalities using conventional staining techniques (Bugno *et al.*, 2009; Yahaya *et al.,* 2019). After staining, chromosomes are identified and placed into their particular pairs using banding methods, and hence numerical and structural abnormalities can be visualised (Bugno *et al.*, 2009). Karyotype analysis, a conventional technique, identifies banding patterns on the chromosome and major structural changes including numerical and structural rearrangements (Pauciullo *et al.,* 2014; Hu, Maurais & Ly*,* 2020). Karyotyping refers to the structured grouping of chromosomes into their respective pairs based on distinctive banding patterns, sizes and centromeric position; and is unique to each species based on number and description (Gersen & Keagle, 2013) (examples in Figure 1-3). As an example, human karyotypes are described as 2n = 46 or for normal males, 46, XY, (Figure 1-2A) and females 46, XX (Gersen & Keagle, 2013).



A **B** **C**

Figure 1-3 G banding karyotypes of cattle (A); pig (B); and horse (C) (Image sources: (A) Iannuzzi, 1996; (B) Gustavsson, 1988; (C) Iannuzzi *et al.,* 2014).

To identify the chromosomal regions or whole chromosomes, a specific nomenclature is used depending on the species. In general, chromosomes are numbered from largest to smallest in size, starting with metacentric chromosomes, and then acrocentric, followed by the short arm

which is designated as p, and the long arm, q (Gersen & Keagle, 2013). The p-arm is always situated at the top and the q arm at the bottom in orientation (Gersen & Keagle, 2013). The region is then numbered and lastly, the band is given a location on a chromosome (Gersen & Keagle, 2013), allowing researchers to individualise chromosomes and identify specific regions within their studies. In situations where there is no p-arm, such as with acrocentric chromosomes, the designation of proximal (p) and distal (d) of the centromere is often used to identify a chromosomal location.

The process of identifying chromosomes and their specific regions, including the use of fluorescent markers, has given rise to the era of molecular cytogenetics, which involves the methods of fluorescence *in situ* hybridisation (FISH) (Gersen & Keagle, 2013; Yahaya *et al*., 2019;). This scientific discipline allowed the Human Genome Project to take shape (Gersen & Keagle, 2013) while coinciding with genome sequencing technologies as well as comparative genomic research.

1.1.2 Methods of study

Staining, also known as banding, techniques, founded in the 1970s, have reduced the number of livestock production reproductive losses, comparative species/breed homology studies (Yahaya *et al.,* 2019) and improved medical research. Staining the chromosomes allows segments to be divided and identified due to the duller and/or brighter areas on the chromosome (which look like bands), based on the abundance of GC- and AT-rich regions (Gersen & Keagle, 2013) (Figure 1-4). The staining method and resolution is important to note when karyotyping an image as different staining techniques fluoresce various regions, and they also have different resolution visibility (Gersen & Keagle, 2013). Most chromosome “bands” have a size of approximately 5-10 Megabases (Mb) DNA (Gersen & Keagle, 2013), yet visualisation of these regions is completely dependent on the staining technique used as some

only stain certain areas of chromosomes, while others stain entire chromosomes (Gersen & Keagle, 2013). Due to this, many different methods of staining have been developed.

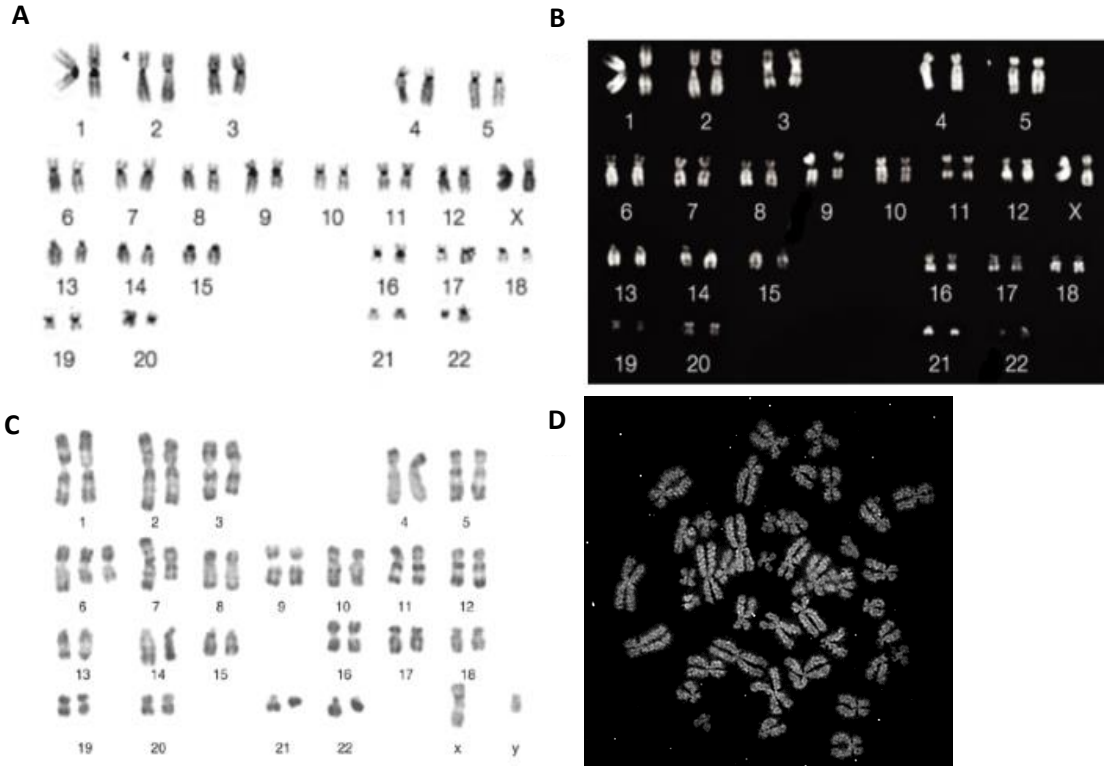


Figure 1-4 Different banding techniques with a human karyotype. G banding (A); Q-banding (B); C-banding (C); and DAPI banding (D). (Image sources: (A), (B) and (C) modified from O'Connor, 2008; (D) Own image).

The most common and widely used technique originated in 1971 with Giemsa staining (G-banding) (Seabright, 1971; Gersen & Keagle, 2013; Hu, Maurais & Ly, 2020) (Figure 1-4A). GTG banded karyotypes are treated with trypsin followed by Giemsa staining, thereby producing a dull (AT-rich heterochromatic) and bright (GC-rich euchromatic) pattern of banding (Gersen & Keagle, 2013). The GC-rich bright bands contain biologically active gene rich, highly evolutionary conserved regions (Gersen & Keagle, 2013). These banding pattern distributions allow scientists to identify unique chromosomes within a karyotype, while simultaneously identifying potential conserved gene blocks (Hu, Maurais & Ly., 2020). Additionally, G-banding identifies chromosome segments of 5-10Mb allowing for easy identification of large chromosomal segment rearrangements (Hu, Maurais & Ly., 2020).

Comparable to G-banding, Quinacrine banding (Q-banding), which was the first banding technique to be invented and still in use today and was used to develop the first human karyotype (Caspersson, Zech & Johansson, 1970), has a similar staining method whereby AT-rich regions fluoresce in duller bands and GC-rich areas fluoresce brightly (Hu, Maurais & Ly, 2020) (Figure 1-4B). Conversely, Reverse banding (R-banding) stains the GC-rich regions, showing up in images as dull bands and conversely, AT-rich regions fluoresce brightly (Gersen & Keagle, 2013). This method has both non-fluorescent and fluorescent (use of acridine orange) techniques and is useful with studies where telomeres are the subject of interest as they often cannot be seen with G- or Q-banding (Gersen & Keagle, 2013). When using acridine orange in fluorescent R-banding, the acridine orange intercalates with the DNA, thereby causing double stranded DNA (dsDNA) to fluoresce green, and single stranded DNA (ssDNA) to fluoresce red. The use of acridine orange intercalation and therefore staining of the DNA, is also used for the Sperm Chromatin Structure Assay®, further described in section 1.4.4. Another method, constitutive heterochromatin banding (C-banding), as the name suggests, only stains the chromatin tightly packed areas around the centromeres; and even though is also uses Giemsa as a staining solution similar to G-banding, a different methodology is utilized to visualize this component of the chromosome (Gersen & Keagle, 2013) (Figure 1-4C). Using this methodology, translocations or chromosomal rearrangements involving or surrounding the centromere, can be studied efficiently (Gersen & Keagle, 2013; Hu, Maurais & Ly., 2020). One of the more recent staining techniques, and one widely used in our laboratory, is the 4,6-Diamino-2-Phenol-Indole (DAPI) fluorescent marker (Figure 1-4D). This technique forms bonds with AT-rich, double stranded DNA (dsDNA) (Gersen & Keagle, 2013), yet can also stain the entirety of the chromosome, allowing overall visualisation of the chromosome when pictured under a fluorescent microscope.

Due to the cheap cost of conducting a banding experiment, it is widely used, but analysis is labour intensive and needs specialised training when compared to new and more recent

molecular techniques (Bugno *et al.,* 2009; Hu, Maurais & Ly, 2020). Banding practices additionally have the limitation of only detecting larger structural changes, thus other methods are required for identifying smaller more cryptic abnormalities (Bishop, 2010).

More refined techniques were needed to not only visualise cryptic chromosomal changes unidentifiable to the naked eye, but also for more specific gene position or gene order changes, which karyotyping may not, or cannot, pick up on (Yahaya *et al.,* 2019). FISH and other *in situ* hybridisation (ISH) techniques developed in the 1980s are known to be more specific, precise, and fast (Bugno *et al.,* 2009; Wolff, 2013) with their uses of labelled probes allowing for improved resolution (Yahaya *et al.,* 2019). FISH methods are used to study a variety of different DNA changes, gene localisations and DNA sequences on same and divergent species allowing for comparative and evolutionary cytogenetic studies (Yahaya *et al*., 2019). There are multiple distinct types of FISH methodologies ranging from individual locus FISH to whole chromosome painting (Figure 1-5).

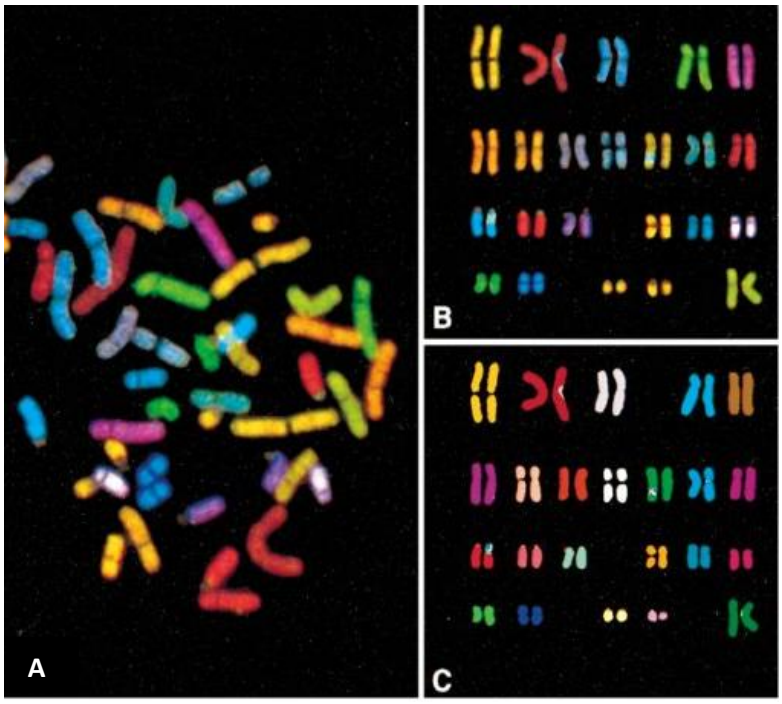


Figure 1-5 Human metaphases labelled with multicolour chromosome painting probes which are only available in humans and mice (A); and the karyotypic arrangement of the labelled chromosomes (B and C) (Image source: Ried *et al.,* 1997).

One of the first FISH techniques which is still used in modern studies, is the use of chromosome painting developed in 1988 (Graphodatsky, Trifonov & Stanyon*.,* 2011; Hu, Maurais & Ly*,* 2020). They are species and chromosome specific probes made from microdissection and polymerase chain reaction (PCR) amplification or flow-sorting chromosomes which are either fluorescently labelled as regions or entire chromosomes, respectively (Hu, Maurais & Ly*,* 2020). Chromosome paints are significant tools as they hybridise to the whole chromosomal set of the metaphase/interphase chromosome preparations of a specific species, thus numerical/structural chromosomal rearrangements can be easily identified (Ried *et al.,* 1998) (Figure 1-5). This technique is useful in disease studies and animal models due to the high specificity and sensitivity of the probes (Ried *et al.,* 1998), consequently allowing researchers to identify previously undetectable abnormalities. Due to this, they can also be used to study homology between different species of mammals, thereby visualizing evolutionary conserved synteny segments between species (Graphodatsky, Trifonov & Stanyon*,* 2011).

In contrast to whole chromosome painting probes, individual locus FISH is a method whereby locus specific probes hybridise to targeted regions or locations on the chromosome (Bishop, 2010; Hu, Maurais & Ly*,* 2020). The primary clones used for investigative studies and testing, as well as for individual locus FISH, such as specific genes or sub-telomeric research, are bacterial artificial chromosomes (BACs) (Wolff, 2013). These fertility plasmids (F-plasmid) of *Escherichia coli* (E. coli) occur naturally and can reach DNA fragment lengths over 300 Kilobases (Kb) as vectors (Dixit *et al.,* 2014), yet on average they are around 100-230 Kb (Osoegawa *et al.*, 1998; Hu, Maurais & Ly*,* 2020). BACs have the added benefit of being stable and the DNA contained within a BAC is easily isolated through plasmid extraction, even if they are smaller than their counterparts, Yeast artificial chromosomes (YACs), another form of clone used for FISH (Osoegawa *et al.,* 1998; Wolff, 2013). Once BACs are selected and isolated for the purpose of annealing to a particular target DNA sequence, they are attached to fluorescent molecules such as Texas Red or Fluorescein isothiocyanate (FITC) so the sequences in question can be

visualised and imaged under a fluorescent microscope (Bishop, 2010) (Figure 1-6). As BACs are so sequence specific, minor genetic abnormalities or chromosomal rearrangements can easily be identified for a wide variety of different research, therapeutic and commercial uses.

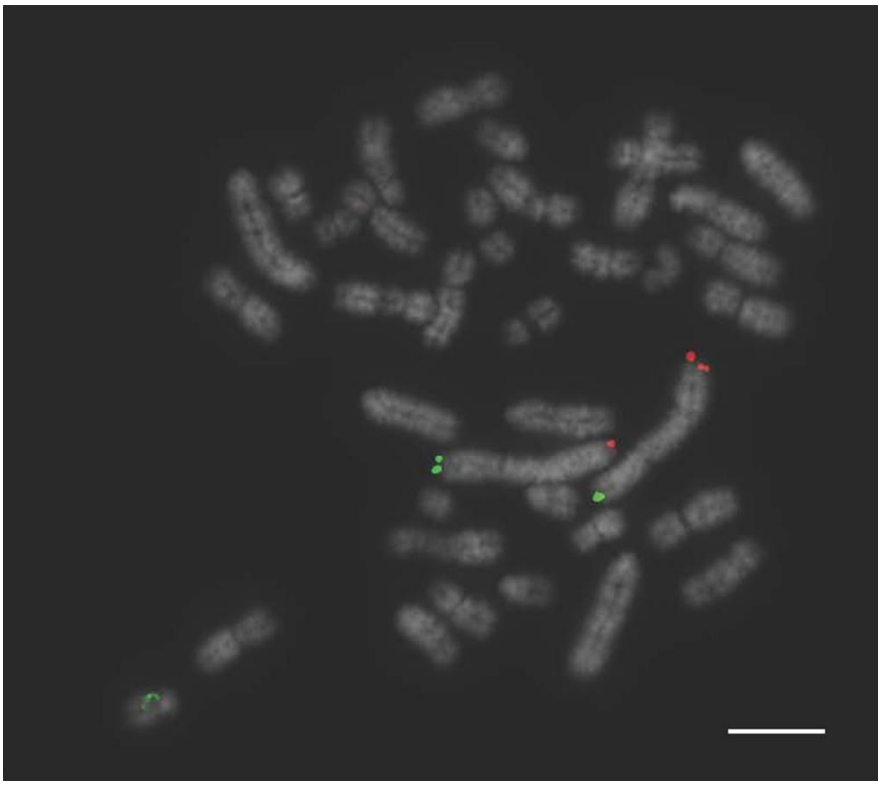


Figure 1-6 Pig chromosome 1 BAC clones CH242-248F13 FITC) and CH242-151E10 (Texas Red) mapping to the correct chromosome and subtelomeric location. Scale bar 10µm (Image source: O'Connor *et al.*, 2017).

Due to their efficiency and diversity of use, a number of laboratories worldwide have generated large BAC libraries in only a number of months, from a number of different complex genomes (Osoegawa *et al.*, 1998) such as humans, cattle, horses, zebrafish and pig. Although FISH with BAC clones is specific, making them the reason they are useful, it can also be their downfall in a comparative study setting, as the probe might not work on cross species (Bishop, 2010). For example, a cattle BAC probe which has been proven to work in the *Bos taurus* (BTA) genome, might not necessarily work on another species of interest, even if the sequence exists. Additionally, BAC clones can be tedious to prepare (Hu, Maurais & Ly, 2020) from the clone stage through to labelling with fluorescent markers; thus, they can be expensive, and

many studies are limited to the probe sets at their immediate disposal (Hu, Maurais & Ly, 2020).

FISH allows researchers to study chromosomal aberrations without intensive cytogenetic training or instruction (Hu, Maurais & Ly, 2020), unlike karyotyping. This ease of use has encouraged its application worldwide in over 400 laboratories for cytogenetic testing (Gersen & Keagle, 2013). Gene fusions, deletions/insertions, translocations, and aneuploidy identified with FISH, aid medical professionals in diagnosing and treating genetic diseases and various cancers (Bishop, 2010; Gersen & Keagle, 2013).

1.1.3 Whole genome sequencing and chromosome level genome assembly

The advent of DNA sequencing transformed the age of genetics presenting scientists the opportunity to study genes and chromosomes at base pair (bp) resolution (Hu, Maurais & Ly, 2020). Structural deviations at nucleotide level can be mapped and diseases can be cross referenced to detailed genomes to better understand the functioning of the genes involved (Deakin *et al.,* 2019). This information can be obtained from a variety of different cell types, using a large number of different sequencing technologies with distinct methods of deciphering the DNA involved (Deakin *et al.,* 2019; Hu, Maurais & Ly, 2020). From Sanger sequencing to scaling and commercialisation of high-throughput next generation sequencing (HT-NGS) techniques, many genomes have been sequenced, assembled, and studied with cross reference to the original approaches of physical mapping.

In 1975, the first sequencing technology was developed and demonstrated by Sanger at the Croonian lecture (reported in Pareek, Smoczynski & Tretyn, 2011). The technique involves annealing amplified/complementary DNA (cDNA) to an oligonucleotide primer followed by using a DNA polymerase enzyme to extend it (Crossley *et al.,* 2020). The DNA polymerase adds a combination of four different deoxynucleotide triphosphates (dNTPs) including dTTP, dATP, dCTP and dGTP, or dideoxynucleotide triphosphates (ddNTPs) which terminate the chain

reaction (ddTTP, ddATP, ddCTP and ddGTP), of which one will be rate-limiting (Crossley *et al*., 2020). The rate-limiting dNTP or ddNTP, will stop the elongation reaction, thereby prompting different length DNA fragments (Crossley *et al.,* 2020).

An example of this method is the automated DNA sequencer 96 capillary ABI Prism 3700, which automatically reloads capillaries with a polymer matrix (Pareek *et al.,* 2011). It generates nucleic acid sequences around 800-1000 bp, yet unfortunately, it cannot differentiate single bp differences in segments over 900 bp and it has relatively low-quality reads in the first 15-40 bp due to primer binding (Crossley *et al.,* 2020). Regardless, in 2003, it was used to sequence the initial human genome after 13 years of work, costing $2.7 billion, and later, the first phage genome (Pareek *et al.,* 2011; Phillippy, 2017; Deakin *et al.,* 2019).

In 2000, 454 Life Sciences was improving on the already developed GS 20, a novel next generation sequencing (NGS) machinery which became commercially available in 2005 as the first NGS technology on the market (Pareek *et al.,* 2011). The initial NGS method was based on a technique called pyrosequencing, or shotgun sequencing (Pareek *et al*., 2011). Unlike Sanger sequencing, it is dependent on the chemiluminescent detection of the release of pyrophosphate when nucleotides are incorporated (Pareek *et al.,* 2011). This method combined with single-molecule emulsion PCR, was used to validate the *Mycoplasma genitalia* genome in a single run at 99.96% accuracy (Pareek *et al.*, 2011) and has been used to produce short read sequences which can in turn make it labour intensive and time consuming (Deakin *et al*., 2019). The short sequence scaffolds need to be anchored to chromosomes thus the larger the genome, the more effort it is to do (Deakin *et al.*, 2019).

Another short-read sequence technology, which was altered from Shotgun sequencing, was developed by Illumina, and is considered second generation NGS (Pareek *et al*., 2011). Five hundred million raw sequences can be generated in a single run of this technology, based on parallel cyclic processing of a sample (Pareek *et al.,* 2011). This method relies on clonal amplicons, clonal beads, or a clonal bridge, making it cost effective and diverse due to the high

number of different preparation techniques (Eisfeldt *et al.,* 2019). Yet, like Shotgun sequencing, it is based on a short-read preparation method and thus lacks detailed resolution to refine contigs (Phillippy, 2017).

To improve on short read sequencing, Oxford Nanopore and PacBio both developed third generation NGS sequencers which can generate over 10 Kb reads in a single run (Phillippy, 2017). Due to the length of these reads, highly continuous assemblies can be generated as standard repeat lengths in genomes are shorter than the reads of third generation sequencers and most bacterial genomes are automatically reconstructed (Phillippy, 2017). Although, these new technologies are highly cost-effective, faster, have higher outputs and can often detect whole chromosomes (Pareek *et al.,* 2011; Baudhuin *et al.,* 2015), they are often associated with increased error rates and need improved assembly techniques (Phillippy, 2017). Regardless, rare variants which have previously been missed, are now being detected with third generation NGS due to the small quantity of starting material (a single molecule) needed for a run (Pareek *et al.,* 2011), making them the most sophisticated capillary sequencers on the market.

Results from sequencing methods are combined to create a full genomic arrangement called an assembly (Liu *et al.,* 2009). The main aim of the assembly of a genome is to have a haplotype-continuous, gapless, and errorless construction of sequences which represents a species. (Phillippy, 2017). A telomere-to-telomere complete human genome sequence has now been achieved by Nurk *et al.* (2022) and the Y chromosome by Rhie *et al.* (2023). When piecing a genome together, highly similar DNA sequences from the same area of the genome are overlapped to create one long sequence, thereby including previously possibly missed areas of the genome (Nagarajan & Pop, 2013). The two ways of genome assembly are either through conservatively (overlap-layout-consensus) or aggressively (de Bruijn graph method) merging scaffolds and contigs into an entire genome (Liu *et al*., 2009). The method used normally depends on the read sequence length. For example, de Bruijn graph techniques are

used more commonly for assembling short reads which produce more errors yet are a more statistically contiguous builds; while overlap-layout-consensus methods are used for more long read type assemblies due to the fewer number of false joins, but they may have more scaffolds and contigs unplaced (Liu *et al.,* 2009; Eldridge*,* 2010; Rizzi *et al.,* 2019).

The vast number of sequencing options and assembly tools proves a challenge for scientists when deciding what to use for their experiment and to piece together the genome they are studying (Nagarajan & Pop, 2013). Repetitive sequences are another difficulty with sequence assembly (Liu *et al.,* 2009). They generate regions of genomic repeated segments which look almost identical, yet come from different parts of the genome, complicating assembly further (depending on repeat length) for either automation and/or manual sequence generation (Liu *et al.,* 2009; Brosnahan, Brooks & Antczak*,* 2010; Nagarajan & Pop, 2013). For example, immune system genes which can be highly polymorphic in mammals, can add to the confusion for an automated sequencing pipeline, especially if they previously haven't been discovered (Brosnahan, Brooks & Antczak*,* 2010).

Regardless of the drawbacks of sequence assembly, in less than a quarter of a century, scientists have accurately, and at a high quality, sequenced 3278 different species from 24 different phylogenetic lines which are freely available on GenBank (Hotaling, Kelley & Frandsen, 2021). This number is almost certainly already out of date. Studies using genome assemblies have changed from purely understanding a sequence (Hotaling, Kelley & Frandsen*,* 2021), to full microbial community reconstruction, analysis of transcriptomes and genomic variant discoveries (Nagarajan & Pop, 2013). Using GenBank, structural variations are compared to reference genomes using methods such as the Basic Local Alignment Search Tool (BLAST) and Comparative Genomics (CoGe) platforms.

The ultimate aim of a genome sequencing effort is a so called "chromosome level assembly" with all the sequences correctly assigned to their rightful place in the chromosome (section 1.3.5). Using these types of chromosome-level assemblies allows for not only improving

conservation efforts, ecology, and maintaining biodiversity, but also is incredibly important in agriculture (Griffin & Bruford, 2020) (Table 1-1, following page). For example, chromosome level assemblies are used in plant and animal studies to improve food security worldwide and ensure conservation genetics can take place for endangered species such as the white and black rhino, and the Gyr and Saker falcons (Griffin & Bruford, 2020). DNA libraries, such as CryoArks, ensure the sustainability of the natural world and the species in it by offering scientists the opportunity to sequence and develop accurate chromosome-level maps of normally “difficult to access” species (Griffin & Bruford, 2020). For both agriculture and conservation purposes, studying these types of maps ensures chromosomal rearrangements causing speciation, reproductive isolation (section 1.1.5) and fertility problems (sections 1.1.5, 1.2.2, 1.2.3) can be identified.

Category	How have traditional genetic studies been used previously?	How can good quality whole genome sequencing improve things?	How do we apply the knowledge for "real world" benefit?
The history and evolution of populations	Used extensively but limited in resolution	Improved accuracy, precision and detail. Need fewer individuals to get an answer	Better understanding of the limitations of the approach. Comparative analysis to understand common evolutionary trajectories
Genetic variation and how organismal groups are adapted to their environment	Minimally used, based on comparisons of populations or at looking at small numbers of genes	Improved detection of the genes involved in adaptation and evolution. Management frameworks became possible	Better characterisation of the genes and genomes
Genetic variation of characteristics that can be measured quantitively (like height and weight)	Limited resolution, dependent on pedigrees and often limited to analysis of small numbers of gene and traits with high heritability	Improved detection of novel genes involved in these traits and the causative mutations within those genes	Identification of genes that are commonly involved in particular traits have been revealed (e.g. the RXFP-2 genes in presence of horns in domestic sheep, the EPAS gene in adaptation to high altitude in many species)
Dividing animals into groups e.g. species, genus, family etc. (so-called "Phylogenomics")	Regularly used, moderate resolution, usually based on a few gene sequences	Much higher resolution, especially for speciation events that occurred rapidly (e.g. vertebrate evolution after the K-Pg extinction event caused by the Chicxulub meteor strike)	Can lead to new discoveries in biodiversity (unsuspected evolutionary pathways, new species, speciation genes, comparative analysis)
Diagnostics and environmental DNA	Regularly used, moderate resolution	Improved identification and detection of DNA in the environment, including parasites and pathogens using "metabarcoding"	Actively taken up, more reproducibility needed
Detecting inbreeding, genetic drift and population viability	Used extensively but very limited in resolution. Very rarely used in population viability analysis	Improved estimates. Novel metrics established (runs of homozygosity). Allow specific regions of the genome to be analysed	Better characterisation of the genes and genomes of species which are affected. Encourage demand from practitioners
Genetic monitoring	Used extensively but limited by DNA quality from non-invasively collected samples	More accurate estimates possible, includes presence/ absence of invasive species, population size, dispersal, and movement	More collaboration between labs and practitioners, need for national monitoring programs (policy-driven)

Table 1-1 The various areas of conservation, ecology and evolution generally studied, methods in which these areas of research can be aided by genomic studies, and how they can be applied globally, are described in the table (Source: Griffin & Bruford, 2020).

1.1.4 A brief overview of patterns of change in mammals

To further understand the cause of genetic mutations an animal or human may carry, causing a possible medical issue, comparative mapping studies using ISH processes have allowed scientists to investigate economically- and/or disease-important traits/phenotypes through cross species studies (Graphodatsky, Trifonov & Stanyon, 2011). Some well-studied comparative models include dog-human and pig-human due to their genetic similarities. Phylogenetic relationship studies to infer karyotype evolution of organisms which are closely related has allowed scientists to build comparative maps for agricultural, laboratory and companion mammals (Graphodatsky, Trifonov & Stanyon, 2011). These studies indicate what parts of chromosomes have evolved, are shared and have been rearranged through evolution and can give an indication of how traits and phenotypes have been derived (Schoen, 2000; Graphodatsky, Trifonov & Stanyon, 2011). For example, closely related species such as cattle and bison, often share larger segments of chromosomes, of which big portions are highly conserved, yet more divergent mammals (such as humans and mice) in general, only share smaller chromosomal regions of shorter distances (Schoen, 2000). Cytogenetic maps allow for the study of genome sequences in these various mammals by mapping exact physical positions on chromosomes, allowing for calculation of linkage disequilibrium for rare or important alleles and for either confirming or disproving a genomic sequence (Farhadi *et al.,* 2013).

Cross species chromosome painting and individual locus FISH probes have allowed for a more accurate measurement and assessment of chromosomal changes in mammals (Graphodatsky, Trifonov & Stanyon, 2011) especially when compared to standard classical cytogenetic karyotyping (Kubiak *et al*., 2020). When species with highly rearranged karyotypes need to be investigated, such as the dog, often chromosome painting probes from a highly studied species like cattle, are preferred over individual locus FISH (Yang *et al.,* 2004). This is because a singular chromosome (or multiple if multi-colour FISH is used) of one species can then be visualised in the other species' chromosomal complement with high accuracy (Graphodatsky, Trifonov &

Stanyon, 2011). If single locus FISH is used, many different probes will need to be positioned along a singular chromosome to get a significant effect, which can not only be expensive, but also more laborious. Moreover, in general terms, while chromosome painting has been effective for “cross species” research in mammals, use of individual clones such as BACs has been limited mostly to birds (Larkin *et al*., 2006; Kiazim *et al*., 2021).

When comparing DNA sequences in cross species FISH, functional genomic sequences are often highly conserved due to their slow rate of evolution, when compared to non-functional sequences (Frazer *et al.,* 2003). Thus, coding regions can be increasingly identified, along with their regulatory functions, in closely related species as well as distantly related mammals (Frazer *et al.,* 2003). This can aid in deciphering divergence or convergence of species and genes from a related ancestor (Frazer *et al*., 2003). Accordingly, these genes are called homologs, reflecting a shorter time since divergence (Frazer *et al.,* 2003). Depending on how homologous genes are inherited, they can either be considered orthologs (genes from two different species which are syntenic conservation in their ancestral species) or paralogs (genes which have different functional elements yet come from a duplicate gene pair) (Frazer *et al.,* 2003). Due to the increased divergence time of paralogous genes, they do not have many conserved sequences when compared to orthologs, which are considered to be increasingly relevant for evolutionary studies, since there is increased probability of conservation of genomic regions (Frazer *et al*., 2003).

Another series of coding sequences which retain function, hence are highly conserved, are protein coding sequences (Frazer *et al*., 2003). FISH probes specific to these sequences, along with repetitive sequence elements, often particular to centromeric/telomeric regions, are used to target and identify functional genes and chromosomal fusions, respectively (Hu, Maurais & Ly, 2020; Frazer *et al.,* 2003). Through combinations of these techniques, chromosomal conserved elements within increasingly divergent species allows researchers to assemble genetic maps of individual species and taxa (Schoen, 2000). If these regions are well studied in

one species, for example the human, and the same sequences are found to have high linkage disequilibrium in another species, there is an increased probability the genes and markers have been conserved in gene blocks, with little rearrangement (Schoen, 2000). These gene blocks are useful for conservation purposes within a species, as well as the study of mammals with little known genomic data.

Cattle (BTA) are one of the most studied agriculturally important mammals. The cattle karyotype has 29 acrocentric autosome pairs and a pair of sex chromosomes (2n = 60) (Figure 1-3A) which is almost identical compared to their main ancestor, the Pecora infraorder, 2n = 58 (Gallagher Jr. *et al.,* 1999; Frohlich *et al.,* 2017). They are found within the Artiodactyla clade and are part of the Bovidae family (Gallagher Jr. *et al.,* 1999). When studied, there was a distinct lack of autosome rearrangements (bar tandem and centric fusions) found within the higher Cetartiodactyla family (Perucatti *et al.,* 2012; Frohlich *et al*., 2017) making them ideal comparative study specimens for mammals with minimal genetic information about them. Compared to the autosomes, any rearrangements seen on the X chromosomes are due to the repetitive sequences found within heterochromatin blocks (Frohlich *et al.*, 2017) which occur much more frequently with cattle. Gallagher Jr. *et al*. (1999) also demonstrated how autosomal rearrangements in the Bovidae family are primarily by Robertsonian centric fusion, while sex chromosomes aberrations were more diverse. Additionally in a study by Larkin *et al*. (2006), seven cattle BAC clones found and mapped on chromosome 19 were linked to human (*Homo sapiens* - HSA) chromosome 17 and mouse (*Mus musculus* - MMU) chromosome 11, indicating how versatile cattle BAC clones are in divergence and evolutionary studies.

To map intrachromosomal breakpoints and rearrangements in evolution, high numbers of cattle BACs distributed across the genome have allowed for accurate mapping of phylogenetic changes (Frohlich *et al.,* 2017) especially for within family species such as Bison (Bonvini subfamily with cattle) and domestic sheep (Caprinae) (Gallagher Jr. *et al.,* 1999). *Bos bison* (BBI - North American bison) and cattle, which are only 1-1.5 million years divergent from one

another, both have the same number of chromosomes (2n = 60), and viable progeny are born from cross species matings with cattle are known to occur (Mikko *et al*., 1997). Further information on this can be found in the next section 1.1.5.

In contrast to the similarity between Bison and cattle, deer, part of the Cervidae family, have a wide diversity of karyotypes ranging from 2n = 70 (*Mazama gouazoubira*) to 2n = 6 (*Muntiacus muntjak vaginalis*) due to a range of tandem fusions and Robertsonian translocations (Frohlich *et al.,* 2017). In a study by Frohlich *et al.* (2017), a combination of 29 cattle whole chromosome paints isolated through flow sorting, as well as 46 cattle BAC probes (chromosomes 1, 3 and X), were used to study 9 different deer species on metaphase preparations. They found 35 conserved chromosomal regions between the nine deer species and cattle, which was novel considering only karyotypic investigations had previously been conducted on cervid species (Frohlich *et al.,* 2017). Cattle chromosome 1 was found to have had complex rearrangements in Cervidae as well as being split into two chromosomes of varying sizes, which is expected as the majority of orthologs found on this chromosome have undergone inversions, translocations, and disruptions through evolution (Frohlich *et al.,* 2017). To improve this chromosome map in deer species, further high-resolution BAC mapping as well as genome sequencing would allow smaller chromosomal changes to be studied, yet many of these species have not yet been sequenced or assembled (Frohlich *et al.,* 2017).

Ovis aries (OAR - domestic sheep) is another comprehensively studied agricultural species of economic importance with 2n = 54 chromosomes also within the Bovidae family (Farhadi *et al.,* 2013). Sheep are well known to have a high number of fecundity genes which are of major interest to breeders, yet specifically in a study by Farhadi *et al.* (2013), the location of growth differentiation factor (GDF9) and bone morphogenetic protein 15, as well as its one receptor (BMP15 and BMPR1B), were investigated using FISH cattle BACs. GDF9 was mapped to cattle chromosome 7q conversely to sheep chromosome 5q, while BMP15 was found on both cattle Xq and sheep Xq (Farhadi *et al.,* 2013). Similarly, BMPR1B was also found on the same

chromosome and arm (cattle/sheep chromosome 6q) in both species. These similarities are to be expected due to both species being part of the same family group, yet this finding allows researchers to further map the areas surrounding these genes to test for conserved gene or haplotype blocks, as well for other genes of interest.

1.1.5 The similarities of chromosome change in evolution and as a cause for infertility in individual animals

Throughout evolution, chromosomes have evolved through various population dynamics and hybridisation. Reproductive isolation, which occurs when a particular species cannot successfully breed with another due to genetic, physiological, behavioural, or geographical reasons, is a primary reason causing infertility in various evolutionary cases. For example, due to differences in chromosomal arrangements, sperm morphology and phenotypes, female horses that mate with male donkeys produce completely sterile, but healthy, mules (Yang *et al.,* 2004; Kubiak *et al.,* 2020). Although viable offspring are produced from these parental species, their reproductive compatibility has decreased to a point of nonexistence, even though they are closely related to one another (Kubiak *et al.,* 2020). This is a form of reproductive isolation. For example, sterile hybrids between horses and donkeys have

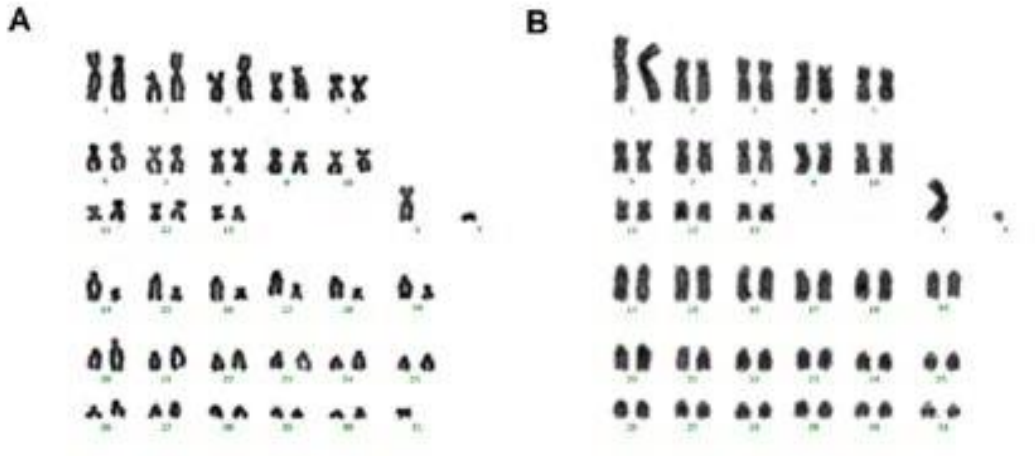


Figure 1-7 A hinny/mule karyotype of 63, X (A) when compared to a stallion of 64, XY (B) (Image source: (A) Terje Raudsepp personal communication; (B) modified from Li *et al.,* 2015).

occurred throughout evolution, creating both mules (2n =63) (Figure 1-7A) which are the offspring of a male donkey and a female horse (mare), and hinnies (also 2n = 63) (Figure 1-7B), which are the progeny of a male horse, a stallion, 2n = 64 and female donkey 2n = 62.(Han *et al*., 2018).

As the time of divergence between species increases, the genetic differences increase and hence reproductive compatibility decreases causing the chance for speciation to occur, to decrease exponentially (Kubiak *et al.,* 2020). A better understanding of what variables cause speciation and/or reproductive isolation, is vital to comprehend evolution as these systems have shaped phylogeny globally (Kubiak *et al.,* 2020). An example of how initial reproductive isolation merged into speciation and consequently created the bison we see today is the hybridisation between cattle and bison. This speciation event in the late 1800s shaped the modern-day bison (Halbert & Derr, 2007). Bison are generally discordant to cattle, producing sterile offspring, yet through a small selection of fertile F1 hybrid progeny bred through advanced backcrosses, morphologically identical bison were created, saving the species from a dramatic bottleneck (Halbert & Derr, 2007). A potential explanation for fertile hybrid establishment, could be due to the high number of breakpoint hotspots and thus chromosomal rearrangements within the Pecoran phylogeny (Frohlich *et al*., 2017). Consequently, one particularly common rearrangement found within the bison-cattle F1 population, such as the 1:29 translocation found in cattle, created a fertile population of hybrids with a fixed homozygous chromosomal arrangement. The 1:29 cattle translocation is further discussed in section 1.3.1.

Humans (2n = 46) have also undergone speciation and reproductive isolation from their predecessors, the great apes (2n = 48). Genetic chromosomal rearrangements due to potentially geographically overlapping (sympatric) or contiguous (parapatric) populations have created mutations to an extent whereby they could not flow from one population to another (Ayala & Coluzzi, 2005). Through this process the mutations accumulated within individual

populations eventually leading to the creation of different species altogether (Ayala & Coluzzi, 2005). This theory of speciation is known as the "suppressed recombination" model that was tested and proven between chimpanzees and humans through comparative studies and gene testing by Ayala & Coluzzi (2005) and others. Nine chromosomes (1, 4, 5, 9, 12, 15, 16, 17 and 18) have undergone pericentric inversions between chimpanzees and humans, and human chromosome 2 is a fusion between chimpanzee chromosomes 12 and 13 (Fan *et al.*, 2002; Ayala & Coluzzi, 2005), which aids in the hypothesis whereby early fixation of chromosomal arrangements created individual populations of apes, eventually leading to speciation (section 1, Figure 1-2).

1.2 Chromosome rearrangements in humans with a specific emphasis on infertility

1.2.1 Types of chromosome rearrangements in general

Within a karyotype, numerical changes are normally due to losses and gains of entire chromosomes (Powell, 2013). Depending on the number of chromosomes present, the mammalian chromosome set can be considered normal (diploid) (Figure 1-8) and contains two sets of chromosomes in pairs (2n) (Powell, 2013). Any deviation from this is considered to be abnormal, or aneuploid (Figure 1-8), and can consist of monosomic (e.g., 45, X in humans), trisomic (e.g., 47, XXY or trisomy 21 in humans) or other variations from the expected normal number of chromosomes for any given species (Powell, 2013; Hu, Maurais & Ly, 2020). The consequences and examples of aneuploidy are further discussed in section 1.2.2.

Numerical and other rearrangements are common problems in mammalian species (Rambags *et al.,* 2005). In a certain set of circumstances, polyploidy occurs, which is when a cell contains more than two whole genomes, e.g., 69, XXY in humans (Madlung, 2013) (Figure 1-8). As discussed, in most mammals, diploid cells are the norm and each species has its own unique set of chromosomes, thus polyploid states are unusual. Although polyploidy isn't often found

in animals, it seems to have been a driving force behind the evolution of most plants and has a major benefit of ensuring fitness due to the sheer number of alleles available, thus the ability to avoid inheritance of adverse recessive mutations (Madlung, 2013).

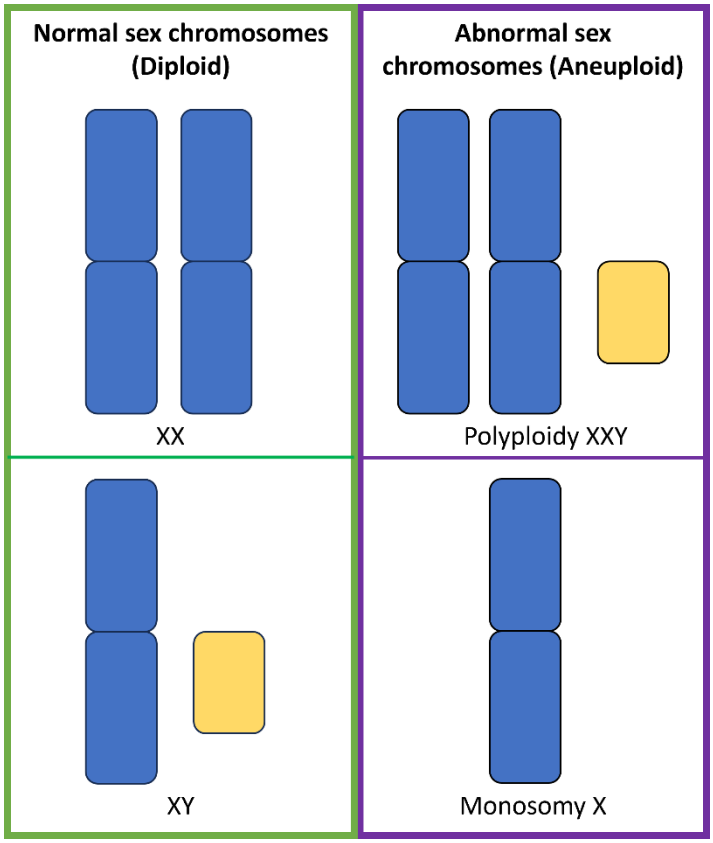


Figure 1-8 Differences between normal diploid (2n) chromosomes when compared to abnormal aneuploid chromosomal states such as polyploidy or monosomy (Own figure).

Within human and other mammalian populations, structural rearrangements are more common abnormalities than numerical ones, as they are compatible with life. Due to this, many of these abnormalities can go unseen for decades, and are only discovered when couples or animals struggle to conceive (Burssed *et al.,* 2022). If a structural chromosomal abnormality occurs, it will be classed as either an unbalanced or balanced rearrangement and can happen between autosomes and/or sex chromosomes (Kaiser-Rogers & Rao, 2013). Individuals who have an unbalanced rearrangement, have either gained or lost genetic information (Kaiser-Rogers & Rao, 2013), and often have significant phenotypic flaws or are miscarried (Powell,

2013). Meanwhile, within balanced abnormalities, no genetic information is lost or gained, hence most mammals in this case are phenotypically normal (Kaiser-Rogers & Rao, 2013).

Some of the most common structural rearrangements within mammals are deletions, insertions, inversions, duplications, and translocations. When the loss of a chromosome segment occurs, it is known as a deletion; and accordingly, an insertion is the gain of a chromosomal segment (Kaiser-Rogers & Rao, 2013). If a chromosome segment reinserts itself after breaking and reorientation by 180° occurs, an inversion has taken place. These can be either paracentric (a singular arm without involvement of the centromere) or pericentric (involvement of the centromere) (Kaiser-Rogers & Rao, 2013). Duplications are another form of unbalanced rearrangements, whereby a chromosomal segment is replicated. Similarly, an isochromosome has two identical arms, due to one arm as a duplication, and one as a loss (Kaiser-Rogers & Rao, 2013).

One of the most researched rearrangements are translocations. This is an irregularity where two chromosomes switch genetic information and can be considered to be unbalanced or balanced, depending on the breakage occurring (Kaiser-Rogers & Rao, 2013). Robertsonian translocations are the most common (~0.1% of human births) and consequently most studied of the different chromosome abnormalities (Zhao *et al.,* 2015). They occur when there is a fusion of the long arms of two acrocentric chromosomes, giving the appearance of a single chromosome with one centromere (Kaiser-Rogers & Rao, 2013) (also see section 1, Figure 1-2). Due to the repetitive gene clusters located on human acrocentric chromosomes, many of these translocations have no phenotypic ramifications (Kaiser-Rogers & Rao, 2013) and can often be difficult to visualise under classical cytogenetic techniques.

1.2.2 Chromosome rearrangements that cause infertility

Rearrangements in the early stages of pregnancy, during gametogenesis or at fertilisation, especially within the embryo in the uterus, can be detrimental (Rambags *et al.*, 2005). For example, males and females who survive past the *in-utero* stage, who have an unbalanced rearrangement, normally have developmental delay and/or many congenital aberrations (Powell, 2013). While humans who have balanced rearrangements generally have fertility issues or are completely sterile (Powell, 2013), due to failure of their germ cells to recombine correctly. Parental-based chromosomal rearrangements are found in 6.65% of couples who suffer from recurrent pregnancy loss, whereby 2-4% of repeat pregnancy losses are specifically due to Robertsonian or balanced reciprocal translocations (Bhatt & Agarwal, 2020).

Infertile men in particular, normally have numerical sex chromosome abnormalities such as Klinefelter syndrome (47, XXY) which is found in 10% of severe oligozoospermic (<15 million sperm/mL) patients and in 5% of severe azoospermic (no sperm in the ejaculate) after cytogenetic analysis (Chakraborty *et al*., 2021). Balanced Robertsonian translocations have also been found in azoospermic patients and in general are some of the most common chromosomal aberrations in infertile males (Chakraborty *et al.,* 2021). Many balanced translocation carriers are phenotypically normal as the breakpoints of these rearrangements are found in repetitive areas of the genome and, consequently, they don't affect the expression of genes within the rearranged chromosomes (Chakraborty *et al*., 2021). Because of this, balanced translocation carriers can on occasion produce progeny without former clinical diagnosis of their genetic abnormality, and unfortunately, due to the high risk of meiosis segregation errors, early infancy mortality is common (Chakraborty *et al*., 2021).

In live-born humans, only three different autosomal trisomies have been found on chromosomes 13, 18 and 21 (Ghosh *et al*., 2022). These numerical abnormalities account for 60-80% of all chromosomal abnormalities associated with early pregnancy losses in the first trimester of pregnancy (Bhatt & Agarwal, 2020). Rambags *et al*., (2005) also noted numerical

changes in chromosomes have also been found in live pig, sheep, and cattle births with varying degrees of ploidy. Of the numerical changes found in mammalian chromosomes, sex chromosome aberrations (45,X; 47,XXX; 47XXY; 47, XYY) are the most common, found at a frequency of 0.2% of all humans (Powell, 2013). This could be due to the increased compatibility with life when compared to autosomal aneuploidies (Powell, 2013), in part because any more than one X chromosome is partly functionally inactivated.

Aneuploid cells are due to incorrect segregation of the chromosomes and subsequent conceptuses often don't survive, either leading to embryonic arrest at the preimplantation stage, failure of implantation or first trimester spontaneous abortion (Ly *et al.,* 2019). Indeed, aneuploidy is the leading cause of IVF failure and pregnancy loss in humans (El Hachem *et al.,* 2017). In rare cases, aneuploid conceptuses survive to term, leading to mental retardation, congenital birth defects and common problems such as Down syndrome (Hassold, Hall & Hunt, 2007).

1.2.3 Why do chromosome rearrangements cause infertility (and reproductive isolation)?

During oogenesis or spermatogenesis, gametes with abnormal chromosomes develop if segregation of chromosomes has occurred incorrectly (Bugno, Jabłońska, Z., & Słota, 2009). If these gametes take part in fertilisation, the resulting embryo will have a rearrangement in respect to genetic mutations or development of a new allele and/or numerical or structural aberrations, which can ultimately lead to embryonic mortality or congenital malformations (Bugno, Jabłońska, & Słota, 2009; Bhatt & Agarwal, 2020). Specifically, during meiosis, when a pair of homologous chromosomes segregates incorrectly such that the pair do not move to a newly formed cell together, nondisjunction occurs and in mosaic individuals, this occurs during mitosis (Sparkes & Crandall, 1972). Nondisjunction is also the cause of most aneuploidies (Bhatt & Agarwal, 2020) and structural abnormalities, which can further interfere with pairing

(Sparkes & Crandall, 1972). It is also the leading cause of the birth of trisomic individuals such as trisomy 13, 18 and 21 (Sparkes & Crandall, 1972) (Figure 1-9). Individuals with any kind of chromosomal abnormality will struggle to conceive due to the unbalanced number of chromosomes within their cells, causing subfertility or complete infertility.

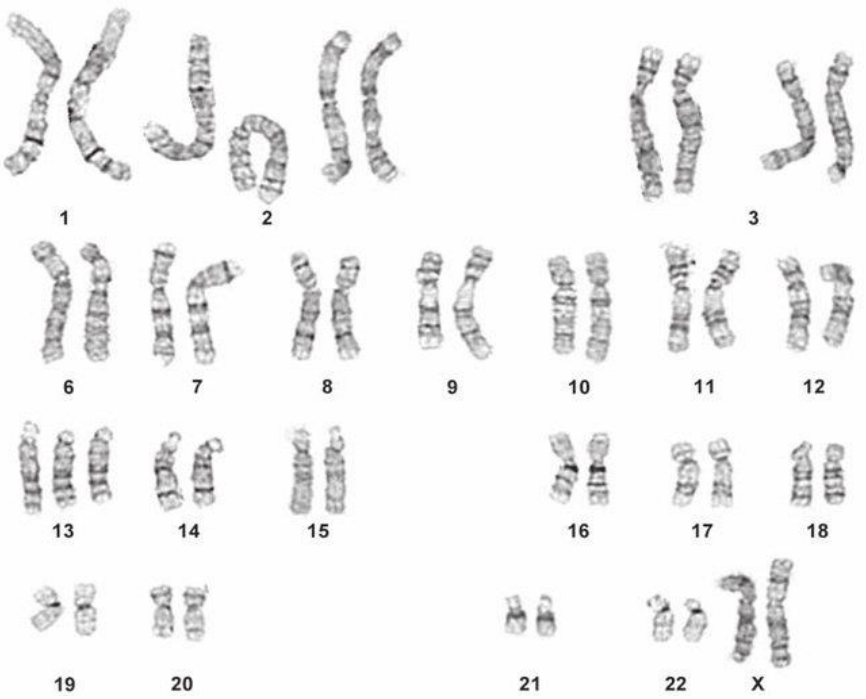


Figure 1-9 Human karyotype demonstrating trisomy 13 (Image source: Hill-Baskin, Lander & Nadeau, 2006).

Furthermore, if structural breakpoints are found in coding regions, they can cause major phenotypic disorders (Burssed *et al.,* 2022). This is because the double stranded breaks (DSB) in the DNA which cause incorrect recombination or repair processes, can create an array of abnormalities within the cells (Burssed *et al*., 2022) (section 1.4.4). Pathological *de novo* causes of these breaks include ionising radiation, DNA stress (mechanical or physical) or oxidative metabolism (Burssed *et al.,* 2022) (section 1.4.3); yet DSBs can also be inherited from parental germ line cells. If a structural chromosomal abnormality occurs, it can happen between autosomes and/or sex chromosomes (Kaiser-Rogers & Rao., 2013).

1.3 Chromosome rearrangements in domestic animals and their use in breeding

1.3.1 Cattle breeding and the 1:29 translocation

In the UK, only just over approximately 50% of the domestic dairy cattle calve each year, even though worldwide cattle production has more than doubled since the 1960s (Jennings, Griffin & O'Connor, 2020). This is a worrying statistic with regards to losses and fertility issues within the population. The low average calving rate could be an indication of the use of sub-fertile bulls within the population (Jennings, Griffin & O'Connor, 2020). This causes a reduction in both meat and milk production sectors and genetic improvement (Jennings, Griffin & O'Connor, 2020). Some of the main causes of subfertility in cattle, as with humans and other mammals, is chromosomal translocations, yet the occurrence of these in the cattle population are underreported (Jennings, Griffin & O'Connor, 2020). This is mainly due to the complex acrocentric heavy karyotype of cattle, and thus, there is high ambiguity when distinguishing between the various chromosomes and translocations are consequently difficult to identify (Jennings, Griffin & O'Connor, 2020).

If sub-fertile bulls are selected and used for breeding, large environmental and financial losses can be seen within the industry (Jennings, Griffin & O'Connor, 2020). In a recent study conducted by Lewis *et al*. (2022), if one reciprocal translocation in a herd is not found, it could cost £7.2 million (incidence of reciprocal translocation predicted at 1.2%). The most common translocation found in cattle is the Robertsonian translocation 1:29 (Jiménez *et al.,* 2022). It is predicted to make up for 0.4% of the translocations in the cattle population (Lewis *et al.,* 2022). Each heterozygous 1:29 translocation impacts the conception rate by 5.1% and by identifying a bull with this issue, a herd breeder could benefit by approximately £2.3 million in a singular generation interval of six years (Lewis *et al*., 2022). A recent study by Jiménez *et al.* (2022) indicated conception rates did not change in populations where the 1:29 translocation was present. This could be an early indication of the homozygous form of 1:29 fixating in the

general cattle population as a new subspecies, eventually leading to evolutionary speciation completely. With the cattle industry having a net worth of £671 billion for dairy cattle and £379 billion for beef (Lewis *et al.,* 2022), it is vital to ensure increased productivity in which more cows conceive through the improved use of artificial insemination (AI), *in vitro*-produced embryos and genetically superior fertile bulls (Jennings, Griffin & O'Connor, 2020). Furthermore, reduction of the number of bulls with the heterozygous 1:29 could and thus far, has, improved conception rates and average calving rate (Jiménez *et al.,* 2022).

Similar to humans, FISH is applied in fertility research, as it is well known that chromosomal aberrations are directly related to high mortality of embryos in mammals, specifically due to the formation of abnormal gametes inevitably causing mutations (Bugno, Jabłońska & Słota, 2009; Jennings, Griffin & O'Connor, 2020; Kubiak *et al.,* 2020). However, chromosomal abnormalities can be identified with individual locus FISH. A cattle screening device using BACs selected on sequence alone was recently developed for commercial use, identifying both homozygous and heterozygous 1:29 translocations, which, as previously mentioned, has already saved the industry large amounts of money (Jennings, Griffin & O'Connor, 2020; Lewis *et al*., 2022). This device was based on novel technology developed by O'Connor *et al*. (2017) which is further discussed in section 1.3.2.

1.3.2 Pig breeding and the FISH revolution

In agricultural screening, both Jennings, Griffin & O'Connor. (2020) and O'Connor *et al.* (2017) noted the need to address chromosomal translocation screening in bulls and boars (*Sus scrofa*, SSC), respectively, due to major economic losses in both industries due to embryonic losses. By using sub-telomeric individual locus FISH BAC probes, they created novel devices to allow for efficient, fast, and inexpensive cryptic translocation screening in the respective species. Each chromosome has a p- and q-arm sub-telomeric probe FITC or Texas Red probe, therefore allowing for easy identification of both overt and cryptic chromosomal rearrangements by

using a fluorescent microscope (O'Connor *et al.*, 2017; Jennings, Griffin & O'Connor, 2020) (Figure 1-10). Wolff (2013) also noted this useful technique in a book by Gersen & Keagle (2013).

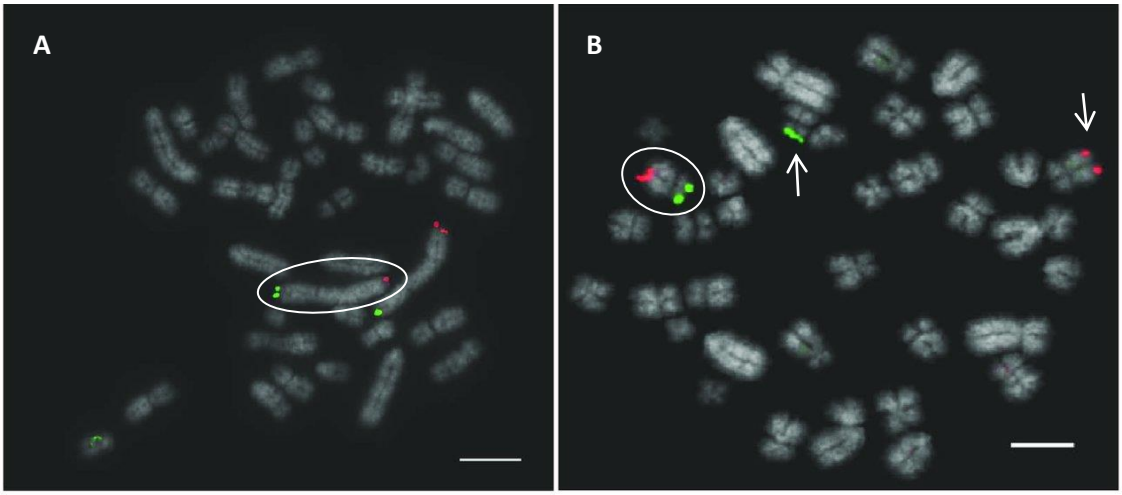


Figure 1-10 Images of pig chromosome 1 subtelomeric BAC probes CH242-248F13 (FITC) and CH242-151E10 (Texas Red) (A). A translocation between pig chromosomes 10 and 7 is indicated with arrows in (B). Porcine chromosome 10 probes CH242-451I23 (FITC) and CH242-517L16 (Texas Red) probes were used for image B. Normal chromosomes are circled in both (A) and (B). Scale bar is 10 µm (Image source: O'Connor *et al*., 2017).

The device developed by O'Connor *et al.* (2017) was created for use on purebred genetically superior boars which had reduced litter sizes; yet no abnormalities were detected with standard karyotyping or in semen parameters, which are the customary ways of identifying a fertility issue in boars (O'Connor *et al.*, 2017). More than 130 balanced chromosomal rearrangements had been identified in pigs with 0.47% of AI boars being affected and thus hypoprolific and exhibiting a reduced litter size (O'Connor *et al*., 2017). Based on research, around half of the boars which show subfertility carry a reciprocal translocation and the only phenotypic indication of this issue is a litter reduction (O'Connor *et al.*, 2017). As discussed in section 1.2.3, due to aneuploidy and unbalanced gamete production, embryo mortality occurs in most mammals resulting in the loss of a single embryo, yet, as pigs are multiparous animals, this is seen as a zygote litter loss of around half the piglets which can cause major financial and environmental issues for breeders and producers worldwide (O'Connor *et al*., 2017).

Using the newly developed multiprobe device, cryptic translocations previously missed by karyotyping were identified and ones diagnosed with karyotyping such as pig t(7:12) and t(1:2) were confirmed demonstrating the efficiency of such a device (O'Connor *et al*., 2017). Additionally, as karyotyping is subjective and relies on good quality chromosomal preparations, FISH devices provide a clear and objective approach to defining translocations and other chromosomal rearrangements. By missing a reciprocal translocation in a terminal boar, with karyotyping or other cytogenetic techniques, could cost a weaner production company £69,802 and a purebred genetics company focusing on the dam line boar, £51,215,378 (Lewis *et al.,* 2021), further emphasising the need to screen other livestock species, such as sheep and horses for reciprocal translocations.

1.3.3 Chromosome abnormalities in horses

Cytogenetic techniques are also vital for mammals such as horses (*Equus caballus* - ECA) due to how expensive they are to buy and keep, with the UK economic equestrian sector currently valued at £4.7 billion (British Equestrian Trade Association, 2022). Karyotyping and chromosome painting have been the main form of routine cytogenetic screening for horses over the past couple of decades, and in recent years, sequencing of the horse genome has given researchers the opportunity to do further in-depth research into the specifics of mapping chromosomes. Karyotypically normal horses have 13 pairs of metacentric chromosomes, 18 pairs of acrocentric chromosomes and one pair of sex chromosomes (2n = 64).

Ghosh *et al*. (2022) demonstrated the need for DNA analysis when analysing abnormalities found in karyotypes in horses. This was not only due to the cryptic nature of the rearrangement, but also to identify the origin of it, maternally, paternally or *de novo* (Ghosh *et al.,* 2022). Many chromosomal rearrangements have previously been found in horses, which have only been identified due to either reduced fertility when mature or developmental/ congenital disorders discovered at birth (Rambags *et al.,* 2005; Pieńkowska-Schelling, Kaul &

Schelling, 2020; Ghosh *et al*., 2022;). Yet these would not have been found without these pathological conditions as indicators, as chromosome screening, and karyotyping, are not standardised analyses conducted prior to breeding a horse. Equine breeding is a big economic undertaking and reproductive performance is the main component for a successful business (Pieńkowska-Schelling, Kaul & Schelling, 2020), as well as improving the welfare of breeding animals by decreasing the number of cycles and reproductive treatments a horse needs to go through. The main generalised routine breeding tests conducted for horses include semen assessments for stallions and currently there are no formal tests on mares except for a basic external and internal reproductive veterinary examination (Pieńkowska-Schelling, Kaul & Schelling, 2020). Due to the influence genetic abnormalities can have on equine fertility, and the pressure of breeding a healthy foal, combined with the long generation interval of a horse (~5 years), it is crucial for routine cytogenetic screening to be done.

The main cytogenetic problems horses have, based on the literature, is almost always related to, or involving, the sex chromosomes. Sex chromosome aneuploidy and X monosomy are the highest reported chromosomal problems, specifically in mares due to clear gonadal dysgenesis (Lear & Layton, 2002; Pieńkowska-Schelling, Kaul & Schelling, 2020). Demyda-Peyrás *et al*. (2014) however noted, many stallions remain undiagnosed as carriers of chromosome rearrangements as they are asymptomatic. This could be a major contributor to underreported early embryonic losses, which most breeders immediately predetermine as the mare. With the vast extent of AI in crossbreed (“sport”) horse breeding worldwide (Aurich, 2012), one stallion could cause significant financial losses for breeders if they were to be a carrier of a chromosomal abnormality.

De novo chromosomal abnormalities, such as a foal with a chromosome 26 acrocentric autosomal trisomy (where 2 out of three chromosomes were fused) studied by Ghosh *et al.* (2022), and an infertile mare with an t(1;16) translocation found by Lear & Layton (2002), are just two examples where abnormalities were found with cytogenetic screening in equines (for

further information, please see section 4). Translocations in horse t(X;15), t(1;3) and t(1;30) have also been described (Lear & Layton, 2002). As with other mammals, chromosomal rearrangements of any kind yield mainly unbalanced gametes thereby causing early death of the embryo, hence *de novo* chromosomal abnormalities are more common than inherited ones, purely due to the nature of replication in the gametes of abnormal carriers.

Another issue in the equine cytogenetic screening of horses, is the lack of clinical knowledge regarding the wide variation of phenotypes witnessed with X chromosome abnormalities (Pieńkowska-Schelling, Kaul & Schelling, 2020). For example, when mosaicism is detected, such as a 64, XX/63, X mare, many horses do not show any phenotypic evidence of subfertility prior to breeding (Pieńkowska-Schelling, Kaul & Schelling, 2020). As with mares, stallion chromosomal abnormalities can also go undetected most of their lives, due to lack of phenotypic evidence; yet some common fertility issues such as cryptorchidism have no cytogenetic basis to them (Vilar *et al.,* 2018). Further necessitating the importance of chromosomal screening techniques or improved genome sequencing technologies.

1.3.4 Chromosome abnormalities in other animals

As covered above, chromosomal rearrangements during evolution are well described and involve both interchromosomal (between chromosomes) and intrachromosomal (within chromosome) changes. The latter occur more in birds and the former more in mammals (Damas *et al.,* 2016; Kiazim *et al.,* 2021). The use of FISH to further establish intrachromosomal changes has recently been further investigated. Single locus BACs are useful for these studies as breakpoints, or highly conserved regions, can be selected as BAC clone sequences, therefore allowing the pattern of change to be directly visualised and mapped. This technique can also be used for species with complex genomic histories or rare species whose genomes have not yet been sequenced. The production of a set of universally hybridising BACs from a well-

studied species such as humans or cattle, could provide a reference for easier, and more direct, sequence analysis and genome mapping. Consequently, genomic scaffold production by bioinformaticians is made into a simpler task as unplaced scaffold location can be determined.

Damas *et al.* (2016) demonstrated this approach with a set of chicken (*Gallus gallus*) BACs which universally hybridise to most avian species and allowed bioinformaticians to anchor scaffolds to these specific chromosomes without the species' reference genomes. The researchers assembled the endangered peregrine falcon (*Falco peregrinus*) and pigeon (*Columbia lilvia*) genomes to chromosome levels which were comparable to the respective avian reference genomes (Damas *et al.,* 2016). Kiazim *et al.* (2021) further established the use of these 74 universally hybridising BACs by using them on eight poorly studied different bird species' macrochromosomes, thereby reconstructing their chromosomes in combination with mathematical analyses and bioinformatics. Additionally, Kiazim *et al.* (2021) reconstructed a presumed ancestor of these birds using the BAC clones, demonstrating the use of this type of methodology for potential conservation purposes.

1.3.5 The importance of physical genome mapping for chromosome analysis using FISH

As useful as bioinformatic platforms and DNA sequencing technologies are to research and clinical investigations, it has been established that sequence information alone does not always accurately assign physical mapping information and not all genome assemblies are as "chromosome level" as they might first appear (Graphodatksy, Trifonov & Stanyon*,* 2011). Due to the expeditious nature of genomic sequencing, assemblies often contain many errors especially in the subtelomeric and repeated regions of chromosomes (O'Connor *et al.,* 2017; Deakin *et al*., 2019;). O'Connor *et al*. (2017) discovered only 45 BACs (55%) mapped to the correct subtelomeric regions of the porcine genome, out of a total of 82 BACs. Additionally, if

BACs did map to the correct chromosome, they were often not at the q-terminus where they had been predicted to be, indicating a high number of mapping mistakes (O'Connor *et al.,* 2017). BACs which had been placed at the q-terminus were in fact small fingerprint contigs which did not have any orientation or full sequence information when pig genome assembly occurred, therefore they were randomly added to the ends of their corresponding chromosomes (O'Connor *et al*., 2017).

Similarly, Zhou *et al*. (2015) found significant assembly errors had been corrected in the updated cattle genome (Btau4.6) when it was compared to the previous version, UMD3.1. This emphasises the necessity to check for which genome has been studied as well as the need for improved sequence data for many of the previously assembled genomes (Zhou *et al.,* 2015). Ambiguity within genome assemblies can majorly affect the studies which have previously been conducted on those genomes (Zhou *et al*., 2015), especially ones which are purely bioinformatics based. Because of this, a combined approach of physical and genomic mapping may be the way forward (Graphodatsky, Trifonov & Stanyon, 2011). While sequence data cannot explain how chromosomal organization, cell-chromosome interaction and genome by environmental interactions occur, cytogenetic data can fill in these gaps in our knowledge (Deakin *et al.,* 2019). This allows scientists to use cytogenetic resources to align sequencing information with chromosome number, morphology and karyotypic information on how a genome is structured (Deakin *et al.,* 2019). If these facts are unknown for a species, abnormalities and gene changes from a phenotypically normal mammal cannot be distinguished from what is rearranged (Deakin *et al.,* 2019), especially in the case of fertility issues where phenotypic knowledge is only gathered later in life.

Physical maps using cytogenetic resources, represent genomes with respect to reference maps of DNA fragments; and locations of markers are required to create them (Dixit *et al.,* 2014). They indicate exact locations of chromosomal features within the DNA and are necessary when studying evolutionary or comparative genomics due to the implications that rearrangements

can have on the genome and thus the study (Dixit *et al.,* 2014). By combining complex genomic BAC libraries with DNA sequencing technologies, as a combined effort of different laboratories with various resources, cost and time can be reduced all while improving overall accuracy and reliability of genomic assembly (Osoegawa *et al.,* 1998).

1.3.6 Genome sequencing technologies to detect chromosome rearrangements

Due to the advancement to single nucleotide sequencing, inherited conditions within mammalian genomes can be clinically analysed for structural variations and genetic disorders (Baudhuin *et al.,* 2015; Eisfeldt *et al.,* 2019; Hu, Maurais & Ly, 2020). This additionally allows scientists to detect balanced and unbalanced structural changes and rearrangements as well as breakpoints (Eisfeldt *et al.,* 2019; Burssed *et al.,* 2022), yet depending on resolution, this can be a long and expensive process when compared to FISH methodology and/or classical cytogenetic techniques (Hu, Maurais & Ly, 2020).

Breakpoint mapping using NGS has caused interpretation challenges for scientists and often other techniques such as physical mapping, PCR or Sanger sequencing is needed to confirm these nucleotide sequences (Eisfeldt *et al.,* 2019; Burssed *et al.,* 2022) which can be an added time cost due to additional experimentation. Because of this, it might be worth investigating combinations of different techniques to explain variation found in the genome and improve assembly build in the most cost-effective manner (Phillippy, 2017; Eisfeldt *et al*., 2019).

1.4 Sperm DNA damage and its relevance

Thus far, the above has dealt with chromosome rearrangements and their relevance to fertility and evolution. A second way at looking at the whole genome in this context is DNA damage in

the sperm. That is, this is both a well-known cause of infertility (in males of course), as well as a possible driver of chromosomal change during evolution.

Male fertility can be defined as the adequate functioning of sperm and the capacity to produce progeny (Petrunkina & Harrison, 2011). According to statistics by Agarwal *et al.* (2015), unexplained infertility cases account for 25% of the cases globally (Agarwal *et al*., 2022a). The conventional method of assessing fertility is through semen analysis, yet it has been scrutinised for its subjectivity and many scientists and clinicians are looking at other potential techniques for evaluating fertility and unexplained infertility. Before considering the role of DNA damage in the clinic (and during evolution) however, an appreciation of the basic biology, as well as contemporary andrological protocols, is necessary.

1.4.1 The spermatozoa

In gametogenesis, specialised and complex sperm (germ) cells are created (Champroux *et al.,* 2016). In human somatic cells, 146 bp of DNA are wrapped around a histone octamer which is combined into a nucleosome and these smaller nucleosome structures amalgamate to make up the highly organised chromatin loop structure (Fraser, 2004; Champroux *et al.,* 2016). The process of condensing chromatin is complex, which include arrangement of chromatin parts, DNA-binding protein transition, changes in transcription, unravelling of the nucleosome and finally condensation of the chromatin (Fraser, 2004). The condensed chromatin, involved in transcription, gene expression, differentiation, cell division and replication, is attached to the nucleus structurally creating the sperm nuclear matrix (Fraser, 2004; Champroux *et al.,* 2016).

Sex specific differentiation and spermatozoa generation takes place in the germinal epithelium in the seminiferous tubules of the testis through spermatogenesis (Champroux *et al.,* 2016; Patel, Leong & Ramasamy, 2018; Hamilton & Assumpção, 2020), which has three main phases: First, spermatogonia are proliferated and differentiated, next, meiosis occurs turning spermatogonia into spermatocytes and then spermatids and this is also known as

spermatocytogenesis, and lastly, spermiogenesis occurs where spermatozoa are generated (Champroux *et al.,* 2016; Sharma & Agarwal, 2018; Peña *et al.,* 2019; Hamilton & Assumpção, 2020).

Spermatogenesis begins at puberty with differentiation and mitosis of totipotent stem cells (spermatogonia) and this process continues throughout the male life (Sharma & Agarwal, 2018). Spermatogonia proliferate, whereby their chromosomes duplicate, the nuclear envelope breaks down and daughter cells form equally (Sharma & Agarwal, 2018). Each spermatogonia then undergoes spermatocytogenesis (meiosis I and meiosis II) to form one diploid primary spermatocyte, then two secondary spermatocytes and lastly, four round haploid spermatids (Sharma & Agarwal, 2018). During meiosis chromosomes pair, crossing over occurs and then there is an exchange of genetic information to form a genome (Sharma & Agarwal, 2018; Peña *et al.,* 2019). Within the final phase of formation, spermiogenesis, the individual spermatids undergo DNA compaction and elongation as well as gaining the standard morphology characteristics of the sperm cell such as a head (DNA and nucleus), midpiece (mitochondrial activity) and tail (for movement) (Sharma & Agarwal, 2018; Peña *et al.,* 2019; Hamilton & Assumpção, 2020). These mature spermatozoa then move into the seminiferous tubule lumen (Peña *et al*., 2019). To aid the mature spermatozoa in fertilising the oocyte and providing sustenance as well as antioxidants for the sperm, seminal fluid, which is produced by the accessory glands, is combined within the male reproductive tract at different stages of ejaculation (Wnuk *et al*., 2010; Patel, Leong & Ramasamy, 2017). Both spermatozoa and seminal fluid can be highly influenced by environmental, physical, and genetic stressors and factors at any point during this process (Fraser, 2004), hence conducting a semen analysis to ascertain the fertility status of a male is so important.

1.4.2 Semen analysis

Fertility potential of a male mammal is conventionally tested by conducting a semen analysis and is an essential component of not only understanding the biological capacity to produce progeny, but also to find any pathological or clinical issues with the male (Agarwal *et al.,* 2019; Agarwal *et al.,* 2022a; Agarwal *et al*., 2022b). A semen analysis is conducted on an ejaculate which contains fluids the majority of which (70%) come from the seminal vesicles, a quarter from the prostate and ~5% for both the spermatozoa and bulbourethral glands, urethral glands, vas deferens and epididymis, respectively (WHO, 2010; Agarwal *et al.,* 2022b). Due to the various origins of different seminal components, quality of sperm and health of an individual can be vitally analysed (Agarwal *et al.,* 2019). Unfortunately, standard semen analyses are also known to be subjective, have low reproducibility and can sometimes be a poor predictor of an individual's fertility (Jiang *et al*., 2009; Agarwal *et al.,* 2019). Regardless, semen analyses are a significant diagnostic test for evaluation of the fertility status of a male and to help understand the basis of a possible issue (Agarwal *et al.,* 2022a).

A semen analysis is conducted in a laboratory by trained professionals (Fraser, 2004) to ensure a correct protocol is followed and maintained (Jiang *et al.,* 2009). In humans, men are requested to provide a sample after a 2-to-7-day abstinence period and are asked to deposit the whole sample into a sterile container as the initial portion of the ejaculate has the primary concentration of spermatozoa (WHO, 2010; Gersen & Keagle, 2013; Patel, Leong & Ramasamy, 2017; Agarwal *et al.,* 2022a). For most male mammals, including humans, after ejaculation, the semen sample is then put into a 37°C incubator for half an hour to an hour post-ejaculation, consequently allowing for liquefaction to occur (WHO, 2010; Agarwal *et al.,* 2022a). A macroscopic and microscopic manual semen analysis then transpires to evaluate various parameters, although in many laboratories, automated (and thus less subjective) machines such as Computer Assisted Semen Analysis (CASA) are taking the place of manual assessments (WHO, 2010; Amann & Waberski, 2014; Agarwal *et al.,* 2022a). Artificial intelligence programs

have also started to be implemented in laboratories conducting semen analysis in an attempt to improve objectivity, specifically in morphology analysis. This has been shown to have overall morphology accuracy values of greater than 87.5% and 89.9%, depending on the program used (Wang et al., 2019b). However, these types of programs have currently not been approved of for commercial use in various accredited andrology laboratories, and manual analysis is still the gold standard.

In the macroscopic manual evaluation, semen is vortexed to ensure the cellular fraction is resuspended, prior to which volume, pH, appearance, and viscosity are analysed (WHO, 2010; Petrunkina & Harrison, 2011; Gersen & Keagle, 2013; Patel, Leong & Ramasamy, 2017; Agarwal *et al*., 2022a). The second component of a semen analysis is microscopic evaluation of the sperm. Sperm count, concentration and motility are evaluated (WHO, 2010; Petrunkina & Harrison, 2011; Gersen & Keagle, 2013; Patel, Leong & Ramasamy, 2017; Agarwal *et al*., 2022a). The volume of the semen sample reflects the overall quantity of seminal fluid produced while sperm count indicates the total number of sperm in the semen sample (Patel, Leong & Ramasamy, 2017). Sperm vitality specifies the percent of viable sperm in the semen sample while sperm morphology shows the number of normal forms of sperm in the semen sample (Patel, Leong & Ramasamy, 2017). Additionally, the presence of sperm agglutination, white blood cells, general morphology and round cells are also investigated as these could indicate a further issue within the body (WHO, 2010; Agarwal *et al*., 2022a). These parameters as a whole, indicate the general fertility potential of a male as they are compared to the World Health Organization (WHO) gold standard values (WHO, 2010; Patel, Leong & Ramasamy, 2017; Agarwal *et al*., 2022a). Regrettably, due to the subjectivity of many of these assessments, and the high percentage of unexplained male infertility, even patients who have parameters within “normal” gold standard WHO values, often have persistent pregnancy failure or male factor infertility (Wnuk *et al.,* 2010; Petrunkina & Harrison, 2011; Evenson, 2016).

One of the most investigated components of a semen analysis is sperm movement. Many studies have indicated, total motile sperm count (a combination of volume multiplied by concentration and motility) has thus far been the most predictive and accurate element for estimating the fertility potential of a man when compared to the three factors individually (Patel, Leong & Ramasamy, 2017). Sperm motility is very sensitive to ion concentration, lifestyle factors and pH and hence its ability to fertilise the oocyte is highly dependent on its ability to reach the Fallopian tubes (Fraser, 2004; Nowicka-Bauer & Nixon, 2020). Asthenozoospermia, a condition which describes up to a third of all male infertility cases, reflects the state whereby less than 32% of the spermatozoa in an ejaculate have forward progressive movement (Nowicka-Bauer & Nixon, 2020). It has been directly linked to oxidative stress (OS) due to the detrimental effects it has on the flagellum and motility-associated signalling pathways, consequently causing motility dysfunction (Dutta, Majzoub & Agarwal, 2019; Nowicka-Bauer & Nixon, 2020). Furthermore, OS has a major impact on the mitochondrial production of energy for the sperm to physically have forward progression (Nowicka-Bauer & Nixon, 2020).

1.4.3 Oxidative stress and lipid peroxidation

ROS and OS are highly correlated with male infertility (30-80% causation) even though ROS are required for normal functioning of the sperm cell through recycling of free radicals (Nowicka-Bauer & Nixon, 2020). ROS are produced in multiple ways. Oxygen is needed by aerobic organisms as fuel for producing energy and ROS products such as superoxide anion O_2^-, as well as OH^- the hydroxyl radical (from Fenton and Haber-Weiss reactions if Fe^{3+} is present) and H_2O_2, hydrogen peroxide (Agarwal, Samanta & Henkel, 2018; Dutta, Majzoub & Agarwal, 2019; Drevet & Aitken, 2020; Nowicka-Bauer & Nixon, 2020). These three molecules can cause instability in the cell as they try to capture stabilising electrons, hence causing oxidation in stable molecules in the cell (Agarwal, Samanta & Henkel, 2018; Drevet & Aitken, 2020). They

also affect the DNA methylation pattern, causing certain transcription factors and genes involved in DNA methyltransferase to alter the signals the cell receives (Agarwal, Samanta & Henkel, 2018) and can be initiated from the sperm cell internally or from external leukocytes, round germ cells and epithelial cells secreted from various locations in the male reproductive system (Nowicka-Bauer & Nixon, 2020).

The nicotinamide adenine dinucleotide phosphate (NADPH) oxidase (NOX) system and electron transport chain (ETC) are the main endogenous ROS generation pathways in the sperm cell (Agarwal, Samanta & Henkel, 2018; Nowicka-Bauer & Nixon, 2020). The ETC generates ROS through the untimely exit of electrons which in turn react with oxygen instead of water at cytochrome c oxidase complex, forming O_2^- and creating pathological damage (Homa *et al.,* 2019; Nowicka-Bauer & Nixon, 2020). In addition to endogenous sources for ROS, exogenous sources such as radiation, toxins, alcohol and smoking cause adverse health effects on males, and consequently further increase ROS and OS in semen parameters (Homa *et al.,* 2019; Nowicka-Bauer & Nixon, 2020). For example, radiation can cause a reduction in semen quality through disruption of intracellular flow of electrons in membranes and smoking elevates seminal leukocyte and ROS levels (Dutta, Majzoub & Agarwal, 2019). In an OS state, ROS are in excess, and increasingly so with a higher number of OS sperm, when compared to the recycling mechanisms present, therefore placing the cell out of homeostasis and into a stressed condition (Dutta, Majzoub & Agarwal, 2019; Drevet & Aitken, 2020; Nowicka-Bauer & Nixon, 2020).

The problem occurs when ROS overwhelm the normal physiological levels within the spermatozoa, consequently limiting the antioxidant defences (Dutta, Majzoub & Agarwal, 2019; Nowicka-Bauer & Nixon, 2020). The sperm cell is “silent” as it cannot synthesise any new proteins due to the lack of cytoplasm and dense DNA packaging, and consequently lacks the mechanisms needed to repair DNA damage (Drevet & Aitken, 2020). Although the sperm is considered “silent”, various RNAs, specifically micro RNAs (miRNAs) and messenger RNA

(mRNA) in sperm have been investigated in recent years to identify if they have a function in the mature sperm and through the fertilisation process (Kiani & Rassoulzadegan, 2013). The function of these different RNAs is relatively unknown, however, they seem to have a significant effect in the enrichment of the oocyte early during post-fertilisation and in general fertilising capacity (Krawetz *et al.*, 2011). Regardless, the sperm is sensitive to damage in multiple forms, especially with regards to reactive oxygen species, pre-fertilisation. Loss of motility caused through the decreased quantity of adenosine triphosphate (ATP) in the cell generates further axonemal damage, morphology issues in the midpiece and loss of viability (Toor & Sikka, 2019). Furthermore, as the number of antioxidant materials in the cell is already sparse in a mature spermatozoon, oxidisable parts of the cell, such as polyunsaturated fatty acids (PUFAs) in the plasma membrane, can easily be overwhelmed in the cell and cause apoptosis (Agarwal, Samanta & Henkel, 2018; Agarwal *et al.,* 2019; Drevet & Aitken, 2020; Nowicka-Bauer & Nixon, 2020). These PUFAs create a fluid membrane which is needed for fusion of the sperm cell during capacitation, acrosome reaction and fertilisation of the oocyte, hence the detrimental impact from ROS on basic sperm function (Aitken, 2017; Agarwal, Samanta & Henkel, 2018; Agarwal *et al.,* 2019; Dutta, Majzoub & Agarwal*,* 2019). Decohexaenoic acid (DHA) is the most prevalent PUFA in human sperm, accounting for over half of all the PUFAs in the sperm membrane (Nowicka-Bauer & Nixon, 2020).

PUFAs are also highly susceptible to LPO, another form of damage caused by ROS in the sperm (Champroux *et al.,* 2016; Aitken, 2017; Dutta, Majzoub & Agarwal*,* 2019). PUFA methylene groups are situated between unconjugated double bonds, inevitably causing the hydrogens attached to the carbons to be the most susceptible to damage from ROS, resulting in LPO if intracellular ROS grows too high (Dutta, Majzoub & Agarwal*,* 2019). LPO impairs flagellar movement, ATP flow in the PUFA plasma membrane and membrane permeability (Dutta, Majzoub & Agarwal*,* 2019; Nowicka-Bauer & Nixon, 2020), specifically due to lipid aldehyde breakdown products of LPO, such as 4-hydroxynonenal (4HNE) and malondialdehyde (MDA) (Nowicka-Bauer & Nixon, 2020). This accounts for approximately 60% loss of membrane fatty

acids (FAs), hence fluidity loss in the membrane (Dutta, Majzoub & Agarwal, 2019; Nowicka-Bauer & Nixon, 2020). 4HNE and MDA can bind to proteins in the ETC of mitochondria, hence prompting additional ROS in a negative autocatalytic apoptotic cycle in the sperm (Aitken, 2017; Dutta, Majzoub & Agarwal, 2019; Nowicka-Bauer & Nixon, 2020). This cycle is also initiated by increased numbers of dead spermatozoa thereby causing heightened quantities of H_2O_2 in the cell, further aiding in the apoptotic cascade, which has been demonstrated in rats (Nowicka-Bauer & Nixon, 2020). Histidine, lysine, and cysteine residue containing proteins, nucleic acids (mainly guanosine in DNA) and lipids in general, are the main points of attack particular to 4HNE, which in turn can produce more H_2O_2 (Nowicka-Bauer & Nixon, 2020). When there are high levels of ROS and 4HNE/MDA present in the sperm, severe DNA damage can occur due to the lack of antioxidant protection mechanisms (Aitken, 2017). These different components can overall lead to reduction of male fertility, loss of sperm function and reduction in membrane integrity (Nowicka-Bauer & Nixon, 2020).

As with ROS affecting spermatozoa movement, LPO does the same through peroxidation of membrane lipids (Nowicka-Bauer & Nixon, 2020). By affecting the PUFA membrane, receptors and enzymes are affected downstream hence the loss of integrity of the spermatozoa membrane, which is also the reason cryopreservation of sperm has been well documented due to the ROS bursts which come with thawing (Nowicka-Bauer & Nixon, 2020). Ion diffusion and ion channel/pump dysregulation has been linked with LPO aldehyde products produced by the PUFA membrane in a similar mechanism to ROS attack through downregulation of these membrane components (Nowicka-Bauer & Nixon, 2020). Additionally, LPO greatly affects mitochondrial membranes which is specifically found in the mid-piece of the sperm morphology, also known as the power centre of the sperm with regards to motility (Nowicka-Bauer & Nixon, 2020). As with most cells, the mitochondria produce the energy the spermatozoa need to function and move, thus any disruption or damage here can cause major consequences for the germ cell (Nowicka-Bauer & Nixon, 2020). If the plasma membrane of the spermatozoa becomes damaged due to ROS or LPO, secondary mitochondrial LPO and ROS

are generated, once again altering the fluid membrane structure causing electron leakage and loss of mitochondrial membrane potential, reducing energy for the spermatozoa and consequently movement (Aitken, 2017; Nowicka-Bauer & Nixon, 2020).

For the sperm cell to recycle and maintain functionality in oxygen homeostasis, certain exogenous enzymes such as glutathione peroxidase (GPX) and superoxide dismutase (SOD) scavenge ROS and LPO aldehydes (Agarwal, Samanta & Henkel, 2018; Nowicka-Bauer & Nixon, 2020). Yet their supply is limited as mentioned, and thus, when the apoptotic cascade begins, they struggle to maintain detoxified conditions (Agarwal, Samanta & Henkel, 2018; Nowicka-Bauer & Nixon, 2020). These molecules trap free radicals and regulate cellular activity (Drevet & Aitken, 2020). Specifically, GPX removes H_2O_2 from the cell which is specifically related to sperm motility and therefore could potentially be a target for therapeutics (Agarwal, Samanta & Henkel, 2018). Lifestyle changes such as removal of alcohol, smoking and reduction of toxins could all be seen as potential corrective measures or targets for clinical application to reduce OS, while antioxidants such as vitamins E and C are known to eliminate ROS due to their scavenging capacities (Dutta, Majzoub & Agarwal, 2019). To investigate these therapeutic targets, the level of OS or LPO within the spermatozoa and semen needs to be evaluated.

In 1987, Aitken and Clarkson were the first people to confirm ROS was produced by human sperm using a chemiluminescent luminol probe (Agarwal, Samanta & Henkel, 2018) and since then, over 30 assays to measure OS in semen have been produced (Agarwal *et al*., 2019). These tests either indirectly measure OS through LPO values or DNA damage, directly by measuring ROS within the sperm cell or measuring the cells antioxidant capacity (Dutta, Majzoub & Agarwal, 2019). Until recently, OS testing was not routinely done due to how complex and expensive it was to test, as well as the lack of a standardised testing protocol available, even with the literature basis indicating the connection between male infertility and OS (Dutta, Majzoub & Agarwal, 2019). Yet, with the advent of new testing procedures such as the Male Infertility Oxidative System (MiOXSYS), which assesses both antioxidants and

oxidants (oxidation-reduction potential, ORP) at a cost-effective and fast rate, clinical biomarkers are finally available for physicians and andrologists alike (Agarwal *et al.,* 2019). MiOXSYS also has reduced inter- and intra-laboratory variability which is one of the main concerns for laboratories when conducting subjective semen analysis and has been proven in multiple studies to be accurate, reliable, and reproducible (Agarwal *et al.,* 2019). It has now been shown that ORP values are significantly negatively correlated with sperm parameters such as motility, concentration, total motile count, and morphology (Agarwal *et al.,* 2019). It is vital to make an accurate diagnosis of male infertility because if an OS DNA damaged sperm manages to fertilise an egg, often the egg does not develop into a healthy embryo, causing embryonic mortality and miscarriage (Fraser, 2004; Agarwal *et al.,* 2019; Hamilton & Assumpção, 2020).

1.4.4 DNA fragmentation damage

In addition to OS directly affecting sperm with ROS in adenosine and guanine bases, it has a vital role in causing DNA fragmentation damage in the form of DSB or single-strand breaks (SSB) within spermatozoa (Aitken, 2017; Drevet & Aitken, 2020). This is because of the ROS imbalance in the semen, along with SOD, block naturally occurring reduction which in turn damages the DNA strands inducing sperm DNA fragmentation (Hamilton & Assumpção, 2020). H_2O_2 and other ROS can also directly cause DNA damage by reaching the nucleus and cause chromatin protein cross-linking and abasic sites (Champroux *et al.,* 2016) (Figure 1-11). Other than OS causing DNA damage, it also occurs due to abortive apoptosis post-meiotically if there is a lack of repair mechanisms present and if there is a strand break, which occurred in spermiogenesis and was not corrected prior to spermatozoa maturation (such as topoisomerase involved in chromatin remodelling (Champroux *et al.,* 2016; Aitken, 2017; Sharma & Agarwal, 2018).

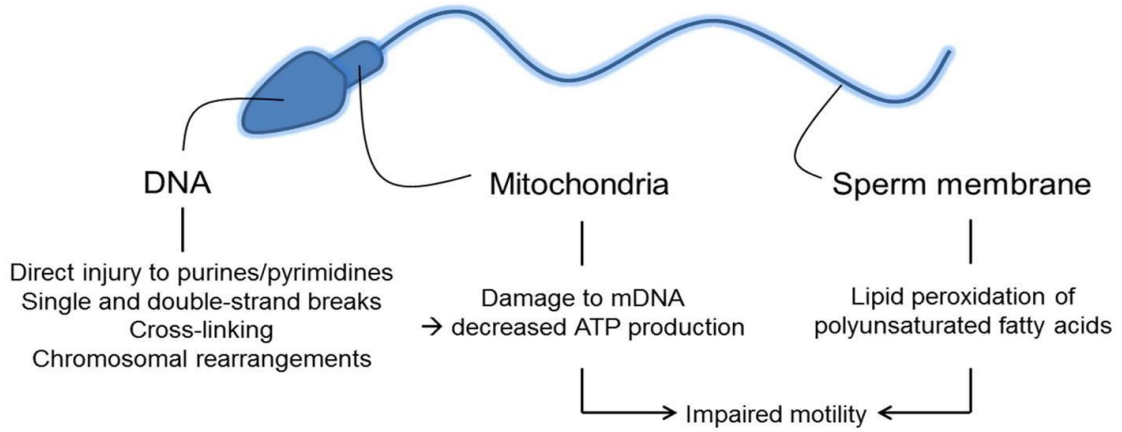


Figure 1-11 The role of oxidative stress (OS) in DNA, mitochondria, and the sperm membrane (Image source: Wagner, Cheng & Ko, 2017).

Topoisomerase II is one of the multiple DNA-based enzymes which fixes DSB, yet if it does not function properly or is overloaded, the DNA breaks can potentially remain in the spermatozoa if not fixed by other similar enzymes (Hamilton & Assumpção, 2020). This specifically takes place during chromatin remodelling within spermiogenesis due to defective protamination or pure breaks within the DNA, which not only leaves the spermatozoa prone to oxidative attack, but also vulnerable to apoptosis (Fraser, 2004; Peña, Ferrusola & Muñoz, 2016; Sharma & Agarwal, 2018). The apoptotic process within spermatogenesis, as well as spermatozoa maturation, spermiogenesis, is needed to remove abnormal germ cells, thereby controlling production of healthy sperm and maintaining the correct proportion of Sertoli cells to germ cells (Champroux *et al.,* 2016; Hamilton & Assumpção, 2020). Sertoli cells induce apoptosis in approximately half to two thirds of germ cells that begin meiosis and hence if this doesn't occur correctly, those cells can reach maturation and are ejaculated with DSB (Champroux *et al.,* 2016). These abortive apoptotic cells are related to poor semen parameters such as low concentration, abnormal morphology, high levels of DNA damage and generally low semen quality in many species including bulls, boars, rams, stallions and humans (Fraser, 2004). Furthermore, increased apoptosis has been related to age in humans with regards to DSB breaks and reduced repair mechanisms in spermatozoa (Fraser, 2004).

Due to all the different means of DNA damage and fragmentation that can occur in the spermatozoa, it is no surprise that different environmental and physiological stressors can

cause DSB, chromosome abnormalities and genetic mutations, consequently negatively affecting male fertility (Fraser, 2004). For example, DNA fragmentation levels within sperm have been found to be significantly raised in asthenozoospermic males when compared to fertile men (Nowicka-Bauer & Nixon, 2020). It has also been shown that DNA damage has been associated with conception issues and poor embryonic development due to the abnormal sperm chromatin structure related to DNA fragmentation (Fraser, 2004; Champroux *et al.,* 2016). Due to the negative relationships between male fertility, DNA damage and conception, testing assays are critical to determine the current level of damage in order to clinically treat the problem for couples to improve their hope and chances of fertilisation (Fraser, 2004; Dutta, Majzoub & Agarwal*,* 2019).

To test for male infertility and to see if the DNA damage in a cell is caused by OS or other factors, simultaneously testing for OS using a system such as MiOXSYS (section 1.4.3) in conjunction with the Sperm Chromatin Structure Assay (SCSA®) for identifying the level of DNA damage, also known as the DNA fragmentation index (DFI), could be an objective alternative to a standard semen analysis (Dutta, Majzoub & Agarwal*,* 2019). It has been demonstrated that the two parameters (ORP and DFI) are significantly positively correlated with one another (Dutta, Majzoub & Agarwal*,* 2019), and as it has been shown that MiOXSYS values are significantly correlated with other semen parameters (Agarwal *et al*., 2019), this could be an alternative worth investigating. Dutta, Majzoub & Agarwal (2019) reported DNA fragmentation was positively correlated with increased MDA levels in semen, indicating a link between LPO and DNA damage, as well as a significant positive correlation between sperm head defects, ORP, DNA fragmentation and infertile men. This substantiates the connection between DNA damage, ORP and male fertility, as a further method of testing for infertility objectively, both of which are currently not included in standard conventional semen analysis (Fraser, 2004; Dutta, Majzoub & Agarwal*,* 2019).

The most popular test for DNA damage is the SCSA® as it measures both chromatin structure abnormalities and sperm DNA fragmentation (Evenson, 2016) in not only humans, but stallions and other male mammals as well (Evenson, 2016; Peña, Ferrusola & Muñoz, 2016). The SCSA® is a useful test as it is fast, gives a whole representation of the sample due to the pure number of cells counted through the flow cytometer, is repeatable and precise and hundreds of different publications have used it as a test for validation of male fertility (Evenson, 2016). The test uses both the green and red fluorescence from the flow cytometer to give an indication of the various populations of cells, whereby red fluorescing cells show denatured single-stranded DNA (ssDNA) and green indicate whole dsDNA (Fraser, 2004; Evenson, 2016). The higher the proportion of red: green, the more damaged and decondensed the sample is with regards to DNA fragmentation and chromatin structure, respectively, indicating the DFI (Fraser, 2004). To test the fluorescence of the semen sample, acridine orange (AO) is used to induce acid denaturation *in-situ*, allowing the user to monitor the sperm chromatin susceptibility and thus integrity (Fraser, 2004), based on the DFI of the sample. The DFI has been shown to be significantly correlated with pregnancy rate *in vitro* and *in vivo*, which indicates its capability as a test of male factor fertility (Fraser, 2004). Although the assay is highly accurate, it needs experienced trained personnel in the test, expensive flow cytometry machinery and unfortunately cannot determine if the chromatin abnormalities are specific to DNA fragmentation (Fraser, 2004). Additionally, some of the literature suggests, DFI is poorly correlated with standard semen analysis parameters of semen quality such as morphology, sperm number, concentration, and motility (Fraser, 2004). From the SCSA® information, High DNA Stainability (HDS) can also be calculated (Evenson, 2016). This is the population of sperm in the semen sample with DNA lacking full protamination and because of this, has a high number of retained histones, and hence, a high staining level (Evenson, 2016). This allows the user to determine the number of sperm chromatin protein defects as a percentage (Evenson, 2016).

Another well used flow cytometric test of DNA fragmentation, is the terminal deoxynucleotidyl transferase-mediated fluorescein-dUTP nick end labelling (TUNEL) test, which was developed in 1993, specifically to test the effect ROS have on DNA integrity (Fraser, 2004; Evenson, 2016). The test labels the 3'-OH free end of DSB and SSB with FITC-avidin and 5'-bromo-2'-deoxyuridine 5'-triphosphate nucleotides (BrdUTP) (deoxyuridine triphosphate - dUTP) by using terminal deoxynucleotidyl transferase (TdT) in an enzymatic reaction (Fraser, 2004). This enzyme incorporates labelled nucleotides by polymerizing the free 3'-OH end of DNA; and using flow cytometry, or less accurately fluorescence/light microscopy, can be detected in spermatozoa (Fraser, 2004; Evenson, 2016). Any fragmented or damaged DNA (chromatin 3'-OH ends) brightly fluoresces and thus can easily be quantified and evaluated, while normal DNA is capped with telomeres and therefore, does not (Fraser, 2004). Other tests for DNA fragmentation have also been produced such as the sperm chromatin dispersion (SCD and HALO) test, COMET assay and standard AO test (Evenson, 2016).

However, aiding men with DNA damage in achieving fertilisation and a successful pregnancy has been a topic of debate for years, as lifestyle changes, natural conception, and even *in-vitro* fertilisation (IVF) are often unsuccessful (Kim, 2018). Intracytoplasmic sperm injection (ICSI) is a promising method to help achieve this goal, yet once again, the method of improving the success rate of this technique through sperm selection is often discussed due to its effect of livebirth rates (Ioannou *et al.*, 2016; Miller *et al*., 2019). Sperm are often only chosen based on their appearance, and thus are not tested for high DNA damage and aneuploidy prior to their selection for ICSI, which is detrimental to patient treatment emotionally and financially, as it is well-known that DNA damage causes recurrent miscarriage and pregnancy loss (Robinson *et al.*, 2012). Robinson *et al.* (2012) demonstrated how the clinics which used the SCSA® and specifically, TUNEL assay, for DNA fragmentation testing, had reduced rates of pregnancy loss, and that there was a significant connection between high DNA damage and miscarriage. Although patients with high DFI have an increased chance of fertilisation with ICSI than with IVF, this is highly dependent on oocyte quality (Robinson *et al.*, 2012).

In the hyaluronic acid (hyaluronan) binding sperm selection (HABselect) trial, the effect of reduced aneuploidy and DNA damage in hyaluronan-selected sperm for ICSI was investigated to see its effect on the proportion of livebirth rates (Miller *et al.,* 2019). It has been shown that HAB-selected sperm for ICSI reduce miscarriage rates, and even though there was a small non-significant (2.2%) increase in livebirth rates, assumed due to reduced miscarriage, there was no statistically significant improvement (Miller *et al.,* 2019; Miller, 2019). It is clear that although there are useful techniques which could aid in the success of ICSI treatment, there first needs to be correct diagnosis, treatment and then the prognosis of the patient and couple. Simply by testing for DNA fragmentation, a patient could potentially be treated for a simple lifestyle change, rather than being referred directly to ICSI. For example, it was shown by Schmid *et al.* (2007) that caffeine-intake was a significant factor in increasing sperm DNA damage due to DSB and it is also well-known that smoking heavily influences DNA damage. DNA fragmentation can lead to gene mutations, chromosomal rearrangements and disomy, further increasing the risk of miscarriage, developmental diseases, or congenital disorders in offspring (Schmid *et al*., 2007; Nili, Mozdarani & Pellestor, 2011; Middelkamp *et al.,* 2020). Based on this information, if a screening service was available to identify these patients with high DNA fragmentation, aneuploidy, and poor chromatin integrity in their sperm, prior to treatment selection, it would provide a relief to patients and couples alike, financially, and emotionally.

1.4.5 Sperm damage studies in stallions

Stallions are highly prized animals, especially if they are successful in their specific discipline in equestrian sports. Because of this they are normally selected for stud based on their breeding pedigree and athletic capabilities, instead of reproductive and fertility levels (Griffin *et al.,* 2019). This is similar to humans as there is a lack of pressure placed on being reproductively fit, and therefore, they have lower per-cycle conception rates at approximately 60%, when compared to other livestock (Griffin *et al*., 2019). This is also the reason they have lower per

cycle conception rates in contrast to other livestock species, especially as mares can only be detected in foal around 10 days to two weeks, after they have been covered (mated) (Griffin *et al.,* 2019). With the short breeding season experienced in horses, successful matings and inseminations are vital to ensure a mare is pregnant due to the high costs involved and hence, robust semen analysis and tests are necessary for equine breeders (Griffin *et al*., 2019). This is especially important with the advent of artificial insemination (AI) as semen is then exported internationally and if it is of poor quality with elevated DNA fragmentation or oxidative damage, there is a higher chance of reproductive failure after lengthy time and travel costs (Gibb & Aitken, 2016).

Conventional semen analysis in stallions normally investigate morphology, motility, and concentration of spermatozoa; yet poor handling and collection methods can often cause a healthy fertile stallion to look sub-fertile and thus, mislead breeders (Griffin *et al.,* 2019). Additionally, very few additional tests are conducted such as DNA fragmentation and oxidative damage assays to find the underlying cause of possible infertility or subfertility. Semen assessments are particularly vital in Thoroughbred stallions as their breeders are not allowed to use AI in the racehorse industry (Griffin *et al.,* 2019), hence further *in situ* testing could greatly benefit these stallions and the industry.

One of the main differences between stallion and human spermatozoa is the method in which stallion spermatozoa produce ATP. Equine spermatozoa use oxidative phosphorylation (OXPHOS) while humans use glycolysis (Griffin *et al.,* 2019; Peña *et al.,* 2019). The OXPHOS system causes the equine spermatozoa to have a 60% higher motility speed than human sperm (Griffin *et al.,* 2019), but due to this system, stallion sperm also has an elevated production of ROS due to the extra mitochondrial activity and high concentration of unsaturated FAs (Squires, 2005; Griffin *et al.,* 2019; Peña *et al.,* 2019). The most paradoxical difference between human and horse sperm is the increased ROS seen in stallions is highly correlated with heightened fertility characteristics and parameters (Peña *et al.,* 2019), such as

rapid motility (R^2= 0.89) and total motility (R^2 = 0.90) (Gibb, Lambourne & Aitken*,* 2014), which is the opposite relationship seen in human spermatozoa. LPO was also positively correlated with rapid motility (R^2 = 0.51) and total motility (R^2 = 0.46) (Gibb, Lambourne & Aitken*,* 2014). Although Griffin *et al.* (2019) have mentioned stallions have a rich matrix of antioxidants, an extended production of ROS does eventually lead to OS, loss of motility and DNA fragmentation, thus it seems stallion spermatozoa have increased energy metabolism particularly of ATP (Wnuk *et al.,* 2010; Gibb, Lambourne & Aitken*,* 2014; Peña *et al.*, 2019). Stallions additionally have a sophisticated structure to track redox homeostasis within their sperm using SOD and GPX and more significantly aldehyde dehydrogenase (Griffin *et al.,* 2019; Peña *et al.,* 2019). The main reason suggested for the differences in energy metabolism and increased velocity of spermatozoa, is due to evolutionary pressure regarding multiple stallions mating with different mares, whereby the fastest sperm would be the first to reach the egg and fertilise it (Griffin *et al.,* 2019). However, this is contradictory to standard equine herd dynamics, whereby a single stallion is in charge of a herd, not multiple. This analogy could lead from cases where a group of young colts finds a herd with a normally older stallion who cannot compete with the younger males, and thus they mate with the mares in the herd, pushing the primary stallion out.

The primary FA in stallion spermatozoa is docosapentaenoic acid (DPA - C22:5) which changes from breeding season to non-breeding seasons throughout the year (Griffin *et al*., 2019). In the breeding season, there is a decrease in the omega to PUFA ratio 3:6 which contributes to the higher elasticity and fluidity of the plasma membrane due to the increased PUFAs, hence aiding the egg-sperm fusion occurring for fertilisation (Griffin *et al.*, 2019). Because of this change, in conjunction with the increased OXPHOS, spermatozoa are more prone to LPO and ROS attack, and consequently OS (Squires, 2005; Griffin *et al.,* 2019). This is especially prominent in older stallions as well as in spermatozoa which have been cryopreserved, as thawing leads to significant osmotic mitochondrial damage (Squires, 2005; Peña *et al.,* 2019).

Cooling, freezing, and thawing spermatozoa which have been cryopreserved also causes DNA damage within the nucleus (Squires, 2005). As with human males, in stallions, it has been shown that fertility differences and breeding soundness between individuals, can be determined by using the SCSA® (Evenson, 2016; Peña, Ferrusola & Muñoz*,* 2016). When using the SCSA® in stallions, morphological abnormalities and the degree of chromatin breakdown have both been negatively correlated with pregnancy rates per cycle (Squires, 2005; Griffin *et al.,* 2019). Additionally, Wnuk *et al.* (2010) demonstrated a negative correlation between DNA damage and total antioxidant capacity of stallion spermatozoa, further demonstrating the need to investigate DNA damage in the equine breeding industry. DNA fragmentation is still rarely used as a proof of semen quality in the equine industry when compared to its use in the human andrology industry. With the growing use of costly procedures such as IVF, embryo transfer, and ICSI, sperm testing for OS and DNA fragmentation are vital for improved overall fertility in the industry (Griffin *et al.,* 2019).

1.4.6 Sperm DNA damage in other animals

AI is used extensively in other agricultural species such as in pigs and cattle, while it is complex in sheep due to their convoluted cervical anatomy (Falchi *et al.,* 2018). Many researchers speculate, that due to the increase in ART, the overall infertility of mammals is rising and so is the DNA damage seen in the sperm (Kumaresan *et al*., 2020). Thus, animals are not naturally selected due to their fertility status (Kumaresan *et al*., 2020). This can have adverse effects on the agricultural industry as a whole, as well as the UK economy (Kumaresan *et al*., 2020). As previously mentioned in sections 1.3.1 and 1.3.2, AI of high genetic merit animals is important, not only to the viability of agricultural practices, but also because one sire can influence the genetics of entire populations globally (Kumaresan *et al*., 2020). Hence, testing for DNA damage in sperm should be a crucial component of semen analysis and sperm viability prior to

dissemination of genetics, especially due to its connection with embryonic losses and miscarriage in most mammals.

Interestingly, ruminant sperm DNA is highly compacted and hence DNA testing in rams and bulls has previously been difficult, especially when determining whether the damage is due to oxidative stress or other factors (Boe-Hansen, Fortes & Satake, 2018; Soria-Meneses, *et al.,* 2022). In rams, Soria-Meneses *et al.* (2022) demonstrated how DNA damage in rams can however be accurately detected with the SCSA®, especially DNA damage related to OS. Additionally, in dairy bulls, DNA fragmentation was responsible for a two-fold difference between high- and below-average-fertility bulls, with the latter having increased DNA damage (Kumaresan *et al.,* 2020). Similarly in non-ruminants, such as boars, field fertility and poor embryo viability was significantly correlated with DNA damage and has had a negative effect on litter sizes in Landrace and Duroc breeds (Khezri *et al.,* 2019). However, DNA damage in boars is generally low and highly individual specific (Boe-Hansen *et al.,* 2005; Khezri *et al.,* 2019). Other factors which have shown to cause an increase in DNA damage, is the use of cryopreservation storage of boar, ram, and bull sperm, as well as environmental toxin and thermal stress effect in ram and bull sperm (Kumaresan *et al.,* 2020).

1.4.7 Sperm DNA breakage and its relationship to evolutionary change

As it has been previously shown (section 1.4.4) how DNA damage can not only cause recurrent miscarriage in individuals, but also has a major effect on chromosomal rearrangements in embryos and can cause genetic abnormalities in offspring. It is no surprise then, that the theory of DNA damage in sperm aiding in speciation and reproductive isolation of species has surfaced. It has been speculated and theorised that environmental conditions and the function the sperm must perform depending on adaptive qualities of a specific species, have caused sperm chromatin structure to change and adjust to stressful situations in order to preserve the sperm integrity (Dominguez, Arca & Ward, 2011; Gosálvez, Holt & Johnston, 2014). For

example, Dominguez, Arca & Ward (2011) suggested that some sperm have the main function of protecting the paternal genome during transit (Echidna), while others have a chromatin structure specific for proper embryogenesis (Common Planigale). Because of these adaptations, the sperm chromatin becomes vulnerable to DNA damage, causing chromosomal structural changes, altering the species entirely over time (Dominguez, Arca & Ward, 2011).

Other studies have found that a protamine deficiency in the sperm is connected to male fertility due to DNA damage, whereby abnormal protamine ratios are linked with heighted DNA fragmentation in species such as the stallion, man, mouse, and bull (Nie *et al*., 2016). In research done by Gosalvez *et al.* (2011), 11 evolutionary divergent species were investigated for DNA damage in their samples, related to their protamine levels. The higher the level of protamine-2 in a specific species, the increased chance of DNA damage found (Gosalvez *et al*., 2011). Conversely, the more stable the sperm integrity was (based on the level of cysteine residues), the more direct the evolutionary relationship was between species, for example animals in the same phylogenetic clade, based on their sperm DNA stability (Gosalvez *et al*., 2011). It is possible, that the higher the level of DNA damage in a species, the increased likelihood of the mammal to rapidly adapt to environments through the instability of the DNA. This could aid scientists in identifying species which are more likely to adapt to environmental differences we are seeing with climate change globally, as well as aid in conservation of these species.

1.5 Regulation of semen analysis

The human andrology industry aids 25% of couples within one year globally, as unexplained male infertility counts for at least 40% of all patients seen (Nowicka-Bauer & Nixon, 2020; Agarwal *et al.*, 2019). Due to how humans have evolved, there is a lack of selection pressure on reproductive qualities and the WHO has estimated approximately 190 million people fight with subfertility globally; a number which is rising (Agarwal *et al.,* 2019). Unexplained male

infertility is defined as the “presence of altered semen characteristics without an identifiable cause” and this is related to the lack of a female cause (Agarwal *et al*., 2019). Ultimately, this is due to the loss of spermatozoa function and therefore acrosome reaction, capacitation and fertilisation cannot occur (Walters *et al.*, 2018).

One parameter which has only recently become of interest, is OS and its impact on/or due to, other diseases/causes (Lanzafame *et al*., 2009). For example, Nowicka-Bauer & Nixon (2020) reported a link between cancer, obesity, heart failure and Alzheimer’s disease with OS, and of course as a biomarker in 35% of men with either varicocele, prostate cancer, or inflammatory cancer (Lanzafame *et al*., 2009; Nowicka-Bauer & Nixon, 2020). Furthermore, ROS were associated with loss of motility, viability and oligozoospermia (low sperm count) (Nowicka-Bauer & Nixon, 2020). As stated previously, varicocele has been seen in 40% of cases of all male subfertility, whereby OS and LPO are crucial acting agents within varicocele (Lanzafame *et al*., 2009; Dutta, Majzoub & Agarwal*,* 2019; Nowicka-Bauer & Nixon, 2020). This could be due to the increased scrotal temperatures seen with varicocele (Lanzafame *et al*., 2009). Other conditions which cause an increase in white blood cells, and thus amplifying the degree of OS in the testis due to inflammation, are hypoxia, ischaemia and injuries to the epididymis or other parts of the male reproductive organs (Lanzafame *et al*., 2009). These seminal leukocytes are a natural defence system against diseases or possible attacks on the testis and can therefore affect spermatogenesis and spermiogenesis by increasing ROS and potentially overwhelming the natural sperm production mechanisms (Lanzafame *et al*., 2009). 55% of oligozoospermic patients (men with a low sperm count) have been reported to have elevated ROS production in addition to reduced motility and sperm function (Aitken, Clarkson & Fishel, 1989; Lanzafame *et al*., 2009).

Evidence additionally suggests, OS causes 30-80% of unexplained male infertility (Agarwal *et al.,* 2019). Due to its connection with multiple clinical pathologies and many being the cause, OS should be incorporated as a biomarker for unexplained male infertility or for men who have

abnormal semen analysis (Agarwal *et al*., 2019) (Table 1-2; Homa *et al*., 2019). This was substantiated by Homa *et al.* (2019) where two different assessments of DNA damage and OS associated with recurrent miscarriage and poor embryo development were assessed, and both tests for each of the individual parameters explained the links between unexplained male infertility, OS and/or DNA fragmentation and miscarriages or lack of embryo development. Similarly, Agarwal *et al.* (2019) found 83.8% of infertile men in a clinical trial had high ORP values. In another study conducted by Agarwal *et al.* (2019) on 3 966 patients, high ORP was significantly associated at a p-value < 0.0001, with abnormal morphology, poor motility and low spermatozoa concentration, and a significant difference was seen between the ORP of fertile men versus sub-fertile men (p < 0.0001). Venkatesh *et al.,* (2009) also found a high correlation between morphological abnormalities and ROS, especially affecting acrosome structure and increased MDA concentrations (Dutta, Majzoub & Agarwal*,* 2019).

	ROS		**sORP**	
	Number of patients	**%**	**Number of patients**	**%**
Normozoospermia ***Normal semen sample***	172	34.6	139	46.2
Oligozoospermia ***Low sperm count***	18	3.6	9	3.0
Asthenozoospermia *Poor* ***sperm motility***	8	1.6	6	2.0
Teratozoospermia *Abnormal sperm morphology*	119	24.0	72	23.9
Oligoasthenozoospermia *Poor sperm count and motility*	1	<1	1	<1
Oligoteratozoospermia ***Poor sperm count and morphology***	69	13.9	32	10.6
Asthenoteratozoospermia *Poor sperm motility and morphology*	29	5.8	11	3.7
Oligoasthenoteratozoospermia *Poor sperm count, motility, and morphology*	49	9.9	15	5.0
Leukocytospermia *High number of white blood cells in semen*	31	6.3	16	5.3
TOTAL	496	-	301	-

Table 1-2 The distribution of semen samples classified according to World health Organization (WHO, 2010) criteria. Reactive oxygen species (ROS) and static oxidation reduction potential (sORP) values are labelled at the top, with the number of patients in each WHO category and their mean ROS or sORP value (%), respectively. (Adapted from Homa *et al.,* 2019).

As these previous studies are indicators of male fertility, the ORP resulting from using the MiOXSYS assay could be paramount for future prediction of male infertility, especially due to it being cost effective, specific, sensitive, and practical (Dutta, Majzoub & Agarwal, 2019; Homa *et al.,* 2019).

Most human LPO assays study 4HNE. 4HNE destabilises proteins within spermatozoa thereby causing an egg-sperm recognition block and accordingly are detrimental to not only fertilisation, but also spermatozoa survival and function (Walters *et al.,* 2018). As with most mammalian male species, LPO products such as 4HNE cause issues with acrosome reaction, reduced motility, and anomalies in morphology (Walters *et al.,* 2018). H_2O_2 is the primary by-product of 4HNE and SOD and GPX are crucial in protecting the sperm by maintaining homeostasis between ROS and H_2O_2, yet they are only present within the midpiece cytoplasm, where the mitochondria are found (Lanzafame *et al*., 2009). When SOD and GPX were in low concentration, a higher ROS production was found in 88.8% of patient seminal plasma and this was directly associated with infertile patients when compared to normozoospermic patients (Lanzafame *et al*., 2009). Yet when SOD was in high concentrations, poor sperm parameters were reported due to a high concentration of H_2O_2 and reducing ROS so low, it was disrupting the normal ROS functioning within the sperm (Lanzafame *et al*., 2009).

As stated in section 1.4.3, if ROS overwhelms the cell, it can cause DNA damage resulting in diseases, embryo mortality and genetic abnormalities (Homa *et al.,* 2019). Irrespective of the method used to determine OS, the connection between DNA fragmentation and OS has been well studied and tested (Homa *et al.,* 2019). High DNA damage and DFI have been found predominantly in patients with varicocele, oligozoospermia and leukocytospermia and in connection with recurrent spontaneous abortions (Agarwal *et al.,* 2016a). The most significant causative relationship was found in patients with high spermatozoa DNA damage specifically who had leukocytospermia (41.7% ± 17.6%), and secondly patients with varicocele (35.7% ± 18.3%), both conditions which are associated with male infertility or sub-fertility (Agarwal *et*

al., 2016a). Based on this information, it is imperative for clinicians to include ROS, ORP and DNA damage screening into routine semen analysis for male mammals in general, to ensure a correct pathology of idiopathic male fertility can be described and treated.

1.6 Specific Aims of this thesis

The study of gross genetic changes and their association with fertility and reproductive isolation a is a large significant field that has many areas with room for improvement and further study. Comparative genomics allows the exploration of highly conserved areas of the genome within closely and distantly related mammals can always be expanded and diversified in by using BAC clones with a view to them becoming tools in a range of species. These BAC clones can consequently be used to not only investigate the accuracy of bioinformatic sequencing platforms within these mammals, but also to test their precision by physically mapping their locations and cross referencing them to the genomic sequencing databases. Genes of interest can be mapped with FISH and can also possibly highlight patterns, orthologues, and conservation throughout phylogenetic relationships. By using the BAC clones specific to a particular species, fertility information related to cryptic chromosomal abnormalities can aid clinicians and geneticists understand the background behind not only phenotypic abnormalities, but also help in future decision making for potential offspring.

When specifically looking at sperm and semen analysis, another aspect of fertility investigation, developing new, simple, and cost-effective techniques to study unexplained male subfertility and infertility in humans, horses and pigs is imperative, as many of these tests are not viable as industrial assays. By using new available technologies such as flow cytometry, previously underrated tests such as for LPO could be improved on and designed for use in clinics and breeding centres when compared to a standard semen analysis. In addition to novel techniques, understanding whether fertility and andrology centres, laboratories and clinics are currently conducting semen analysis to standard is also an important aspect of investigation

which hasn't been looked into within the fertility industry. Because of these reasons, the specific aims of this thesis are:

Specific aim 1. To assess a panel of 48 sequence- and conservation score-based homologous cattle BAC probes in both phylogenetically similar and more distantly related mammals for cross species hybridization patterns. This is partly to observe evolutionary differences between species and partly to assess the potential for these cattle probes to be used to assess chromosome abnormalities affecting fertility in other species.

Specific aim 2. To develop a novel approach for accurate detection of chromosomal rearrangements affecting equine fertility.

Specific aim 3. To develop a novel flow cytometric assay for membrane lipid peroxidation (LPO) in human and equine sperm and to the hypothesis that there is a correlation between that, and DNA damage (and, by extension, fertility) in stallions and men.

Specific aim 4. To determine whether there are any differences in semen analyses between laboratories and clinics based on the regulatory body with which they are registered, as well as what type of entity they are.

2. Methods and Materials

2.1 Chromosome preparation and karyotyping slide set up

In order to address both specific aims 1 (section 3) and 2 (section 4) chromosomes needed to be karyotyped. This section describes the methods used to achieve this.

2.1.1 Chromosome harvesting/Blood culture and DNA extraction

Equine and bovine blood was received in lithium heparin tubes. 0.5ml of blood samples were placed into 9.5ml of 37℃ prewarmed Gibco™ PBMax™ Karyotyping medium (ThermoFisher Scientific, Waltham, United States) in T25 flasks upright and incubated for 3 days for cattle and 4 days for horses.

37℃ prewarmed KaryoMAX™ Colcemid™ Solution in phosphate buffered saline (PBS) (ThermoFisher Scientific, Waltham, United States) was added to the T25 flasks and left in the incubator for 30 minutes. This is done in order to collect the cells in the metaphase stage, preventing sister chromatids separating in anaphase (Keagle & Gersen, 2013). The cell suspensions were then moved into 15mL falcon tubes, centrifuged at 1,900 revolutions per minute (RPM) for 10 minutes prior to removal of the supernatant, and then the cell pellet was resuspended. 37℃ prewarmed 75mM potassium chloride (KCl) was then added to the cell suspension in a dropwise manner while gentle agitation was achieved for a maximum of 12 minutes. 3:1 100% methanol to 100% acetic acid was then added dropwise with gentle agitation until a total of 14mL is achieved. As a hypotonic solution, KCl causes the water to move through osmosis into the cell as the cell cytoplasm has a higher salt concentration, thereby causing the cells to swell (Keagle & Gersen, 2013). The mixture was then gently inverted and spun down at 1,900 RPM for 10 minutes. The supernatant was then removed, and pellet resuspended before being drawn into a Pasteur pipette. 5mL chilled fixative (3:1 100% methanol to 100% acetic acid) was added to the tube before the suspension was gently released into the fixative to halt the hypotonic solution's action. The sample was once again

centrifuged at 1,900 RPM for 10 minutes and the process of fixation and cleaning repeated another two times. The fixative additionally lyses any other red blood cells in the sample (Keagle & Gersen, 2013). Once this was completed, the sample was resuspended in 5mL fixative and stored at -20℃.

2.1.2 Karyotype slide preparation

Blood culture- or fibroblast-based chromosome preparations were centrifuged for 10 minutes at 1,900 RPM and the supernatant was then discarded. 0.5mL fixative for resuspension of the pellet was added. 10μL of the chromosome preparation was then pipetted on a Superfrost slide before addition of 10μL of fixative and left to dry. The slides were then placed through a room temperature 70% acetic acid wash for 7 seconds to remove debris and left to dry. DAPI VECTASHIELD antifade medium (Vector Laboratories, Newark, United States) was added to 22x50mm coverslips which were inverted onto the slides and covered to develop for 10 minutes in the dark. The DAPI stains the chromosomes so they can be visualised under a fluorescent microscope. The slides were stored in the fridge until use.

2.2 Production of labelled fluorescent probes

This section is relevant for both specific aims 1 (section 3) and 2 (section 4). Fluorescently labelled probes were the backbone of these sections' focus.

2.2.1 Selection of equine BAC probes

128 equine BACs (Table 2-1, page 99) were selected from the CHORI-241 clone placement library on EquCab3.0 (GCF_002863925.1) CloneDB Release ID 102. Horse BACs were selected based on concordance, unique positioning, 130-190 Kb length and end-sequence/subtelomeric placing. Four subtelomeric BACs were selected for each chromosome, two on the p-arms and two on the q-arms for the metacentric chromosomes (1-13, and X), and proximal (p) or distal

(d) to the centromere for the acrocentric chromosomes (14-31). The BACs were received from Professor Terje Raudsepp at Texas A&M, United States, in LB agar stab form.

2.2.2 LB broth preparation, DNA purification and further processing of Equine BACs

To prepare a culture medium for the equine BACs, a 2YT LB broth was made by adding 8g Tryptone, 5g Yeast extract and 2.5g sodium chloride (NaCl) to 500mL deionized water (ddH_2O). The LB broth was then autoclaved to sterilise it and once the solution had cooled sufficiently, 125μL of the 100x stock solution Chloramphenicol (Sigma-Aldrich, St. Louis, United States) was added to the LB broth for a final Chloramphenicol concentration of 25 μg/mL. 10μL of the agar stabs were removed and placed into sterile 4.5mL polystyrene tubes containing 3.5mL of the LB broth under sterile conditions. The test tubes were cultured overnight at 37℃ with loosened two-stopper vented caps to allow the bacteria to grow under aerobic conditions.

2mL of overnight liquid cultures were pipetted into Eppendorf tubes and spun at 8000 RPM for 3 minutes. The supernatant was then removed from the pellet and a QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) used to purify the DNA.

The purified DNA was then stored at 4℃ and shipped to CytoCell Limited (Oxford Gene Technology, United Kingdom) in labelled and Biofilm sealed Eppendorf tubes for amplification, fluorescent labelling, probe purification and quality check in a proprietary method of processing. Once completed, they were sent back to the University of Kent, ready to be used.

Glycerol stock solution was made by adding 250mL absolute glycerol to 250mL LB broth and then autoclaved. When the solution was sufficiently cooled, 125µL Chloramphenicol 100x stock solution (Sigma-Aldrich, St. Louis, United States) was added to the solution, and 750μL of the final solution was distributed into Eppendorf tubes. 750μL of overnight cultures was then pipetted into the Cryovial tubes in a 1:1 ratio of glycerol to bacterial culture and pipette mixed.

This was to make a stock of the individual bacterial cultures for future use. These were then stored at -80°C.

Chr	p probe	Insert size (bp)	Genomic location		q/d probe	Insert size (bp)	Genomic location
1	CH241-155D20	164 003	Chr1:654,145–818,147		CH241-444C19	159 000	Chr1:187,838,620–187,997,619
	CH231-218A20	185 844	Chr1:29,528–215,371		CH241-198L13	163 746	Chr1:187,956,836–188,120,581
2	CH241-60K15	153 174	Chr2:575,486-728,659		CH241-38N11	163 086	Chr2:120,928,785–121,091,870
	CH241-135D13	184 396	Chr2:409,265–593,660		CH241-171D17	169 865	Chr2:121,130,726–121,300,590
3	CH241-140E5	154 392	Chr3:184,065–338,456		CH241-157L23	164 503	Chr3:120,689,928–120,854,430
	CH241-16O19	171 258	Chr3:419,963-591,220		CH241-268H12	183 740	Chr3:120,353,669–120,537,408
4	CH241-226H13	144 962	Chr4:14,187–159,148		CH241-238E5	155 496	Chr4:109,131,116–109,286,611
	CH241-58P11	149 412	Chr4:211,230–360,641		CH241-69N15	162 937	Chr4:108,912,259–109,075,195
5	CH241-326B6	182 273	Chr5:258,001–440,273		CH241-278A14	156 244	Chr5:96,493,518–96,649,761
	CH241-284D9	189 313	Chr5:68,679–257,991		CH241-205O16	159 131	Chr5:96,334,839–96,493,969
6	CH241-243H2	146 863	Chr6:243,304–390,166		CH241-238O1	171 113	Chr6:86,833,203–87,004,315
	CH241-118P7	171 219	Chr6:354,864–526,082		CH241-182N17	176 766	Chr6:86,955,428–87,132,193
7	CH241-445A22	156 618	Chr7:371,011–527,628		CH241-194A19	159 824	Chr7:100,044,786–100,204,609
	CH241-73B22	164 557	Chr7:90,546–255,102		CH241-466E8	161 094	Chr7:100,498,430–100,659,523
8	CH241-138N6	173 137	Chr8:234,006–407,142		CH241-207B3	166 136	Chr8:97,186,426–97,352,561
	CH241-150C15	188 780	Chr8:509,178–697,957		CH241-447E16	162 675	Chr8:97,346,743–97,509,417
9	CH241-78C23	149 085	Chr9:92,040–241,124		CH241-147D16	170 151	Chr9:85,229,500–85,399,650
	CH241-238A5	161 072	Chr9:190,197–351,268		CH241-214B6	185 600	Chr9:84,973,475-85,159,074
10	CH241-292C11	149 915	Chr10:174,306-324,220		CH241-470F5	187 503	Chr10:84,578,941–84,766,443
	CH241-213H3	174 755	Chr10:55,442–230,196		CH241-452L20	177 902	Chr10:84,555,285–84,733,186
11	CH241-89I23	140 186	Chr11:233,609-373,794		CH241-162E1	157 577	Chr11:60,946,962–61,104,538
	CH241-243E9	164 494	Chr11:31,311–195,804		CH241-346F11	188 857	Chr11:61,388,796–61,577,652
12	CH241-338I24	155 734	Chr12:312,216-467,949		CH241-441F18	155 340	Chr12:36,346,681–36,502,020
	CH241-332I23	182 473	Chr12:29,136–211,608		CH241-189J7	172 729	Chr12:36,765,219–36,937,947
13	CH241-462D15	171 200	Chr13:278,167–449,366		CH241-463O14	167 736	Chr13:43,492,133–43,659,868
	CH241-205B14	163 669	Chr13:205,193–368,861		CH241-7M5	191 068	Chr13:43,549,762–43,740,829
14	CH241-226M19	148 070	Chr14:82,577–230,646		CH241-10G24	176 382	Chr14:94,162,576–94,338,957
	CH241-118J1	154 010	Chr14:336,403–490,412		CH241-145C12	189 708	Chr14:94,363,809–94,553,516
15	CH241-75N19	170 949	Chr15:727,971–898,919		CH241-206P14	153 132	Chr15:92,641,601–92,794,732
	CH241-224I12	170 458	Chr15:369,503–539,960		CH241-52I17	158 318	Chr15:92,345,802–92,504,119
16	CH241-326J20	179 335	Chr16:256,755–436,089		CH241-462M22	176 465	Chr16:88,201,709–88,378,173
	CH241-222K9	187 579	Chr16:290,003–477,581		CH241-385L23	184 177	Chr16:88,601,651–88,785,827
17	CH241-464K16	146 335	Chr17:436,094–582,428		CH241-163E4	154 238	Chr17:80,430,742–80,584,979

	CH241-360G6	164 378	Chr17:582,512–746,889		CH241-404G13	170 668	Chr17:80,550,284–80,720,951
18	CH241-293H3	182 602	Chr18:209,621–392,222		CH241-187P4	142 654	Chr18:82,259,388–82,402,041
	CH241-157L16	172 859	Chr18:650,324–823,182		CH241-95G13	171 587	Chr18:82,118,703–82,290,289
19	CH241-457F24	183 877	Chr19:189,534–373,410		CH241-452B11	147 890	Chr19:62,464,488–62,612,377
	CH241-70D6	181 538	Chr19:60,302–241,839		CH241-82J19	157 786	Chr19:61,917,220–62,075,005
20	CH241-470E7	155 477	Chr20:625,871–781,347		CH241-219C9	178 215	Chr20:65,088,761–65,266,975
	CH241-71D19	168 910	Chr20:261,062–429,971		CH241-449F5	166 201	Chr20:64,943,532–65,109,732
21	CH241-117G15	150 821	Chr21:670,538–821,358		CH241-46M8	181 797	Chr21:58,308,787–58,490,583
	CH241-464H3	158 770	Chr21:498,471–657,240		CH241-466D12	171 597	Chr21:58,143,106–58,314,702
22	CH241-186N17	156 930	Chr22:233,885–390,814		CH241-231A16	187 111	Chr22:50,721,525–50,908,635
	CH241-6L7	173 849	Chr22:511,136–684,984		CH241-158E14	183 335	Chr22:50,447,462–50,630,769
23	CH241-46E18	189 855	Chr23:536,075–725,929		CH241-90I9	143 026	Chr23:55,351,969–55,494,994
	CH241-27L9	172 604	Chr23:206,865–379,468		CH241-343M4	162 688	Chr23:55,120,023–55,282,710
24	CH241-46J10	137 559	Chr24:602,432–739,990		CH241-50M21	160 052	Chr24:47,269,611–47,429,662
	CH241-228K14	169 553	Chr24:53,600–223,152		CH241-464J16	172 135	Chr24:47,429,711–47,601,845
25	CH241-200A1	170 559	Chr25:362,861–533,419		CH241-464F20	180 515	Chr25:39,666,086–39,846,600
	CH241-195H18	168 654	Chr25:490,886–659,539		CH241-188F8	131 012	Chr25:39,817,752–39,948,763
26	CH241-254K24	146 049	Chr26:1,129,047–1,275,095		CH241-265D2	169 415	Chr26:42,560,224–42,729,638
	CH241-281A21	148 111	Chr26:1,297,473–1,445,583		CH241-224I24	170 248	Chr26:42,853,681–43,023,928
27	CH241-206A7	172 638	Chr27:600,979–773,616		CH241-222E18	161 911	Chr27:40,024,494–40,186,404
	CH241-162E13	186 721	Chr27:890,453–1,077,173		CH241-92B21	170 057	Chr27:39,793,149–39,963,205
28	CH241-223A13	174 434	Chr28:829,432–1,003,865		CH241-271I18	152 583	Chr28:47,042,872–47,195,454
	CH241-294K9	177 655	Chr28:1,003,907–1,181,561		CH241-244K4	178 176	Chr28:46,783,884–46,962,059
29	CH241-219A13	175 403	Chr29:278,507–453,909		CH241-292C13	136 870	Chr29:34,356,965–34,493,834
	CH241-56C22	162 816	Chr29:560,678–723,493		CH241-136O18	176 639	Chr29:34,579,399–34,756,037
30	CH241-17B6	157 295	Chr30:232,691–389,985		CH241-294F5	186 935	Chr30:31,092,115–31,279,049
	CH241-411P18	170 841	Chr30:416,062–586,902		CH241-161P14	182 175	Chr30:30,762,630–30,944,804
31	CH241-381J7	173 697	Chr31:116,137–289,833		CH241-336J2	166 273	Chr31:25,548,264–25,714,536
	CH241-24L6	181 435	Chr31:308,786–490,220		CH241-170N2	174 892	Chr31:25,069,347–25,244,238
X	CH241-469J20	157 878	ChrX:978,025–1,135,902		CH241-20G11	174 840	ChrX:127,485,640–127,660,479
	CH241-159K1	161 070	ChrX:267,646–428,715		CH241-457E2	174 227	ChrX:126,713,938–126,888,164

Table 2-1 CHORI-241 clone placement library found on EquCab3.0 (GCF_002863925.1) CloneDB Release ID 102. The chromosome, probe name and location (p- or q-arm; proximal or distal: p/q/d) and insert size (bp)of each BAC are shown. Probes for equine chromosomes 1-13 and X are labelled based on their arms (p-/q-arm) and probes for equine chromosomes 14-31 are labelled based on their relative distance to the centromere (proximal: p/distal: d). Chr: Chromosome.

2.2.3 Selection of bovine BAC probes

50 cattle BACs (Table 2-2) were selected based on in silico genomic sequence analysis and conservation score by Dr Denis Larkin from The Royal Veterinary College (RVC). Conservation score (CS) was obtained by the PhastCons (Phylogenetic, Analysis, with Space/Time models) program used for detecting conserved evolutionary components based on a multiple alignment score. This score is based on a phylogenetic relationship between several aligned evolutionarily conserved elements, thereby scoring a sequence from 0 to 1 on a conservation scale. These analyses were performed by Dr Larkin and his group at RVC. The BACs with the highest CS were obtained from the CHORI-240 clone placement library. Repetitive content was analysed in these BACs, but it was not the determining factor for algorithmic selection for conservation score. These BACs had been received in Luria-Bertani (LB) agar stab form.

Chr	BAC ID	p/d	Insert size (bp)	Genomic location
1	CH240-475L23	p	137 702	Chr1:85,193,736–85,331,438
	CH240-377G11	d	136 376	Chr1:115,300,273–115,436,649
2	CH240-420D19	p	146 453	Chr2:52,169,035–52,315,488
	CH240-244I9	d	122 310	Chr2:18,261,827–18,384,137
	CH240-386C22	p	133 985	Chr2:34,919,377–35,053,362
	CH240-196L19	d	186 825	Chr2:87,812,076–87,998,901
	CH240-514B6	p	Not found	Not found
3	CH240-465O11	d	148 530	Chr3:96,042,026–96,190,556
	CH240-474H7	p	162 274	Chr3:56,673,448–56,835,722
	CH240-288K11	d	162 832	Chr3:84,418,392–84,581,224
	CH240-297K13	p	159 996	Chr3:33,084,551–33,244,547
	CH240-379P12	d	126 325	Chr3:45,119,285–45,245,610
4	CH240-60H16	p	143 803	Chr4:27,468,431–27,612,234
5	CH240-339P15	d	137 864	Chr5:103,700,624–103,838,488
6	CH240-124I9	p	126 112	Chr6:14,449,210–14,575,322
8	CH240-88P10	d	211 245	Chr8:27,537,709–27,784,954

	CH240-18F3	p	116 048	Chr8:55,547,292–55,663,340
	CH240-182G15	d	173 456	Chr8:96,617,426–96,790,882
9	CH240-412N22			
	CH240-341J24	p	131 074	Chr9:38,991,185–39,122,259
11	CH240-256G3	d	171 982	Chr11:27,059,107–27,231,089
	CH240-258M12	p	151 073	Chr11:61,348,283–61,499,356
	CH240-288F24	d	166 612	Chr11:93,985,525–94,152,137
12	CH240-329H2	p	155 205	Chr12:40,505,886–40,661,091
14	CH240-402O18	p	142 578	Chr14:29,258,032–29,400,610
	CH240-396P6	d	170 394	Chr14:58,828,004–58,998,398
16	CH240-208E15	d	Not found	Not found
17	CH240-26E21	p	193 383	Chr17:11,493,983–11,687,366
18	CH240-339M3	d	150 042	Chr18:19,451,839–19,601,881
19	CH240-333I1	p	170 993	Chr19:64,340,477–64,511,470
	CH240-171A7	d	Not found	Not found
	CH240-97L3	p	136 254	Chr19:48,976,202–49,112,456
21	CH240-380F23	p	167 831	Chr21:44,039,468–44,207,299
	CH240-344K23	d	131 112	Chr21:10,516,367–10,647,479
22	CH240-124B16	d	121 205	Chr22:38,775,745–38,896,950
23	CH240-310I12	p	Not found	Not found
24	CH240-33I13	d	168 248	Chr24:31,457,849–31,626,097
	CH240-305N4	p	136 079	Chr24:54,580,003–54,716,082
25	CH240-451P4	d	164 910	Chr25:30,020,596–30,185,506
26	CH240-368N15	d	159 290	Chr26:21,514,385–21,673,675
	CH240-244D2	p	181 120	Chr26:42,921,450–43,102,570
	CH240-224G7	p	151 920	Chr26:32,975,712–33,127,632
27	CH240-457O14	d	147 460	Chr27:12,923,864–13,071,324
28	CH240-394O23			
	CH240-236P3	p	151 121	Chr28:29,583,582–29,734,703
29	CH240-226K16	d	160 480	Chr29:44,624,003–44,784,483
X	CH240-128C9	p	157 332	ChrX:8,620,597–8,777,929
	CH240-29N7	d	191 229	ChrX:17,646,083–17,837,312

	CH240-359O3	p	123 986	ChrX:121,586,116–121,710,102
	CH240-48F6	d	158 819	ChrX:74,820,800–74,979,619

Table 2-2 CHORI-240 BACs selected for conservation score and checked for insert size (bp), genomic location and bioinformatic positioning on the National Center for Biotechnology Information (NCBI). BACs highlighted in red were not tested as they failed during culture. BACs labelled as "not found" could not be identified on the NCBI yet were still tested and included in the study. Each probe was labelled as either proximal (p) or distal (d) to the centromere.

All 50 probes were searched for on National Centre for Biotechnology Information (NCBI) and checked for length (Kb), range (Kb) and bioinformatic positioning (Table 2-2). None of the BACs were sequenced prior to isolation. To identify the BACs on the NCBI, the accession numbers of the BACs were found in the archives function. This was followed by using the BLAST function to identify temporary RID's for each probe prior to visualising each probe using the NCBI genome browser on the *Bos taurus* ARS-UCD1.3 assembly.

2.2.4 LB agar preparation and plating and purification of Bovine DNA

To prepare agar plates for the bovine BACs to culture, 16g of LB agar powder (Invitrogen, Waltham, United States) was added to 500mL of ddH_2O. This was then autoclaved, and 300μL Chloramphenicol 100x stock solution (Sigma-Aldrich, St. Louis, United States) was added to the solution once it had sufficiently cooled (final concentration of 60µg/mL). This solution was then poured into agar plates and left to set. For each individual bovine BAC, a sterilised pipette tip was used to remove the bacteria from the agar stab, and then streaked onto the LB agar plate. The plates were then placed in a 37°C hotbox to culture overnight.

2mL of autoclaved and cooled 1x PBS was used to wash the bacteria from the agar plates and then the solutions were transferred into Eppendorf tubes. It was assumed all bacterial colonies per clone per plate were the same and no mutations were present. A QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) was then used to purify the colonies by following their standard protocol.

2.2.5 Probe DNA amplification

A NanoDrop™ 1000 spectrophotometer (ThermoFisher Scientific, Waltham, United States) was used to analyse the 260/280 purity ratio and DNA concentration of each DNA sample. When analysing the purity ratio, values of ~1.8 for DNA and ~2.0 for RNA were considered to be "pure", respectively. It was important to check these values prior to amplification to ensure there was enough DNA in the sample. Amplification of the individual DNA samples was conducted using an altered protocol, yet with the reagents from the GenomiPhiV2 DNA Amplification Kit (GE Healthcare, Chicago, United States). 27μL of sample buffer was mixed with 3μL purified BAC DNA in 0.5mL Eppendorf tubes which were then incubated in a thermocycler at 95℃ for 3 minutes before directly being placed on ice. 30μL enzyme-reaction buffer was added to each sample solution. To figure out the volume of Phi29 DNA polymerase enzyme needed for the protocol, a ratio of 3μL (DNA sample) x 1.2 x number of tubes was calculated. The buffer was calculated by 9 x the enzyme volume. These samples were incubated in a thermocycler at 30℃ for 90 minutes, and the enzyme was then inactivated for 10 minutes at 65℃. These samples are then placed on ice.

12μL sodium acetate/ Ethylenediaminetetraacetic acid (EDTA) buffer (3M sodium acetate; 0.5M EDTA; both at pH 8.00) and 60μL ddH_2O was added to the sample. 300μL absolute ethanol was then added and centrifuged at 11,000 RPM for 15 minutes before the supernatant was removed. 500μL 70% ethanol was added and then followed by 2 minutes of centrifugation at 11,000 RPM. Once again, the supernatant was removed, and residual ethanol was then left to evaporate at room temperature. 60μL 10mM Tris-hydrochloric acid (HCl) buffer (pH 8.00) was then added to the pellet to resuspend it and then stored at 4℃.

2.2.6 Nick translation labelling

The DNA samples were analysed using a NanoDrop™ 1000 spectrophotometer (ThermoFisher Scientific, Waltham, United States) for appropriate 260/280 purity ratio and DNA concentration. To reach a final concentration of 166.5μg/μL, 10mM Tris-HCl buffer (pH 8) was added to the samples to dilute them. 12μL of the diluted BAC DNA, 8μL Nucleotide Mix A (CytoCell Ltd., Oxford Gene Technology, United Kingdom), 49.5μL MBG H_2O, 10μL 0.01M Dithiothreitol (DTT), 5μL DNase I (0.01 U/mL), 4μL DNA polymerase I (10U/mL) and 10μL 0.01M nick translation buffer (CytoCell Ltd., Oxford Gene Technology, United Kingdom) were added to an Eppendorf tube to a final volume of 100μL. Depending on the probe, either 1.5μL Texas Red-12-dUTP (Invitrogen, Waltham, United States) or 1.5μL Fluorescein-12-UTP (FITC) (Roche, Basel, Switzerland) was added to the mixture. These were pulsed to mix, and then followed by a 15℃-incubation period of 2 hours in the thermocycler. A 65℃-heat inactivation step was used to deactivate the DNase enzyme before the mixtures were placed on ice.

To test the digestion length of each DNA sample (<500 bp), an agarose gel of 1.5% was made. 100 ml 1xTris-borate-EDTA (TBE) (Sigma-Aldrich, St. Louis, United States) was added to 1.5g Agarose (Bio-Rad, Hercules, United States), after which, 2μL SYBR Safe (Invitrogen, Waltham, United States) was added post cooling. Once set, 2μL 100bp ladder (Promega, Madison, United States) and individual 2μL BAC DNA samples were all mixed with 2μL 6x loading buffer (Promega, Madison, United States) prior to loading in the gel. The gel was run at 90 V/58mA for 25 minutes.

2.2.7 Probe purification

To clean the probes from any excess fluorescent markers which were not attached during nick translation and purify the probes, a QIAquick Nucleotide Removal Kit (Qiagen, Hilden, Germany) was used. 80µL of buffer PNI was added to the PCR tube containing the probe,

gently mixed, and transferred to a 2mL Eppendorf tube, before addition of 900µL of buffer PNI. This was done as the volume of the PCR tubes was too small for the volume of buffer needed for the cleaning process. 700µL of the probe and buffer PNI solution was then transferred to the quick-spin columns provided by the kit and spun at 6 000 RPM for one minute in the tabletop centrifuge. The flow-through was then discarded. The remaining 300µL was then added to the column and again, spun for 1 minute at 6 000 RPM, and the flow-through discarded. To wash the column, 750µL PE buffer was added to the column, and spun for one minute at 6 000 RPM, once again discarding the flow-through. The column was then further spun for another minute at 13 000 RPM for one minute to ensure all remaining PE buffer was removed. The columns were moved to a new 1.5mL centrifuge tube and 100µL MBG H_2O added. This was left to stand at room temperature for 5 minutes prior to spinning the columns a final time at 13 000 RPM for one minute. Following centrifugation, the columns were discarded, and the purified probes were stored at 4°C. No quality control procedures, such as probe sequencing, were conducted post-purification due to lack of funding. The probes were deemed to have worked if they hybridized to the chromosomes (section 2.3).

2.3 Fluorescence *in situ* hybridisation (FISH)

2.3.1 Bovine probe preparation

Probe mixtures were made by adding a combination of FITC and Texas Red labelled probes. 5.5μL Hybridisation solution I (CytoCell Ltd., Oxford Gene Technology, United Kingdom), 1.5μL bovine hybloc (Applied Genetics Laboratories, Melbourne, United States) to block repetitive content, 1.5μL Texas Red labelled probes and 1.5μL FITC labelled probes were combined to make a dual colour FISH master mix.

2.3.2 Dual colour FISH hybridisation

Blood culture or fibroblast-based chromosome preparations from various species depending on the experiment, were centrifuged for 10 minutes at 1,900 RPM and the supernatant was then discarded before addition of 0.5mL fixative for resuspension of the pellet. 10μL of the chromosome preparation was then pipetted onto Superfrost slides before addition of 10μL of fixative and left to dry. The slides were placed through a room temperature dehydration series, two minutes in each solution, of 2xSSC (ThermoFisher Scientific, Waltham, United States), 70% ethanol, 85% ethanol and lastly 100% ethanol before being left to dry.

22x22mm coverslips had 10μL probe master mixes pipetted onto them prior to their inversion onto the slides containing chromosome preparations. They were then sealed with Fixogum (rubber cement) prior to being heated on a Hybrite hotplate for 5 minutes at 37℃. For denaturation of the template DNA, the slides were placed for 5 minutes at 75℃ on a second hotplate, before being placed directly into a 37℃ humidified chamber for 24 hours for same-species hybridisation. For cross-species application, the 75℃-denaturation step is only done for 2 minutes, and 37℃ hybridisation is extended to 72 hours.

Either 24 hours (same species FISH) or 72 hours (cross species FISH) post-incubation (second day FISH), the rubber cement is removed from the slides before being washed in 72℃ 0.4xSSC for 2 minutes (heated wash only completed for same-species experiments). Slides were immediately placed in a 2xSSC + 0.05% Tween-20 solution for 30 seconds at room temperature. DAPI VECTASHIELD antifade medium (Vector Laboratories, Newark, United States) was added to 22x50mm coverslips which were inverted onto the slides and covered to develop for 10 minutes (Figure 2-1). The slides were stored in the fridge until use. The process of slide preparation per experiment was repeated a minimum of three times to validate the results.

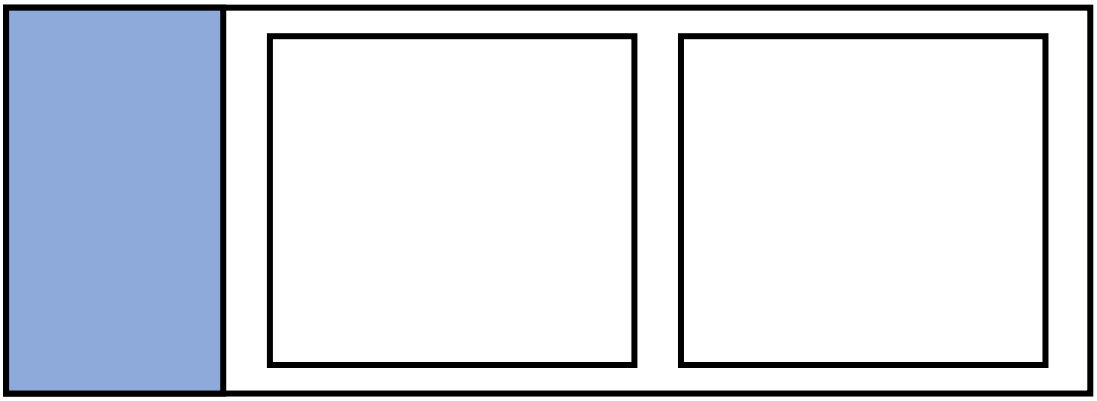


Figure 2-1 A standard microscope slide layout, including two cover slips indicated by the two squares.

2.3.3 Bovine BAC clone FISH experiments

Initial FISH testing involved establishing whether the 50 bovine BACs hybridised to same-species bovine chromosome preparations. Only 48 probes were confirmed to work when tested on same-species bovine chromosome preparations, therefore these BACs were used for further experiments. The 48 probes were then tested on 11 different species of animals following cross species FISH hybridization. Four of these species, sheep, bison, pig and

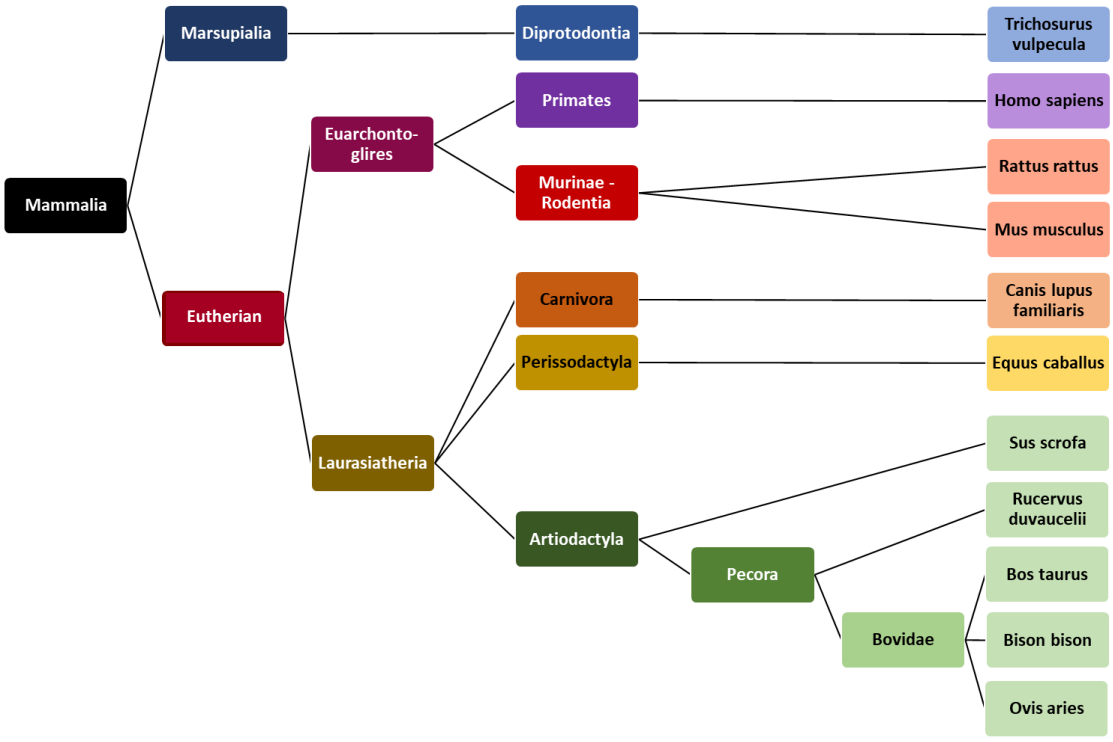


Figure 2-2 The phylogenetic tree of the 11 species tested (far right boxes) in specific aim 1 (section 3) with the 48 cattle BACs.

Rucervus duvaucelii (RDU - barasingha) are all within the group of Artiodactyla, other than the cattle; and the other 6 (horse, human, mouse, *Canis lupus familiaris* [CFA - dog], *Rattus rattus* [RRA – black rat] and *Trichosurus vulpecula* [TVU – common brushtail possum]) species are from different orders of increasing divergence (Figure 2-2).

2.3.4 Equine BAC clone FISH experiments

Initial testing involved checking the 64 FITC probes and the 64 Texas Red probes on equine chromosome metaphase preparations to evaluate the degree of hybridisation of these subtelomeric BACs, and to confirm whether they hybridise to the correct chromosomal locations as indicated by the NCBI. The equine chromosome preparations were karyotyped (section 2.1.2) prior to testing, to make sure a normal 64, XX horse was being used as a standardisation. It is assumed this horse had no cryptic translocations. Once all probes were tested individually against a “normal” horse, each probe was karyotyped to make sure they were hybridising in the correct location according to the NCBI. If probes were hybridising to the correct location, singular FITC and Texas Red probes found on the same chromosome were paired to establish if they were overlapping or could hybridise simultaneously to the chromosomes. This also allowed a final check to be made to ensure they were both correctly karyotyped as on the same chromosome. The probes were then paired per chromosome and placed onto octochrome (section 2.3.5) and multiprobe (section 2.3.6) devices to consequently test on 19 horses which had previously been karyotyped to investigate subfertility or infertility issues.

2.3.5 Octochrome slide preparation and template hybridisation (equine)

This section refers to specific aim 2 (section 4). Commercial slides (CytoCell Ltd., Oxford Gene Technology, United Kingdom) with an octochrome configuration have eight number labelled

boxes per slide. A specialised commercial template slide (CytoCell Ltd., Oxford Gene Technology, United Kingdom) was required to combine chromosome preparations with probe mixtures on individual squares on the octochrome. Chromosome suspensions were spun down for 10 minutes in a centrifuge at 1,900 RPM, prior to removal of the supernatant and addition of 0.5μL fixative for resuspension of the pellet. 5μL chromosome suspension was added to each square on the slide, followed by 5μL fixative, to spread the metaphases across the squares. The slides were then placed through a dehydration series of four different solutions, for two minutes each of 2xSSC (ThermoFisher Scientific, Waltham, United States), 70% ethanol, 85% ethanol and 100% ethanol at room temperature. This was to ensure the metaphases are prepared for hybridisation and they are dehydrated to the slide.

On each square of the template slide, 4μL of individual dual colour probe mixtures were pipetted (with no equine hybloc addition as it was not commercially available) and then added to the octochrome device (Figure 2-3). To initiate hybridisation, the "sandwiches" were placed on a Hybrite hotplate for 10 minutes at 37℃, before DNA denaturation for 5 minutes at 75℃. The slides are then placed in a 37℃ humidified chamber for 24 hours.

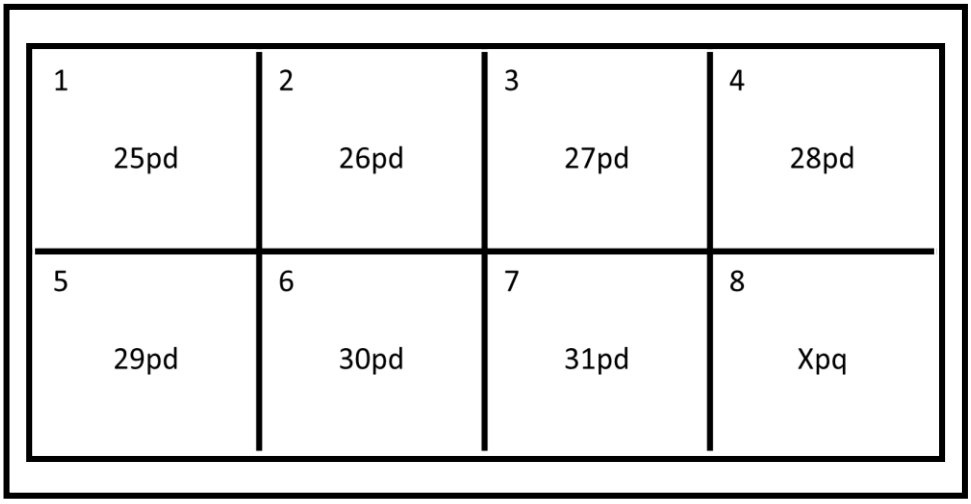


Figure 2-3 Octochrome device layout. Each box (top left of each square) indicates a different chromosome (middle of each square). Boxes 1-8 were used for equine chromosomes 25-X. Proximal (p) to the centromere probes are labelled with FITC and probes distal (d) to the centromere were labelled with Texas Red. The p-arm was labelled in FITC, and the q-arm was labelled in Texas Red for equine chromosome X.

24 hours post-incubation, the template slide was removed from the octochrome, and the octochrome was placed through the same second day FISH protocol as in the previous section.

2.3.6 Multiprobe preparation and template hybridisation

As with section 2.3.5, this methodology is specific to specific aim 2 (section 4). As with octochrome devices, specialised commercial multiprobe slides (CytoCell Ltd., Oxford Gene Technology, United Kingdom) contain a set number of labelled boxes (24 boxes) on a single slide. A specialised commercial template slide (CytoCell Ltd., Oxford Gene Technology, United Kingdom) was required for hybridisation to occur. Blood culture chromosome preparations were centrifuged for 10 minutes at 1,900 RPM prior to removal of supernatant and addition of 0.5mL fixative for resuspension. 2μL of chromosome preparation was added to each box on the multiprobe slide and then 2μL fixative added for fixation. The slides were placed through a dehydration series of 2xSSC (ThermoFisher Scientific, Waltham, United States) and 70%, 85% and absolute ethanol at room temperature at 2 minutes per solution.

On the template slide, 2μL of individual probe mixtures were pipetted in each box prior to

1 1pq	**2** 2pq	**3** 3pq	**4** 4pq	**5** 5pq	**6** 6pq	**7** 7pq	**8** 8pq
9 9pq	**10** 10pq	**11** 11pq	**12** 12pq	**13** 13pq	**14** 14pd	**15** 15pd	**16** 16pd
17 17pd	**18** 18pd	**19** 19pd	**20** 20pd	**21** 21pd	**22** 22pd	**23** 23pd	**24** 24pd

Figure 2-4 Multiprobe device layout with each box (number in the top left of each square) indicating a different chromosome (middle of each square). Boxes 1-24 were used for equine chromosomes 1-24. Metacentric equine chromosome 1-13 p-arm probes were labelled with FITC, and q-arm probes were labelled with Texas Red. Acrocentric equine chromosomes 14-24 were labelled as proximal (FITC, p) and distal (Texas Red, d) to the centromere for each chromosome.

inversion onto the multiprobe slide (Figure 2-4). This “sandwich” was heated on a 37℃ Hybrite hotplate for 10 minutes to aid hybridisation, preceding probe, and DNA denaturation for 5 minutes at 75℃. The “sandwiches” were placed in a humidified chamber for 24 hours.

24 hours post-incubation, the template slide was removed from the multiprobe, and the multiprobe slide was placed through the same second day FISH protocol as in the previous section 2.3.5. The multiprobe and octochrome devices were compared with regards to their hybridisation results to see if there is a difference between them using an unpaired t-test at $p < 0.05$.

2.4 Microscopy and Image Analysis

2.4.1 Microscopy

All microscopy work was visualised using an Olympus BX61 epifluorescent microscope. A cooled charge-coupled device (CCD) camera was used to capture images on SmartCapture 3 software (Digital Scientific UK, Cambridge, United Kingdom) with Texas Red, FITC and DAPI filters at a x100 magnification. Microscopy was used for specific aims 1 (section 3), 2 (section 4) and 3 (section 5).

2.4.2 Karyotyping

Images for karyotyping were captured and exported to SmartType 2 software (Digital Scientific UK, Cambridge, United Kingdom) for analysis. Where it was available, the International System for cytogenetic nomenclature for a specific species, was used to karyotype an animal. If it was unavailable, reference images from published literature were used to attempt to karyotype a species. A minimum of 5 karyotypes were created for each animal from a specific species evaluated for specific aims 1 (section 3) and 2 (section 4).

2.5 Cross species analysis

This section describes methodology used for specific aim 1 (section 3). All correlations calculated for this section were deemed significant if $p < 0.05$ and were based on the strength of the correlation (R) and the number (N) of data points included in the calculation of the correlation.

2.5.1 FLPter measurements

20 measurements were made for each of the 48 probes on each of the 11 species. The (i) full length of the chromosome (n = 5 measurements), (ii) the individual arms (n = 5 measurements) and (iii) the distance from the centromere to the probe was measured (n = 10 measurements). The same method of measurement was used for each individual image for each probe in order to mitigate user error. ImageJ (version 1.51r, Rasband W., National Institutes of Health, United States) was used to analyse the fractional length relative to the p-terminus (FLPter) measurements, which compares the relative distance of the probe location to the centromere

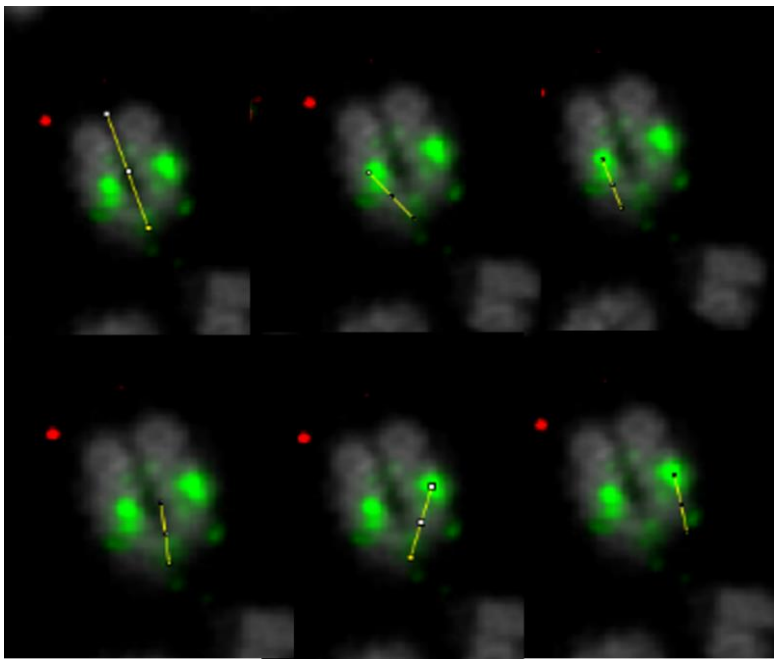


Figure 2-5 Examples on the same chromosome of how different FLPter measurements were made. The top left picture indicates how a full chromosome measurement was completed and the other images are related to measurements from the centromere to the probe (yellow line).

based on measurements. The ratio of the individual probes (iii) to both the chromosomal length (i) and to the length of the chromosomal arms (ii) was determined by the mean (μ) of all the measurements taken (Figure 2-5). Further description of the use of these values can be found in section 2.5.5.

2.5.2 Ideogram generation

Using Microsoft PowerPoint, ideograms were developed based on karyotype images. Banding patterns were recreated using measurement analysis and visual interpretation to gain the most accurate image, with a fractional- and ratio-based technique. These images were validated by comparison of multiple karyotypic images to consider any variation between the metaphase spread banding.

2.5.3 FLPter measurements on ideograms

By applying the FLPter value to the ideograms developed, and by using a reference genome with known BAC order, the probe movement cross-species by chromosomal rearrangement was determined. The FLPter value was applied to the ideogrammatic chromosome axis relative to the p-arm of the chromosome, to develop a visual representation of characterisation without the need for sequencing information. Standard deviations (s.d., σ), standard errors (SE) and accuracy values for the FLPter measurements when compared to the genomic locations according to the NCBI were calculated. A paired t-test was used to test significant differences between the average FLPter measurement made and the genomic location according to the NCBI database (p-value < 0.05).

2.5.4 BLAST searches

The 48 cattle BAC clone sequences were searched individually using the "BLASTn" function from the NCBI on six different species, namely bison, sheep, horse, dog, human and mouse. Query cover, percent identity and location were noted for each of the species to give an indication of the accuracy of the sequence allocation within each of the species. These results were compared to the image analysis completed using the fluorescent microscope described in section 2.4.1, and the FLPter measurements made.

2.5.5 Evolutionary genes of interest

Genes of agricultural, fertility, functional or medical interest were identified in the specific probe sequences found on the cattle ARS-UCD1.3 genome assembly. The genes found in cattle within the highly conserved regions, were checked against five highly researched genomes which can be found in Table 2-3 (sheep, horse, dog, human and mouse). If orthologues were identified, chromosomal and sequence locations were then compared by comparative mapping, BLAST searches and FLPter values to confirm if the highly conserved sequences in the species had the genes in them. Evolutionary connections between the orthologues and species

Species	Genome assembly
Bos taurus (BTA) – Domestic cattle	ARS-UCD1.3
Equus caballus (ECA) – Domestic horse	EquCab3.0
Homo sapiens (HSA) – Human	GRCh38.p14
Rat rattus (RRA) – Black rat	Rrattus_CSIRO_v1
Canis familiaris (CFA) – Domestic dog	CanFam3.1
Ovis aries (OAR) – Domestic sheep	Oar_rambouillet_v1.0

Table 2-3 The genomes for each species which were researched in this section, for evolutionary conserved genes using BLASTn. This was done with the original cattle BAC clone sequences, a gene search, followed by an orthologue search.

were investigated.

A comprehensive list was compiled of all the genes found within these cattle BAC clone regions. Genes were only further investigated if there was published literature on the specific gene in the individual species and had either one or multiple of the following three criteria:

(i) an FLPter value which was <20% of the difference between the chromosomal position and measurement made as well as where the chromosome on the NCBI location matched the BLAST search chromosomal location

(ii) the chromosome on the NCBI location matched the BLAST search chromosomal location yet had no FLPter measurement due to lack of a hybridisation signal or the signal/FLPter measurement was unclear

(iii) the FLPter value was <30% of the difference between the chromosomal position and measure made.

2.6 Semen information

The following subsections pertain to specific aim 3 (section 5).

2.6.1 Stallions

Stallion semen was provided by Stallion AI Services Ltd. (Whitchurch, United Kingdom) and received in either chilled at ~15℃ (n = 12) (in extender) and/or liquid nitrogen straws (in cryoprotectant) maintained at -196℃ (n = 30) from different stallions (N = 20). Table 2-4 describes which stallions had their semen chilled and which were frozen, how many straws (frozen)/vials (chilled) were tested, as well as what type of cryoprotectant (including DMSO) the frozen semen samples were placed into, either egg- (n = 18) or milk-based (n = 8) (proprietary information). The type of cryoprotectant used was estimated based on the colour of the sample when thawed.

Stallion	# Frozen	# Chilled	CP		Stallion	# Frozen	# Chilled	CP
A	2	5	Milk		K	2	0	Egg
B	2	0	Milk		L	2	0	Egg
C	2	0	Egg		M	2	0	Egg
D	2	0	Milk		N	2	0	Milk
E	2	0	Egg		O	1	1	Milk
F	2	0	Egg		P	1	1	Egg
G	2	0	Egg		Q	0	2	Unknown
H	2	0	Egg		R	0	1	Unknown
I	2	0	Milk		S	0	1	Unknown
J	2	0	Egg		T	0	1	Unknown

Table 2-4 The list of stallions (labelled with letters A-T), with a specific number of frozen and chilled samples per stallion. The cryoprotectant (CP) used for each stallion is also listed. The cryoprotectant was selected by Stallion AI Ltd specific to each stallion and specific details were proprietary information, however we were provided with the CP to be egg- or milk-based. Therefore, this was assumed based on the colour post-thaw.

The choice of which type of cryoprotectant to use was dependent on which one worked best for the individual stallion's sperm and semen characteristics. This was determined by Stallion AI Services Ltd. It is well known that both egg- and milk-based cryoprotectants have membrane-stabilising effects which aid in protecting the sperm against cold shock (Gibb & Aitken, 2015).

Pregnancy results were obtained for n = 6 of the stallions, three were considered bad (horses B, G and N) and three were considered good (horses E, F and J). For frozen semen, a poor pregnancy rate is considered less the 50% success per cycle, and/or less than 75% of pregnancies by the end of the season, on multiple cycles (per mare). Sample collection dates ranged from 19/04/2011 to 29/06/2021.

Prior to receiving chilled samples, they were diluted by Stallion AI Services Ltd on a 1:1 basis with a pre-warmed 37°C milk- or egg-based extender before being placed into a 4°C fridge and then shipped in an ice box to the University of Kent.

Frozen samples were collected from the stallions by Stallion AI Services Ltd and centrifuged with INRA 96™ at ~400 x g for 10 minutes at room temperature, prior to removal of the extender, and resuspension of the sperm pellet in the egg/milk-based cryoprotectant. The frozen sample straws ranged from 200-300 million sperm/mL concentration per straw. They were shipped from Stallion AI Services Ltd to the University of Kent (Canterbury, United Kingdom) in a liquid nitrogen shipping container and then moved into a -196℃ liquid nitrogen storage container.

2.6.2 Humans

Human semen was provided by Andrology Solutions Ltd. (London, United Kingdom) and The Doctors Laboratory (TDL) (London, United Kingdom) in cryopreserved aliquots (n = 44). The men involved had varying abstinence periods ranging from 2 to 5 days prior to producing a sample on site at TDL. Samples were placed into a 37°C incubator to allow liquefaction for 30 minutes prior to standard WHO regulation semen analysis. Extra semen samples which were not used for the semen analysis were snap frozen with no extender into 2mL cryovials within 1 hour of ejaculation and placed immediately into a -80°C freezer. These aliquots were collected by the University of Kent in a dry ice shipping box and samples were immediately placed into a -80°C freezer at the University of Kent for further testing.

All men had previously signed forms allowing for their semen to be used for scientific testing purposes and each individual was assigned a number in order to maintain anonymity. Available semen analysis information on all the patients was collected from the TDL database. This data was based on the fresh semen collected post-ejaculation at the TDL premises, prior to snap freezing the samples in cryovials at -80℃ in the TDL laboratory. This information allowed for comparison between fresh and frozen samples.

2.7 Semen parameters

A LensHooke X1 PRO Semen Quality Analyzer (CASA) (Bonraybio, United Kingdom) was used to obtain semen quality information about the samples prior to analysis for specific aim 3 (section 5). 30µL of sample was placed on the LensHooke cassette and inserted into the CASA device. Volume, colour, and liquefaction time were manual input items, but the CASA analysed pH, sperm concentration (million sperm/mL), total sperm number, progressive motility (PR, %), total motility (TM = PR + NP, %), non-progressive motility (NP, %) and normal morphology (%). As this CASA system is designed for human sperm, stallion morphology could not be analysed, yet other equine factors were still investigated for both species.

A MiOXSYS system (Caerus Biotechnologies, Geneva, Switzerland) was used to obtain ORP (mV) information about each of the samples prior to analysis. A MiOXSYS sensor was placed into the machine and then 30µL of sample was transferred onto the sample applicator port of the MiOXSYS sensor. The analyser started to process the sample once it had detected it. An ORP measurement was produced for the sample, and as it does not distinguish between stallion and human sperm, it could be used for both species. A mV/million ORP value was calculated, if an exact CASA concentration number was given.

2.8 Lipid Peroxidation Assay

Similarly, to section 2.8, these protocols are relevant to specific aim 3 (section 5).

2.8.1 Commercial LPO kit description

A cell-based, non-gametic LPO assay kit (ab243377) from Abcam (Abcam, Cambridge, United Kingdom) was tested to potentially be used as a future LPO kit for semen. This kit was based on a ratiometric LPO Sensor (LPS) (proprietary information) which indicated LPO due to 4HNE in cells by changing its fluorescence from red to green. By measuring the fluorescence change,

an estimate of LPO can be evaluated. The protocol stated the user should analyse the results with a flow cytometer and fluorescent microscope at their own discretion.

The kit included a vial of H_2O_2 (1 M, 4000x) to use as a positive control and a booklet which described the methodology of the experiment. This kit was aimed to be used on live non-gametic cells in a 37℃ culture-based situation. Sperm cells should be alive when they are thawed post-freezing, hence testing them with a live cell-based assay was important when selecting a commercial test to optimise.

2.8.2 Optimisation of commercial LPO kit for stallion and human semen

A 10x working solution of the LPS was made by diluting the 500x stock solution in PBS and not Hank's balanced salt solution with Hepes (HHBS) as suggested in the methodology. This component of the protocol was altered as PBS is often used to clean sperm cells due to its non-toxic capacity and it also stops cells from lysing (Martin *et al.*, 2006). As a positive control, the recommended final working concentration (based on the protocol booklet) was 250 µM (1x) solution of hydrogen peroxide added to the cells for 30 minutes. A 1x final concentration of LPS was advised to be used and this was dependent on the volume of the cells used, not the number of cells in the sample, for example 10 µL LPO sensor added to 90 µL of cells. Based on the SCSA®, a concentration of 2 million sperm/mL was used as a standard for the experiment. It was noted this could affect the outcome of the results, as different individual stallions and males have varying numbers of sperm within their semen samples, which would also be the case for non-gametic cells.

The initial tests were conducted to find the ideal conditions for testing the sperm with this commercial LPS and kit. (1) Different concentrations of H_2O_2 (1 M (n = 5), 100 mM (n = 2), 10 mM (n = 2), 1 mM (n = 4), 500 µM (n = 29) and 250 µM (n = 9)) were diluted to obtain initial information on LPO damage as a positive control using the $C_1V_1 = C_2V_2$ formula. (2) Different concentrations of sperm (10 million/ml (n = 2), 5 million/ml (n = 3) and 2 million/ml (n = 3))

were tested to find the ideal sperm concentration for testing. (3) Different concentrations of the LPS (1x (n = 3), 0.5x (n = 3), 0,2x (n = 2) and 0.1x (n = 3)) were evaluated to determine if the concentration of the sensor needed to be lowered to give a better indication of the sperm in the sample, as it is known that sperm cells are significantly smaller than most other cells grown in culture (Son *et al*., 2017). (4) Exposing spermatozoa to H_2O_2 in incubation for changing time lengths (30 minutes (n = 2), 1 hour (n = 3) or 2 hours (n = 2)). (5) Testing samples at different times after exposing the sample to LPS and washing (Immediate testing (n = 7), 30 minutes (n = 2) and 1 hour (n =7) after LPS washing).

Only one of the parameters 1-5, which can be found in Table 2-5, were tested at a time to ensure there was an accurate representation of the results. This also allowed the discrepancies seen for that particular experiment to be pinpointed to a specific change made. Variables used for each experiment other than the one being tested can be seen in Table 2-6.

Test	V1	V2	V3	V4	V5	V6
(1) H_2O_2Concentration	1M	100mM	10mM	1mM	500µM	250µM
(2) Sperm Concentration (million/mL)	10	5	2	-	-	-
(3) LPO Sensor Concentration	1x	0.5x	0.2x	0.1x	-	-
(4) H_2O_2Incubation Time	30 min	1 hour	2 hours	-	-	-
(5) Length of time after LPS washing	No time	30 min	1 hour	-	-	-

Table 2-5 Tests conducted for the optimisation of the non-gametic lipid peroxidation (LPO) kit for sperm. Standardised sample variables are demonstrated in Table 2-6. Highlighted values are noted as the controls in each experiment based on what was suggested by Abcam, for comparison purposes. V: Variables.

As the optimal LPO testing conditions included (1) 500 µM H_2O_2, (2) no specific sperm concentration as long as the FC flow rate was less than 300 cells/second, (3) 1x LPS concentration, (4) 30 minute H_2O_2 incubation for positive controls and (5) immediate testing after LPS washing, these parameters were used for subsequent testing (n = 75). The results for

both the optimisation of the LPO kit and the subsequent patient testing are discussed in specific aim 2 (section 4).

Test	H_2O_2	M/mL	LPS	H_2O_2 IT	Testing
(1) H_2O_2 Concentration	-	2	1x	30 min	No time
(2) Sperm Concentration (million/mL)	500 µM	-			
(3) LPO Sensor Concentration		300 cells/sec	-		
(4) H_2O_2 Incubation Time			1x	-	
(5) Length of time after LPS washing				30 min	-

Table 2-6 The standardised variables for each test described in Table 2-5. M/ml: Million sperm/mL. IT: Incubation time. Testing: Time to testing after lipid peroxidation sensor (LPS) washing. No time: Immediate testing. 300 cells/sec: 300 cells/second flow rate on the flow cytometer.

2.8.3 LPO laboratory protocols

The original LPO protocol describes growing the cells intended for analysis overnight in a 37°C incubator with 5% carbon dioxide (CO_2). As cells were not grown, for stallion semen analysis, the liquid nitrogen stored straws were placed for 60 seconds into a 37°C-water bath to liquify the samples, while human semen cryovials were thawed at room temperature for 15 minutes. The cryovials were not placed into a 37°C incubator as they had previously undergone liquefaction prior to snap freezing.

The samples were gently pipette mixed prior to moving 250 µL of the sample into a new Eppendorf tube and 1 mL of PBS was added to the sample. The semen was centrifuged in a tabletop centrifuge at 400x g for 10 minutes and the supernatant removed, so only the sperm remained. Fresh sterile PBS was then added to the pellet depending on the concentration of sperm in the sample (based on the information provided), what the final concentration of the sample was going to be for that test (for example, 2 million/ml, 5 million/per etc.) and the number of repeats that were going to be run (at least three, two repeats and one positive control). These were then aliquoted into Eppendorf tubes.

Positive control cells were treated with H_2O_2 (depending on the concentration of H_2O_2 being tested/used) at the same volume as the sample volume per Eppendorf tube. On the commercial protocol, no specific volume of H_2O_2 was stipulated and thus this 1:1 ratio of sperm and PBS volume to H_2O_2 volume was selected. The positive control tube and the sperm Eppendorf tubes were then placed into a 37℃ incubator for 30 minutes. LPS was added to the cells depending on the concentration of LPS being tested/used (for example 1x, 0.5x etc.) and as per the commercial protocol, were incubated for a further 30 minutes at 37℃ (yet not including 5% CO_2).

The samples were then washed with 1 mL PBS instead of HHBS, spun at 400 x g in a tabletop centrifuge for 10 minutes and the supernatant was removed. This step was repeated three times to thoroughly wash the sample. It was noted that centrifuging sperm often causes extra damage to the sperm and can often separate the heads from the tails, thus damage reported in the results could be an overestimate of the actual LPO damage present in the spermatozoa.

The sample was then moved into a flow cytometry tube and then analysed with a BD Accuri C6 Plus Flow Cytometer (Becton, Dickinson and Company, Franklin Lakes, United States) by monitoring the 488/530 nm (FITC) and 488/572 nm (phycoerythrin, PE) channels. This was done within two hours of staining the samples with the LPS. Further analysis information is discussed in section 2.8.4.

2.8.4 LPO Data Analysis

Figure 2-6 was given as an example in the Abcam LPO assay (cell-based) protocol booklet (ab243377) of what to expect when analysing the commercial LPO cell-based non-gametic kit results. On a graph of red/green ratio (x axis) to counts (y axis), for the positive control, a narrow bell curve with a low ratio of red/green should be expected, thus the more damaged the cells are, the lower the red to green ratio should be. For the samples being tested, a flatter and wider bell curve should be visualised with a higher red/green ratio.

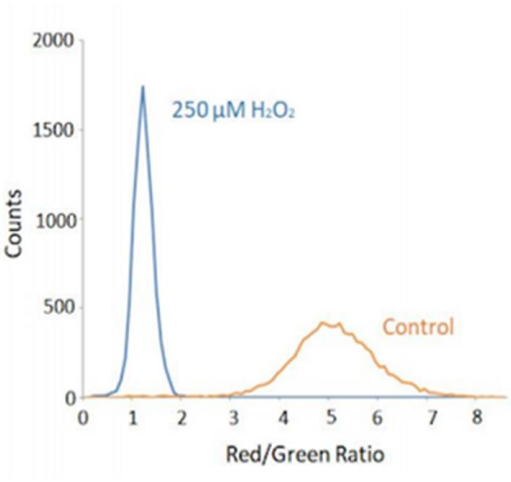


Figure 2-6 Abcam lipid peroxidation (LPO) assay graph of red/green ratio (x axis) to counts (y axis). The positive control using 250 µM hydrogen peroxide (H_2O_2) is indicated on the left curve (blue), and sample control on the right curve (orange). (Image source: Abcam, 2019).

Therefore, it is assumed, red fluorescing cells are considered to be healthy cells and green fluorescing cells are LPO damaged cells. No information was given on how to gate the population and determine what the cell population was on a flow cytometer.

Analysis was done based on research by Aitken *et al.* (2007) (Figure 2-7) and what was suggested by the Abcam commercial protocol booklet (Abcam, 2019) (Figure 2-6). Data was analysed as described below.

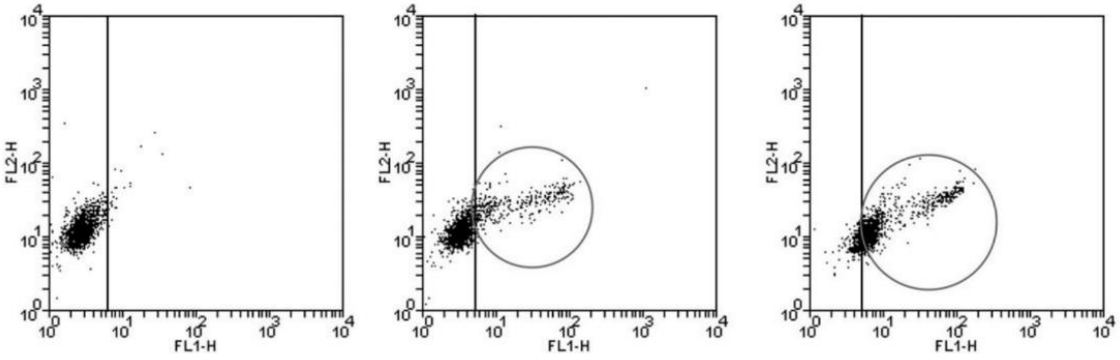


Figure 2-7 Graphs from Aitken *et al*. (2007) demonstrating the analysis of the impact of time and ferrous ion promotion on lipid peroxidation (LPO) in sperm with BIODIPYC_11, also an lipid peroxidation sensor (LPS). From left to right: Flow cytometry analysis of control cells (left); and flow cytometry analysis demonstrating the progressive increase in BIODIPYC_11 green fluorescence in human sperm samples treated with 80 µM Fe(II) for 15 minutes (middle) or 60 minutes (right). FL1-H (x axis) is fluorescence intensity and FL2-H (y axis) are events. (Image source: Aitken *et al*., 2007).

(1) A forward scatter (FSC) (x axis) versus side scatter (SSCa) (y axis) plot was created to give an indication of the cell size and complexity in the experiment as a quality control to ensure there

were whole spermatozoa in the sample. By using this plot, a differentiation was made between the cells in the population which are larger and of higher complexity (upper right quadrant, Q1-UR) as sperm cells are complex cells which are larger than the other artefact particles in the sample, which are debris (upper left quadrant, Q1-UL). If a sample is highly damaged and many sperm heads have been separated from their tails, it is expected to see more Q1-UL than in a healthy sample, and there will be a less definitive differentiation between the two populations (Figure 2-8A). If there was no definitive population, the sample was excluded from further analysis.

(2) A FITC (x-axis) versus FSC (y-axis) was developed to give an indication of the damaged cells in the sample. A population of cells should be present in the data which are of larger size (increased FSC) and damaged (increased FITC). These were highlighted (E2) and percentage of the plot shown by E2, noted for future calculations (Figure 2-8B).

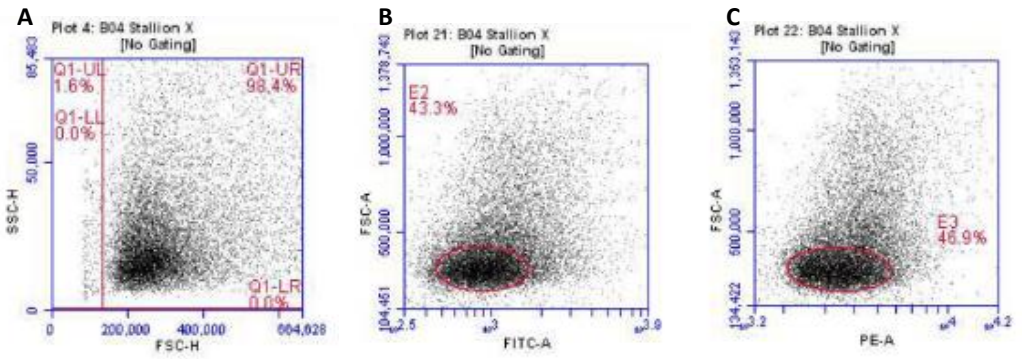


Figure 2-8 An example of the methods used to determine the sperm population and lipid peroxidation (LPO) ratio. (A) A forward scatter (FSC) vs side scatter (SSCa) plot as a quality control (QC) for sperm in the sample (Q1-UR); (B) FITC vs FSC graph used to indicate damaged cells in the sample whereby E2 shows the sperm population; and (C) PE vs FSC graph used to indicate damaged cells in the sample whereby E3 shows the sperm population.

The area selected for E in (2) and (3) was determined by the positive control used for a specific man or stallion, and not changed again for the individual sample run, as to give an accurate representation of the data. By changing the E for each graph within a specific run, the results become unreliable. Repeats of the same run were also conducted using the same positive control E gating parameters initially determined.

(3) A PE (x-axis) versus FSC (y-axis) plot was made to give an indication of the healthy cells in the population. Similar to graph (2), there should be a population of spermatozoa present in the graph which are bigger (increased FSC) and healthy (increased PE) (Figure 2-8C). They were highlighted (E) and percentage of the plot indicated by E, used for the following calculations.

To calculate the LPO ratio, the number of healthy spermatozoa to the number of LPO damaged cells based on the percentage of the plot indicated by respective PE (Figure 2-8C) and FITC (Figure 2-8B) E-values was done as follows:

$$\frac{Percentage\ of\ plot\ of\ PE\ (healthy\ cells)}{Percentage\ of\ plot\ of\ FITC\ (LPO\ damaged\ cells)} = LPO\ ratio$$

For each sample, a minimum of 2 repeats were conducted for accuracy and reliability of the results. These results were averaged for each specific test and a mean was determined as well as standard deviation. A minimum, maximum, median, interquartile range (IQR) and Pearson correlation coefficients ($p < 0.05$) were also developed for the data. Statistical testing included testing the samples against their positive controls and other tests such as (1), (2), (3), (4) and (5) from Table 2-6 (section 1.8.2). An analysis of variance (ANOVA) was developed for test (1), between four variables (No H_2O_2, 250 µM H_2O_2, 500 µM H_2O_2 and 1 M H_2O_2).

Stallion sperm was also evaluated for the following tests to identify possible effects on LPO and to correlate it to various sperm parameters. (1) Effect of cold storage (Frozen and chilled); (2) Breed effect (Warmblood, crossbreed ("sport") horse, Racing/endurance, other); (3) Pregnancy outcomes (good or bad); (4) Cryoprotectant type (egg- or milk-based) (Table 2-7). Samples for each stallion were averaged depending on what was being tested, for example if stallion A had n = 2 chilled samples on the same collection date, the average for those samples would be indicated in the results for each variable. Mean (µ), maximum, minimum, median, standard deviation (σ), IQR and Pearson correlation coefficients (R, $p < 0.05$) between different semen parameters were analysed for each test. T-tests were done to compare the means of

the different variables tested, where $p < 0.05$ was considered significant. The same tests were conducted for DFI and ORP for section 2.9.2.

Test	**V/O/E 1**	**V/O/E 2**	**V/O/E 3**	**V/O/E 4**
Effect of cold storage	Frozen	Chilled		
Breed effect	Warmblood	crossbreed ("sport") Horse	Racing/ Endurance	Other
Pregnancy rate	Good	Bad		
Cryoprotectant type	Egg-based	Milk-based		

Table 2-7 Different tests completed with the stallion semen and the variables in each test. V/O/E is dependent on the test. V: Variable; O: Outcome; or E: Effect.

2.9 DFI and HDS

Sperm was analysed for DFI (%) and HDS (%) by Marie Claire Aquilina using the following methodology. However, I did all the data and statistical analyses relevant to specific aim 3 (section 5).

2.9.1 SCSA® protocol

For DFI and HDS analysis, the SCSA® published protocol by Evenson (2016) was used. An acid detergent solution was made up of 4.39 g NaCl, 20 mL 2.0 N HCl and 0.5 mL Triton-X 100 (0.1%, ThermoFisher Scientific, Waltham, United States), purified water was added to a volume of 500 mL and this solution was adjusted to pH 1.2 with 5 N HCl. The AO stock solution (Invitrogen, Waltham, United States, 10 mg/mL) was diluted with purified water to 1 mg/mL. An AO staining buffer was made of 370 mL 0.1 M citric acid buffer, 630 mL 0.2 M Na_2PO_4 buffer, 372 mg EDTA (disodium, 1 mM) and 8.77 g NaCl (0.15 M) by mixing overnight to ensure the EDTA was completely dissolved in the solution. This solution was then adjusted to a pH of 6.0 using saturated NaOH. A staining solution of AO was made up from 600µL AO stock solution and 100 mL of AO staining buffer.

Thawed semen samples were measured directly for DFI (%) and HDS (%) after thawing. The samples were separated into three different tubes, and each was diluted with Tris-HCl, NaCl and EDTA (TNE) buffer to attain an approximate dilution of 1-2 million sperm/mL/Eppendorf tube. Each tube had the following steps done individually, as not to add potential artefacts such as OS and DNA breakage to the sample. 400µL of acid detergent solution was added to the sperm suspension and then vortexed. After 30 seconds, 1.2 mL of AO solution was added, vortexed and then analysed on a BD Accuri C6 Plus Flow Cytometer in the 488/530 nm (FITC) and 488/572 nm (PE) channels.

2.9.2 Data analysis

Sperm flow cytometric data was collected after running the sample for 3 minutes at a flow rate of less than 300 cells/second. If the flow rate was higher than this, a fresh and more diluted sperm sample was made. 5000 or more events were collected per sample, and 3 repeats were conducted based on the other tubes from the same sample.

The number of red (ssDNA) fluorescing cells (x axis) versus the number of green (dsDNA) fluorescing cells (y axis) can be seen in Figure 2-9A. This diagram by Evenson (2016) describes how the more damaged the DNA in the sample is, the redder fluorescing cells will be seen overall and there will be a colour shift from green (normal) to red (fragmented). Sample gating was conducted according to Evenson (2016) in Figure 2-9B. Figure 2-10 describes the methods of calculating %DFI and %HDS (Evenson *et al.*, 1985). These methods were used for each sample and then the %DFI and %HDS were averaged over the three to give a mean and s.d. (µ) for each sample. Medians, IQR, and Pearson correlation coefficients (R, $p < 0.05$) were also determined for, and between, DFI, LPO, ORP, TM, PR, vitality, and morphology.

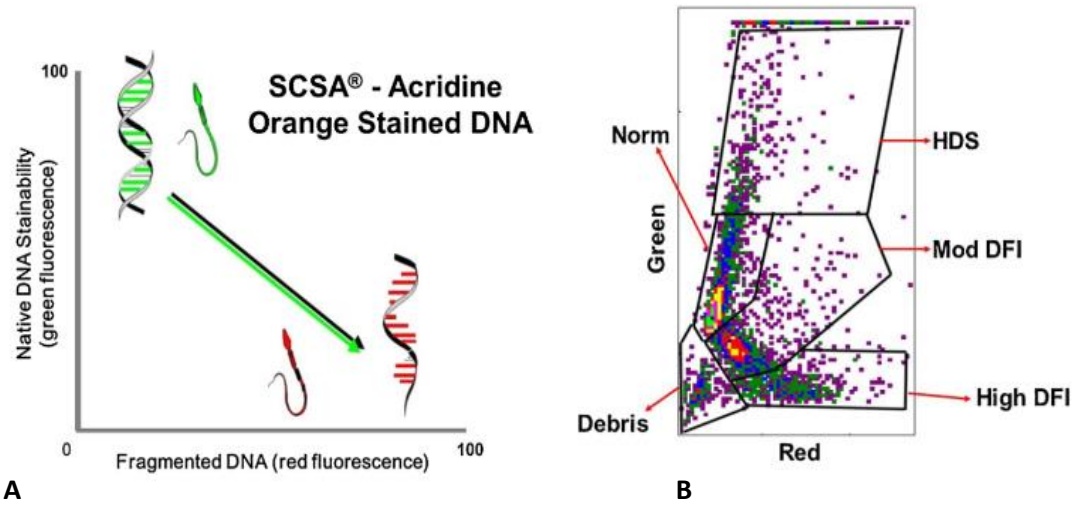


Figure 2-9 Fragmented DNA or single stranded DNA (ssDNA) (red fluorescence) (x axis) versus the native DNA stainability or double stranded DNA (dsDNA) (green fluorescence) (y axis) (18A). The change from green to red is used to calculated Sperm Chromatin Structure Assay® (SCSA®) DNA fragmentation index (DFI) (%). Gating of the SCSA® sperm data (18B) used from Evenson (2011). Norm: Normal sperm; HDS: High DNA Stainability; Mod DFI: Moderate DFI. (Image source: (A) Evenson, 2016; (B) Evenson, 2011).

$$\text{DFI} = \frac{\text{red fluorescence}}{\text{total (red + green) fluorescence}}$$

X DFI = mean of DFI population (1-1024 units)
SD DFI = standard deviation of DFI population
%DFI = % cells outside main sperm population
Moderate DFI and High DFI

Figure 2-10 The formula for DNA fragmentation index (DFI), with the main variables calculated from this formula, below it, including the mean of the DFI in the population, standard deviation of the DFI population, and the percentage DFI. (Image source: Evenson *et al*., 1985).

2.10 Semen analysis accreditation survey

The following subsections are important for specific aim 4 (section 6).

2.10.1 Respondents

A cross-sectional email survey was sent out in May 2021 using Jisc Online Surveys (Jisc, Bristol, United Kingdom) (Supplemental Table I). The survey was sent to participants in the UK National External Quality Assessment Service (NEQAS, Sheffield, United Kingdom) for Andrology (n = 184), all members of the Association of Reproductive and Clinical Scientists (ARCS, Brentford, United Kingdom) members (n = 682) and all fertility clinics licensed by the

Human Fertilisation and Embryology Authority (HFEA, London, United Kingdom) (n = 117). A general introduction to the survey was provided prior to taking part.

A total of 108 responses were recorded. Any clinics or laboratories which are not part of the United Kingdom were excluded (n = 1) as this survey was intended to represent the current local coverage of fertility clinics and laboratories. One response (n = 1) was excluded as it was not answered by a trained laboratory member of staff. Some laboratories would have been contacted more than once if they were registered with ARCS, HFEA and/or NEQAS. Individual survey answers were analysed for overlaps and repeats within the answers to determine whether there were any persons answering more than once and none were found.

Response	Initial category	New category
766430-766421-81310895	HFEA (Group 3)	UKAS (Group 1)
766430-766421-81311209	Neither UKAS nor HFEA (Group 4)	UKAS (Group 1)
766430-766421-81301551	Neither UKAS nor HFEA (Group 4)	UKAS (Group 1)
766430-766421-81310341	UKAS (Group 1)	Neither UKAS nor HFEA (Group 4)

Table 2-8 The survey responses which were automatically classified into initial categories by the Jisc Online Survey Tool, and then were manually set into new categories based on the specific answers the respondents gave. Response numbers (left), initial category (middle) and new category (right) are shown.

Respondents were allocated to groups for ease of understanding. Group 1 had United Kingdom Accreditation Service (UKAS, Staines-upon-Thames, United Kingdom) only accredited laboratories; Group 2 contained both UKAS and HFEA accredited organizations; Group 3 was HFEA only; and Group 4 respondents had neither HFEA nor UKAS accreditation. Based on these groups, the following answers were moved, 766430-766421-81310895; 766430-766421-81311209; 766430-766421-81301551 and 766430-766421-81310341. Answers 766430-766421-81310895 and 766430-766421-81311209 were moved to (iii) due to application for United Kingdom Accreditation Service (UKAS) at the time of sending out of the survey and 766430-766421-81301551 was moved to (iii) as it was CPA classified and thus UKAS only, due

to the move from CPA to UKAS. 766430-766421-81310341 was removed from the UKAS only category, due to being accredited for microbiology, but not andrology. These are seen in Table 2-8.

2.10.2 Bias

In order to reduce bias, the survey made clear that the answers should be given by only one individual from each laboratory by a person holding a permanent laboratory position familiar with semen analysis (determined by question 2 in the survey), and that the answers should pertain to current laboratory practice, rather than the personal views of the respondent.

2.10.3 Design

The survey questions were designed to demonstrate how laboratories in the UK perform semen analysis and how the results are used to diagnose and manage the patient. At the time of sending this survey, the current WHO guidelines for semen analysis were published in 2010 (5th edition). A complete list of the questions in the survey is provided in Supplemental Table I. Each question in the survey had to be answered to continue. Comment sections were made available for justification of answers if necessary. The survey was designed to take no more than 10 min to complete, and all responses were collected anonymously.

2.10.4 Ethics

It was determined that Ethics and research committee approval was not required to undertake this cross-sectional email survey evaluation as no patients were contacted.

2.10.5 Analysis tool and statistical analysis

The data was analysed using the Jisc Online Survey Analyse tool. Statistical analyses were performed using a two-tailed Chi-square test at significance level of $p < 0.05$ for both within group, out of group and between category analyses. A Yate's correction for continuity was used to compensate for deviations for results less than $n = 5$.

3. Specific Aim 1 - Production of a panel of universally hybridising cattle BAC clones specifically chosen for sequence selection and conservation score, to expand research of mammalian comparative genomics

3.1 Background

3.1.1 Genetic variation and evolutionary breakpoint regions

As mentioned in section 1.1, each eukaryotic species has its own specific karyotype, which both gives an indication as to how its genome evolved and can be used to diagnose certain genetic disorders such as infertility (O'Connor *et al.,* 2018). Genomic changes and chromosomal rearrangements therefore define species to some degree and, technically, were first studied by karyotyping, followed by FISH, microarrays, and genome sequencing. Combining classical and molecular cytogenetics with modern day sequencing can generate chromosome level genome assemblies, facilitate detailed comparative genomics between species as well as provide more diagnostic accuracy for individual genome rearrangements.

Using comparative genomics in evolutionary research, has defined various evolutionary models and structural changes which can give an indication of genomic synteny interferences, such as evolutionary breakpoint regions (EBRs) and Homologous Synteny Blocks (HSBs) (Graphodatsky, Trifonov & Stanyon, 2011; Capilla *et al*., 2016; Deakin *et al*., 2019). EBRs have a higher chance of breaking and consequently rearranging, yet the underlying nature of instability is not well understood (Deakin *et al*., 2019). There are many theories on what causes speciation, yet EBRs are known to have a direct impact on reproductive isolation due to the number of repetitive sequences found in these regions, but whether this is due to the genomic instability of these regions or due to the specific chromosome's ability to change remains to be

determined (Capilla *et al*., 2016; Deakin *et al*., 2019). Transposable elements and segmental duplications found in repetitive regions are both heavily involved in non-allelic homologous recombination which causes structural rearrangements (Capilla *et al*., 2016; Deakin *et al*., 2019). Repetitive regions are also gene dense, thus it can be assumed functional-, adaptive-, fertility- and reproductive-compatibility diminishes when a breakage occurs as these genes are transferred into to another region of the genome thereby potentially improving specific gene expression and potentially adaptation characteristics (Graphodatsky, Trifonov & Stanyon, 2011; Capilla *et al*., 2016; Deakin *et al*., 2019). By choosing BAC clones that define EBRs, there is potential to study evolutionary change further and, more importantly, how and why these changes come about in different species (Graphodatsky, Trifonov & Stanyon, 2011).

3.1.2 Combining sequencing and molecular cytogenetics

Sequencing and molecular cytogenetics have allowed comparative genomics to evolve as a discipline over the last 40 years, into comprehensive information banks of bioinformatics and physical based information (Damas *et al*., 2016), using known well studied genomes to be overlaid onto relatively unknown species for further research (section 1.1.4). Both agricultural and medical fields have benefitted from the use of comparative mapping with the rapid identification of detrimental disease-causing genes and quantitative trait loci (QTL).

One of the main issues with sequencing and physical mapping, is the lack of connectivity between the two in many studies. For example, sequence scaffolds of a high-quality nature have been ordered into full genomes, yet many of these have not been checked on a chromosomal scale (Damas *et al*., 2016). These assemblies are also known to contain multiple errors and a combined effort is needed to correct these mistakes and thereby improve the assembly (Zimin *et al*., 2009). The aim of genome assembly is to produce a sequence from the p to the q terminus, in the correct order – a so-called chromosome-level assembly (Damas *et al*., 2016). Often, sequence scaffolds are incorrectly mapped or cannot be placed, resulting in

excess scaffolds (Mantere, Kersten and Hoischen, 2019). To try to aid this issue, sequence algorithms such as reference assisted chromosome assembly (RACA), are used to place these genomic fragments into full assemblies by creating long contigs (Kim *et al*., 2013). This algorithm requires a fully assembled reference genome from a singular species with which it subsequently orientates and places the next generation sequencing (NGS) scaffolds into fully predicted chromosome fragments (PCF) (Kim *et al*., 2013). Using a bioinformatics system such as the Evolution Highway Chromosome Browser, these PCFs can be visualised and compared to other species' genomes (Damas, Corbo & Lewin, 2021).

In 2017, PCFs created by RACA, in combination with validation through computational methods, permitted the first arrangement and visualisation of scaffolds on metaphase chromosomes through avian BACs (Damas *et al*., 2016), of pigeon, peregrine falcon, budgerigar, saker falcon and ostrich. The assemblies were analogous to proven mapping and sequencing methods and individually comprised more than 80% of the individual species' genomes (O'Connor *et al.,* 2018). Larkin *et al*. (2006) has used this combined technique in mammalian analyses, by utilising 19 cattle BACs and mapping them to mink chromosomes with FISH. Similarly, five synteny blocks were detected between human chromosome 17 and cattle chromosome 19 with this technique and the same BACs (Larkin *et al*., 2006), indicating a novel method of chromosomal analyses for comparative studies. Genomic conservation can thereby be selected for, depending on hybridisation success of BACs, for *in silico* analysis of evolutionary divergence. Through this technique of generating chromosome-level assemblies, a universal set of avian BACs was created.

3.1.3 Genomes and BACs

As discussed in section 1.3.4, Damas *et al*. (2016) developed a batch of universally hybridising chicken BACS which further initiated the ability to map bird genomes physically to a standard comparable to their respective reference genomes, aiding bioinformaticians with the large

number of unplaced genomic scaffolds (UGS) for those species. The opportunity to align and assemble genomes with the use of universally hybridising BACs, which are highly conserved through selection with the reference assisted chromosome assembly (RACA) algorithm, presents a unique method of study comparable to the Damas *et al*. (2016) work. The goal of this study was to use the BACs selected by the Royal Veterinary College (RVC) (described in section 2.2.1) and the previous work done by Jennings & Griffin (2019), to investigate if a universally hybridising set of cattle BACs for mammalian research can be adopted for evolutionary research and chromosomes can be mapped to reference sequence standard.

To achieve this, 10 species from two of the super-ordinal Placentalia clades, Euarchontoglires (Eutherians) - Primates (human) and Rodentia (mouse and rat); and Laurasiatheria – Cetartiodactyla (cattle, bison, pig, sheep, barasingha), Perissodactyla (horse) and Carnivora (dog); and one Marsupialia (possum), were selected for investigation due to their differing chromosomal traits such as karyotype diploid number, chromosomal morphology, and divergence time (in millions of years). Five of the 11 species selected form part of the Artiodactyla Order (including cattle, bison, sheep, barasingha and pigs) to test species evolution and synteny between species of the same order. One Carnivora (dog), one

Species	**Common name**	**Order**	**Super-ordinal clade**	**Divergence from** cattle **(MYA)**
BTA	Cattle	Artiodactyla	Laurasiatheria	0
BBI	American Bison	Artiodactyla	Laurasiatheria	4.88
OAR	Sheep	Artiodactyla	Laurasiatheria	24.80
RDU	Barasingha	Artiodactyla	Laurasiatheria	27.30
SSC	Pig	Artiodactyla	Laurasiatheria	62.00
CFA	Dog	Carnivora	Laurasiatheria	78.00
ECA	Horse	Perissodactyla	Laurasiatheria	78.00
HSA	Human	Primates	Euarchontoglires	96.00
MMU	House Mouse	Rodentia	Euarchontoglires	96.00
RRA	Black rat	Rodentia	Euarchontoglires	96.00
TVU	Brush-tailed Possum	Diprotodontia	Marsupial	159.00

Table 3-1 Species are depicted with their common name, order they belong to, and the divergence time from *Bos taurus* in millions of years apart (millions of years apart, MYA). BTA: *Bos taurus*. BBI: *Bison bison*. OAR: *Ovis aries.* RDU: *Rucervus duvaucelii*. SSC: *Sus scrofa*. CFA: *Canis familiaris*. ECA: *Equus caballus*. HSA: *Homo sapiens*. MMU: *Mus musculus*. RRA: *Rattus rattus*. TVU: *Trichosurus vulpecula*. Divergence times were calculated using TimeTree 5: An expanded resource for species divergence times (Kumar *et al*., 2022).

Perissodactyla (horse), one Primate (human), two Rodentia (mouse and rat) and lastly, one Diprotodontia (possum) were chosen for out of order evaluation. Table 3-1 demonstrates the species selected as well as their common name, the order they are part of among the evolutionary divergence times from the cattle reference. Figure 3-1 supports the phylogenetic tree of all the species tested from Table 3-1.

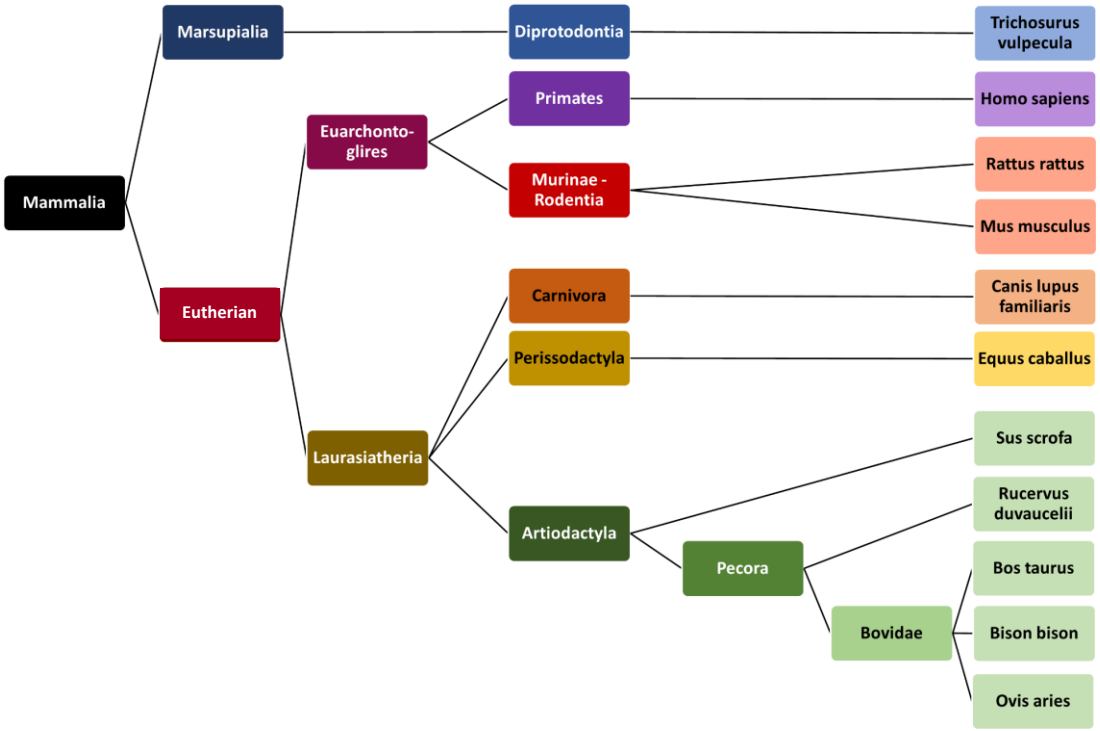


Figure 3-1 All the species tested (far right) in this specific aim, are shown in a phylogenetic tree indicating their various nodes of similarity and divergence from one another. Mammals in the same colour boxes, form part of the same order.

In previous work by Jennings & Griffin (2019), it was found that sequence-based BAC selection alone does not increase hybridisation success rate alone and more information was needed on the BACs. Hence the selection on conservation score for the BACs tested in this study. The intention was to map cattle BAC probes across highly divergent species and inferring their sequenced genomes *de novo* to a chromosome level.

3.1.3.1 Cetartiodactyla

Cattle, sheep, pigs, barasingha and bison, all form part of the Artiodactyla order of species, also known as even-toed ungulates. They are part of the mammalian class of which the cattle BAC clones were selected which originated ~53 million years ago. Approximately 200 different species of Artiodactyl's exist and comprise cattle, sheep, bison, barasingha and pigs, the first three of which belong to the Bovidae family, and first four, Pecora infraorder. It is known that for most of the Artiodactyla species, a small number of chromosomal rearrangements led to their reproductive isolation from one another many million years ago (Graphodatsky, Trifonov & Stanyon, 2011). This makes cattle an ideal model organism to use in comparative genomic studies, especially within order.

Many species, including three out of the five (cattle, pig, and sheep) selected for this study, are well known for their agricultural importance and hence have been studied comprehensively (sections 1.1.4, 1.3.1 and 1.3.2) (Zhou *et al*., 2015). Because of their value economically and the wide variety of chromosomal numbers and rearrangements within the Artiodactyl order, they are excellent to study for evolutionary understanding of how Robertsonian translocations can impact fixation of homologous chromosomal rearrangements in a species, which can then lead to speciation. This has been highlighted through the number of FISH and sequencing studies which have taken place in the last decade.

The cattle genome was first sequenced in 2009 by The Bovine Genome Sequencing and Analysis Consortium which cost $53 million and included researchers from 25 different countries (Zimin *et al*., 2009; Zhou *et al*., 2015). Btau4.0 was the initial sequence and it combined whole genome shotgun reads, BAC clone placements and various comparative maps from humans, thus giving a total size of 2.77 Gb (Zimin *et al*., 2009; Zhou *et al*., 2015). It covered 89% of the assembled contigs on all 30 chromosomes (29 autosomes and X chromosome) and the bovine genome has recently been updated to the UMD3.1 assembly by a different group of researchers with an alternative assembly algorithm (Zimin *et al*., 2009;

Zhou *et al*., 2015). The two genomes have significant differences between the two and hence the NCBI has recently allocated the UMD3.1 assembly as the designated “reference assembly” for cattle, removing ambiguity on which to use for further studies (Zhou *et al*., 2015). The UMD3.1 assembly was therefore used in this study. By combining the use of this reference assembly, karyotyping, fractional length relative to the p-terminus (FLPter) measurements and the 48 cattle BAC clones selected for conservation score, current gaps or incorrectly mapped locations in the assembly could be corrected.

When comparing the cattle assembly to other Artiodactyls, the closest relative to cattle in this study is the bison. When this study was conducted, only a poorly assembled reference genome (which was sequenced in 2014) was available, and there were multiple unplaced genomic scaffolds (UGS). However, recently the bison genome has been updated to chromosome-level by Stroupe *et al.* (2023). Similarly, when this study was conducted, the barasingha had no reference genome information available, and this is still the case as a lot of details are unknown about the Barasingha, except for its basic evolutionary history. Conversely, pigs have a 98.2% completed reference genome (Sscrofa11.1) assembled in 2017 by The Swine Genome Sequencing Consortium. Pigs have a genome size of 2.5 Gb and only 1.6% of the genomic information is missing. This could be due to its agricultural importance and the number of physical mapping studies which have contributed to assembly improvement (O’Connor *et al*., 2017). Lastly, the sheep genome used for this study is the ARS-UI_Ramb_v2.0, which was recently assembled in 2021. It has a much higher quality than the pig reference genome assembly, of 99.1%, and is 0.1 Gb smaller than the cattle genome (2.7 gigabases, Gb). The sheep reference genome assembly has a much higher contig N50 length than cattle, of 43.2x and 25.9x, respectively. This could be due to the recency of the updated published genome and the improved assembly methods, when compared to the cattle.

3.1.3.2 Perissodactyla

Perissodactyla, also known as the single-toed or odd-toed ungulates, originated 55-40 million years ago and includes the Equidae group of mammals, specifically the horse (Steiner & Ryder, 2011). They have a genomic history of major chromosomal rearrangements after which came rapid divergence and speciation, yet these genomic changes have fixed in recent years (Graphodatsky, Trifonov & Stanyon, 2011) and the order has a chromosomal number range from 2n = 32 in Mountain zebra, to 2n = 76-80 in the New World tapirs (Steiner & Ryder, 2011). Majority of Perissodactyla are either vulnerable, near threatened, endangered, or critically endangered (Steiner & Ryder, 2011) and thus research into their reproductive isolation could be crucial for conservation of these species.

Horses in particular form part of the *Equus* genus which demonstrate a variety of hereditary conditions (>90) and due to 48% of the horse chromosomes indicating conserved synteny with a singular human chromosome, they could be ideal model organisms for human disease (Wade *et al*., 2009). The horse genome is made up of 46% repetitive sequences, with <1% of segmental duplication, majority of which are intra-chromosomal, indicating a low number of rearrangements in recent history (Wade *et al*., 2009). The first sequenced horse genome, EquCab2, was released in 2007 at 2.33 Gb and in 2014 it was updated to include more data and contiguity, being improved to 2.41 Gb and renamed EquCab3 (Kalbfleisch *et al*., 2018). EquCab2 benefited from the use of FISH information and various sequencing algorithms and was improved with Illumina short-read sequencing data and further cytogenetics work leading up to 2014, thus it is comprehensive as out of all 32 chromosomes, only horse chromosome 6 is made up of more than one scaffold (Kalbfleisch *et al*., 2018). Additionally, the equine genome has the added advantage that it has been improved on by the same group of researchers, thereby reducing the chance of conflicting results unlike in the cattle genome assemblies.

3.1.3.3 Carnivora

Carnivores are one of the most investigated orders when studying speciation due to evolutionary fragmented and chromosomally rearranged genomic organizations of the 296 different species phylogenetically linked to an ancestor which originated approximately 55 million years ago (Graphodatsky, Trifonov & Stanyon, 2011; Nie *et al*., 2012; Hassanin *et al*., 2021). Dogs (2n = 78), part of the Canidae family and *Canis* genus, are model organisms for human disease studies due to their historically unconserved genomes and wide variety of heritable problems (Breen, Bullerdiek & Langford, 1999). As dogs were the first species to be domesticated by humans, they have been subjected to highly similar selection pressures as humans, sharing similar disorders and thus making them ideal candidates to study when compared to their highly conserved genomic counterparts (Wang *et al*., 2019a). Although cattle have undergone intense selection throughout domestication, they do not share the same selection pressures as dogs have undergone. This is due to their role in agricultural practices as food and dairy producers, whereas dog domestication was for companionship, sport, and aesthetics, living with humans' day in and day out, thereby the evolution of similar disorders. This is demonstrated chromosomally in a study by Webber and Ponting, 2005. For example, dogs and humans share similar conserved regions and genes in the subtelomeric regions of their chromosomes which are hotspots for synonymous nucleotide substitution rates (Webber & Ponting, 2005).

Dogs have the most rearranged karyotypes in their order as well as the highest diploid number (Nie *et al*., 2012), hence in previous years, identifying the various chromosomes when karyotyping has been difficult (Breen, Bullerdiek & Langford, 1999). This is especially true due to the similarities in banding patterns and size of the chromosomes (Breen, Bullerdiek & Langford, 1999). Sequencing of the dog genome first occurred in 2005 with the CanFam3.1 reference genome (7.5x) and was subsequently updated in 2014 by including more euchromatic regions (Wang *et al*., 2021). Another reference genome released was the ROS_Cfam_1.0 (GCF_014441545.1) in 2020 by The Roslin Institute. It covers 2.4 Gb of total

sequence length (56.5x) and this is the one this study is based on. In 2021, a consortium began research on a new reference genome, the Dog10K_Boxer_Tasha_1.0 based on the original CanFam3.1 dog, which is predicted to be completed by 2026 (Jagannathan *et al*., 2021).

3.1.3.4 Rodentia

The largest order is the Rodentia with >2000 different species, making up >40% of the mammalian species (Graphodatsky, Trifonov & Stanyon, 2011) originating approximately 55 million years ago; and is separated into two main infraorders of the Sciurognathi and Hystricognathi, depending on the relation between the incisors and the angle of the lower jaw (Capilla *et al*., 2016). Rodentia genomes have been significantly structurally reshuffled and understanding the dynamics of these changes would indicate how speciation has occurred throughout history (Graphodatsky, Trifonov & Stanyon, 2011; Capilla *et al*., 2016). Additionally, the order is known to have decreased recombination rates and increased rates of nucleotide substitution when compared to other Laurasiatheria (Capilla *et al*., 2016).

Mice (2n = 40) are part of the Murinae subfamily and *Mus* genus, which is the most studied genus of the Rodents diverging from humans about 90 million years ago (Kumar *et al*., 2017). Its counterpart, the rat (2n = 38) is also part of the Murinae subfamily, but part of the *Rattus* genus. Mice and rats diverged around 20 million years ago (Kumar *et al*., 2017). Like the dog and horse, the mouse is also used frequently as a model organism for disease and biology due to physiological with humans, as well as its ability to be used for genetic modification and gene editing. In 2020, both the current reference genome assemblies for mouse (GRCm39) and rat (Rrattus_CSIRO_v1) were released. The mouse genome assembly (2.7 Gb) is substantially bigger than the rat (2.4 Gb) and has one of the highest contig N50 (59.5x) and qualities (99.5% complete) than any of the other mammalian assembled genomes. The rat in comparison is only 1.6x contig N50 and is 96.4% complete. As the mouse is used extensively in research, this could be the reason it is of such high quality when compared to the rat.

3.1.3.5 Primate

Primates originated around 85-90 million years ago with the original ancestral primate karyotype containing 2n = 50 (Stanyon *et al*., 2008). The order contains two main branches, the Haplorrhini and Strepsirrhini, of which humans form part of the former (Stanyon *et al*., 2008). Hominoidea superorder primates, an ancestor of humans from approximately 10 million years ago, have syntenies formed by a fusion of chimpanzee chromosomes 14/15 to create the homo sapiens chromosome 2 (2n = 48), and are part of the “great apes'' including other species such as chimpanzees and gorillas (Stanyon *et al*., 2008).

In 2000, the initial human genome sequence assembly was released, only covering the euchromatic fractions of the genome (Lander *et al.*, 2001). Since then, the most recent reference genome publication was in 2013, with the most recent patch in 2022, GRCh38.p14, by the Genome Reference Consortium (Nurk *et al*., 2022). It is 3.1 Gb in size with a contig N50 of 57.9x and is 99.2% complete, less than the well-studied mouse. Unlike most genomes which originated through the use of shotgun sequencing, the human genome was assembled with the use of BACs, which were isolated, mapped and then sequenced to different parts of the genome. Using this technique, the results were more contiguous and thus accurate, yet the costs associated with it were immense and repetitive sequences were underrepresented (Liu *et al*., 2009; Nurk *et al*., 2022). Regardless, the use of BACs in genome sequence assembly has been invaluable and their use cannot be underestimated.

3.1.3.5 Diprotodontia

Common brushtail possum (2n = 20) are part of the Diprotodontia order of the Marsupial clade. They are the only mammal in this study, which is not part of the Eutherian clade, allowing further investigation to identify any regions of similarity dating back to when mammals first branched into two different clades. Marsupials all originate from a singular ancestor (2n = 14) and the largest order within this clade is the Diprotodontia, with over 125

different species (Westerman, Meredith & Springer, 2010). Even though the genomic size of marsupials is comparable to eutherian mammals, they have much smaller diploid numbers, with significantly larger chromosomes, additional to the largely conserved and stable 18 autosomal and X sex chromosomal "blocks" which are seen in most marsupials (Rens *et al*., 1999; Westerman, Meredith & Springer, 2010). Possums are the only Phalangeridae species in its family which has had both cross-species chromosome painting (through flow sorting chromosomes) analysis and same species chromosome painting conducted on it. As little is known about this family of marsupials, any additional findings could be useful for future research of this marsupial family (Rens *et al.*, 1999).

The most recent possum genome, mTriVul1.pri, was sequenced and assembled in 2020 at 3.4 Gb (the largest of all the assemblies in this study) and a 4.3x contig N50 contiguity with 96.8% of it completed. Unlike most Eutherian mammals, the possum X sex chromosome is the smallest chromosome within a karyotype.

3.1.4 Specific aims

Specific aim 1 was to assess a panel of 48 sequence- and conservation score-based homologous cattle BAC probes in both phylogenetically similar and more distantly related mammals for cross species hybridization patterns. This is partly to observe evolutionary differences between species and partly to assess the potential for these cattle probes to be used to assess chromosome abnormalities affecting fertility in other species.

- **Specific aim 1a.** To confirm that the 48 cattle BAC probes hybridise to cattle chromosomes as well as to four similar "within Order" mammals
- **Specific aim 1b.** To evaluate the degree of hybridisation of the 48 BAC probes on the chromosomes of six other mammals not within Order

- **Specific aim 1c.** To create and map ideograms of the evolutionary changes between the chromosomes of the species studied determine the FLPter measurements of each probe on the individual chromosomes
- **Specific aim 1d**. To investigate whether common genes of economic and agricultural importance are found within the highly conserved areas of the 48 BAC probes through bioinformatic alignment and ideogram evaluation in cattle

3.2 Results

3.2.1 Specific aim 1a. To confirm that the 48 cattle BAC probes hybridise to cattle chromosomes as well as to four similar "within order" mammals

Initial quality control checks completed in cattle, whereby the probes being tested were paired with BACs known to hybridise to the correct and same assumed chromosome, indicated 42 out of the 48 BACs hybridised to the correct assumed locations, yet three BACs did not hybridise to any chromosomes (CH240-196L19, CH240-182G15 and CH240-288F24). Of the remaining three BACs, two did not hybridise to the assumed correct chromosomes (CH240-244I9, CH240-124I9) and one was centromeric universal (CH240-97L3). Even though six cattle clones did not hybridise to any chromosomes, even after pairing them with different probes which did work and still no signal was seen, they were not excluded from the study.

After verification in the cattle, the other four Artiodactyl species were tested. Two of the cattle probes did not hybridise on any Artiodactyl chromosomes (CH240-196L19 and CH240-97L3). CH240-124I9 hybridised to two different pairs of cattle chromosomes yet was only on one small acrocentric chromosome in both sheep and barasingha indicating a possible fission of this lowly conserved (Conservation score = 3.743) region in cattle. Table 3-2 shows the hybridisation success rate results. Three probes did not hybridise in bison, seven BACs on sheep, 10 on barasingha and 30 on pig. Overall, the hybridisation success in the pig was significantly lower than the other four Artiodactyl species when compared using a Chi-squared

test (37.500%, $p < 0.0001$), which could be due to an excessive quantity of cell debris in the sample, thereby causing disturbance with hybridisation. Figure 3-2 demonstrates a successful hybridisation between cattle chromosomes and sheep chromosomes.

	Cattle **(QC)**	**Bison**	**Sheep**	**Barasingha**	**Pig**
Divergence (MYA)	0	4.88	24.80	27.30	62.00
Diploid number (2n)	60	60	54	56	38
Hybridisation success (%)	91.7%	93.8%	85.4%	72.9%	37.5%

Table 3-2 Species divergence (millions of years apart, MYA) from *Bos taurus*, the diploid chromosome number (2n) and hybridisation success rate (%) from the 48 bovine BACs, is stipulated along with the different within order (Artiodactyl) species tested.

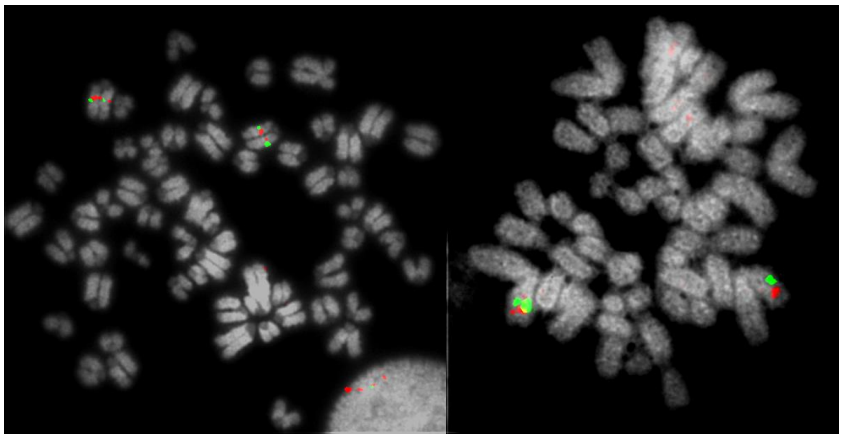


Figure 3-2 Cattle chromosome 3 probes CH240- 474H7 (FITC) and CH240-465O11 (Texas Red) hybridising to cattle chromosome 3 (left) and sheep chromosome 1 (right) (Magnification x1000) (Image source: Own images).

The Pearson correlation coefficient between the divergence time and hybridisation rate of the Artiodactyls, was found to be a strong negative significant correlation of -0.963 as the R value is larger than 0.7 ($p = 0.008$). This indicates as divergence time increases, hybridisation success rate decreases. It was noted that hybridisation success is highly dependent on the divergence time based on a coefficient of determination value of 0.927. Interestingly, there was an identical, yet positive, significant Pearson correlation coefficient between hybridisation success and 2n of 0.963 ($p = 0.008$). Again, there is a connection between the two, as when one increases, as does the other. There was also a very low correlation between the conservation

score (CS) of the BAC probes and their hybridisation success across the five Artiodactyl species (R = 0.244, p = 0.095).

Based on these preliminary results, it is hypothesised that an Artiodactyl based hybridising cattle BAC set is highly promising. Further analysis on out of order species with the cattle BACs set was then decided for subsequent work. Table 3-3 indicate the raw results obtained by microscopy. These results include the BAC ID, chromosome number of each BAC, whether the BAC was proximal or distal to the centromere, its conservation score and whether it worked (Yes) or didn't work (No) on the Artiodactyls analysed. Additional information includes whether the BAC hybridised to the correct chromosome or not (Wrong chr such as CH240-244I9) or if the probe was considered "universal" (CH240-97L3) in its hybridisation during quality control.

BAC ID	Chr	p/d	CS	QC on Cattle	Sheep	Bison	Barasingha	Pig
CH240-475L23	1	p	8.430	Yes	Yes	Yes	Yes	Yes
CH240-377G11	1	d	4.117	Yes	Yes	Yes	Yes	Yes
CH240-420D19	2	p	22.674	Yes	Yes	Yes	No	Yes
CH240-244I9	2	d	20.66	Wrong chr	Yes	Yes	No	No
CH240-386C22	2	p	7.572	Yes	Yes	Yes	Yes	No
CH240-196L19	2	d	5.399	No	No	No	No	No
CH240-514B6	2	p	3.730	Yes	Yes	Yes	Yes	No
CH240-465O11	3	d	7.060	Yes	Yes	Yes	Yes	Yes
CH240-474H7	3	p	6.150	Yes	Yes	Yes	Yes	Yes
CH240-288K11	3	d	6.100	Yes	Yes	Yes	Yes	Yes
CH240-297K13	3	p	3.957	Yes	Yes	Yes	Yes	No
CH240-379P12	3	d	3.833	Yes	No	Yes	Yes	Yes
CH240-60H16	4	p	4.897	Yes	Yes	Yes	Yes	No
CH240-339P15	5	d	4.050	Yes	Yes	Yes	No	Yes
CH240-124I9	6	p	3.743	Yes	Yes	No	Yes	No
CH240-88P10	8	d	5.330	Yes	Yes	Yes	Yes	Yes
CH240-18F3	8	p	4.750	Yes	Yes	Yes	Yes	Yes
CH240-182G15	8	d	4.518	No	Yes	Yes	Yes	No
CH240-341J24	9	p	4.327	Yes	Yes	Yes	Yes	No
CH240-256G3	11	d	4.050	Yes	Yes	Yes	Yes	Yes
CH240-258M12	11	p	8.768	Yes	Yes	Yes	Yes	Yes

CH240-288F24	11	d	5.750	No	Yes	Yes	Yes	No
CH240-329H2	12	p	5.106	Yes	Yes	Yes	No	Yes
CH240-402O18	14	p	4.448	Yes	Yes	Yes	Yes	Yes
CH240-396P6	14	d	6.996	Yes	Yes	Yes	Yes	Yes
CH240-208E13	16	d	6.522	Yes	Yes	Yes	Yes	No
CH240-26E21	17	p	5.957	Yes	No	Yes	No	No
CH240-339M3	18	d	4.263	Yes	Yes	Yes	Yes	No
CH240-333I3	19	p	5.708	Yes	Yes	Yes	Yes	No
CH240-171A7	19	d	4.550	Yes	Yes	Yes	Yes	No
CH240-97L3	19	p	3.831	Universal	No	No	No	No
CH240-380F23	21	p	9.781	Yes	Yes	Yes	No	No
CH240-344K23	21	d	8.669	Yes	Yes	Yes	No	No
CH240-124B16	22	d	7.498	Yes	Yes	No	No	No
CH240-390I12	23	p	5.038	Yes	No	Yes	No	No
CH240-33I13	24	d	8.270	Yes	Yes	Yes	Yes	No
CH240-305N4	24	p	4.950	Yes	Yes	Yes	Yes	No
CH240-451P4	25	d	5.572	Yes	No	Yes	Yes	Yes
CH240-368N15	26	d	6.739	Yes	Yes	Yes	Yes	No
CH240-244D2	26	p	5.952	Yes	Yes	Yes	Yes	No
CH240-224G7	26	p	5.071	Yes	Yes	Yes	Yes	No
CH240-457O14	27	d	3.848	Yes	Yes	Yes	Yes	Yes
CH240-236P3	28	p	4.475	Yes	Yes	Yes	Yes	No
CH240-226K16	29	d	6.355	Yes	Yes	Yes	Yes	Yes
CH240-128C9	X	p	5.600	Yes	No	Yes	Yes	No
CH240-29N7	X	d	4.296	Yes	Yes	Yes	Yes	No
CH240-359O3	X	p	4.148	Yes	Yes	Yes	Yes	No
CH240-48F6	X	d	4.107	Yes	No	Yes	Yes	No

Table 3-3 Successful hybridisations using highly conserved bovine BACs on Artiodactyl species. The table contains the bovine BAC ID's, assumed chromosomal location (pre-quality control), location of proximal (p) (FITC) or distal (d) (Texas Red) sequence to the centromere; and conservation score (CS). Species include: the quality control (QC) on Cattle; and normal within-order (Artiodactyl) FISH investigation on American bison; domestic sheep; barasingha; and domestic pig.

3.2.2 Specific aim 1b. To evaluate the degree of hybridisation of the 48 BAC probes on the chromosomes of six other mammals not within Order

The cattle BACs were hybridised to 6 phylogenetically different mammals. The success rate varied from 25% (possum) to 72.9% (horse) and was once again highly positively, yet not

significantly, correlated with species diploid number (R = 0.752, p = 0.085) and highly negatively and not significantly, associated with the divergence time (R = -0.724, p = 0.104) (Table 3-4, Figure 3-3 and Table 3-5). These are slightly lower results than those noted within the Artiodactyl order, and none of the results are significant. As the horse and dog species are also part of the Laurasiatheria superorder, along with the Artiodactyls, and their hybridisation success rates were moderate to high, 64.6% and 72.9%, respectively, a Laurasiatheria superorder panel of BACs could be identified for use. Unlike the dog and horse, the mouse, rat, and possum had low hybridisation success with the cattle BACs, and thus the use of these probes on Marsupials or Euarchontoglires could be limited. Although this is the case, within the Euarchontoglires, humans had an identical success rate to the dog of 64.6%, thus further investigation into the Primate branch could be useful.

	Horse	Dog	Human	Rat	Mouse	Possum
Divergence (MYA)	78.00	78.00	96.00	96.00	96.00	159.00
Diploid number (2n)	64	78	46	42	40	40
Hybridisation success (%)	72.9%	64.6%	64.6%	39.6%	29.1%	25.0%

Table 3-4 Species divergence (millions of years apart, MYA) from *Bos taurus*, diploid chromosome number (2n) and their respective overall hybridisation success rates (%) from the 48 bovine BACs tested, along with different out-group order animals (non-Artiodactyl).

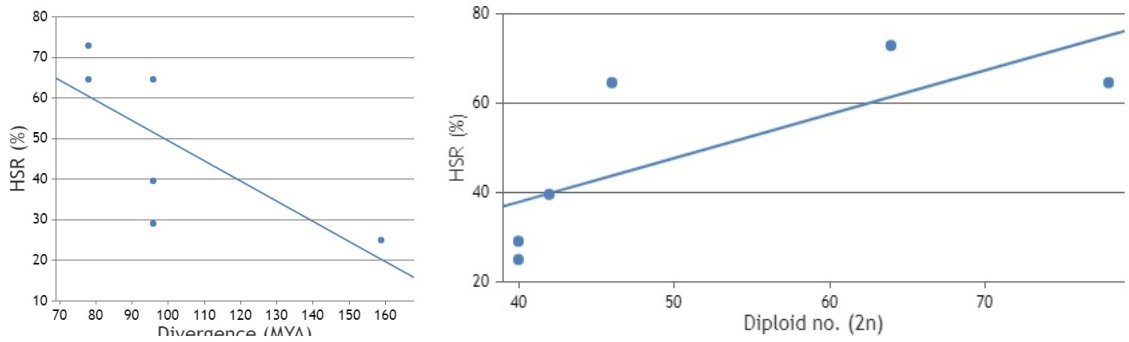


Figure 3-3 Scatter plots of the (A) Divergence time (millions of years apart, MYA) of each of the species from cattle (x axis) versus their respective hybridisation success rates (HSR) (%) (y axis); and (B) the species' diploid numbers (no.) (2n) (x axis) versus their respective hybridisation success rates (HSR) (%) (y axis). The line of best fit indicated the correlation between the two variables. Each species is represented by a point on the graph.

BAC ID	Chr	p/d	CS	Horse	Dog	Mouse	Rat	Huma	Possu

								n	m
CH240-475L23	1	p	8.430	Yes	No	No	Yes	No	No
CH240-377G11	1	d	4.117	No	No	No	No	No	No
CH240-420D19	2	p	22.674	Yes	No	Yes	Yes	Yes	Yes
CH240-244I9	2	d	20.66	No	Yes	No	No	Yes	No
CH240-386C22	2	p	7.572	Yes	Yes	Yes	No	Yes	Yes
CH240-196L19	2	d	5.399	No	No	No	No	No	No
CH240-514B6	2	p	3.730	Yes	No	No	No	Yes	No
CH240-465O11	3	d	7.060	No	No	No	Yes	Yes	No
CH240-474H7	3	p	6.150	Yes	No	Yes	Yes	Yes	Yes
CH240-288K11	3	d	6.100	Yes	No	Yes	Yes	Yes	Yes
CH240-297K13	3	p	3.957	No	No	No	Yes	Yes	No
CH240-379P12	3	d	3.833	Yes	No	No	Yes	No	No
CH240-60H16	4	p	4.897	Yes	No	No	No	No	No
CH240-339P15	5	d	4.050	Yes	Yes	No	Yes	No	No
CH240-124I9	6	p	3.743	No	No	No	No	No	Yes
CH240-88P10	8	d	5.330	Yes	Yes	Yes	Yes	Yes	No
CH240-18F3	8	p	4.750	Yes	Yes	No	Yes	Yes	No
CH240-182G15	8	d	4.518	No	No	No	No	No	No
CH240-341J24	9	p	4.327	Yes	Yes	No	No	No	No
CH240-256G3	11	d	4.050	Yes	Yes	Yes	Yes	Yes	Yes
CH240-258M12	11	p	8.768	Yes	Yes	Yes	No	Yes	Yes
CH240-288F24	11	d	5.750	Yes	Yes	No	Yes	Yes	No
CH240-329H2	12	p	5.106	Yes	Yes	No	No	Yes	No
CH240-402O18	14	p	4.448	Yes	No	No	Yes	Yes	No
CH240-396P6	14	d	6.996	Yes	Yes	No	No	Yes	Yes
CH240-208E13	16	d	6.522	Yes	Yes	No	Yes	Yes	No
CH240-26E21	17	p	5.957	Yes	Yes	No	Yes	No	No
CH240-339M3	18	d	4.263	Yes	Yes	No	No	No	No
CH240-333I3	19	p	5.708	Yes	Yes	Yes	No	Yes	No
CH240-171A7	19	d	4.550	Yes	Yes	Yes	No	Yes	No
CH240-97L3	19	p	3.831	No	No	No	No	No	No
CH240-380F23	21	p	9.781	Yes	Yes	Yes	No	Yes	No
CH240-344K23	21	d	8.669	Yes	Yes	No	No	Yes	No
CH240-124B16	22	d	7.498	Yes	No	Yes	No	Yes	Yes
CH240-390I12	23	p	5.038	Yes	Yes	No	No	Yes	No
CH240-33I13	24	d	8.270	Yes	Yes	No	No	Yes	No
CH240-305N4	24	p	4.950	No	No	No	No	No	No
CH240-451P4	25	d	5.572	Yes	Yes	No	No	No	No
CH240-368N15	26	d	6.739	Yes	Yes	No	Yes	Yes	Yes
CH240-244D2	26	p	5.952	No	Yes	No	No	Yes	No
CH240-224G7	26	p	5.071	No	Yes	No	No	Yes	No
CH240-457O14	27	d	3.848	Yes	Yes	No	Yes	No	No

CH240-236P3	28	p	4.475	Yes	Yes	No	No	Yes	Yes
CH240-226K16	29	d	6.355	Yes	Yes	Yes	Yes	Yes	No
CH240-128C9	X	p	5.600	Yes	Yes	Yes	No	Yes	No
CH240-29N7	X	d	4.296	Yes	Yes	Yes	No	Yes	No
CH240-359O3	X	p	4.148	No	Yes	No	No	No	Yes
CH240-48F6	X	d	4.107	No	Yes	No	No	No	No

Table 3-5 Successful hybridisations using highly conserved bovine BACs on non-Artiodactyl order, Perissodactyla (horse), Carnivora (dog), Rodentia (mouse and rat), Primates (human) and Diprotodontia (possum) species. The table contains the bovine BAC ID's, assumed chromosomal location (pre-quality control), location of the proximal (p) (FITC) or distal (d) (Texas Red) BAC sequence; and conservation score (CS).

Divergence time and hybridisation rate of all species was found to be a strong negative correlation of R = -0.716 and this was not a significant result in this study (p = 0.07) (Table 3-6 and Figure 3-3). Although hybridisation was weak between highly divergent species and cattle, as with the pig sample, the rat, mouse, and possum samples were of lower quality due to a higher concentration of cellular debris, which inadvertently decreased the likelihood of optimal hybridisation of the cattle BACs to metaphases and hence visualisation of the results. As mentioned in section 2.3, all experiments were repeated a minimum of three times on the same sample to validate results. For rat, mouse, and possum, only one of each sample had a mitotic index high enough to use for these experiments and thus the hybridisations could not be repeated on separate samples if there was too much cellular debris.

X variable	Y variable	Order tested	R	Significance
Hybridisation success	Divergence time	Artiodactyl only	-0.963	0.008*
		Non-Artiodactyl	-0.724	0.104
		All orders	-0.716	0.070
Hybridisation success	Diploid number	Artiodactyl only	0.963	0.008*
		Non-Artiodactyl	0.752	0.085
		All orders	0.406	0.366

Table 3-6 The calculated Pearson correlation coefficients (R) for each of the variables (x and y) are indicated, as well as groups tested against one another. *Significant at a p < 0.05.

One probe, CH240-256G3 (cattle chromosome 11), worked across all 11 species. This BAC had a below average conservation score value (Conservation score, CS) of 4.050. Three BACs, CH240-196L19, CH240-305N4 and CH240-97L3, did not work on any of the non-Artiodactyl chromosomes and they had a mean CS of 4.528. Additionally, CH240-196L19 (CS = 5.399) did not work in any Artiodactyl species either, and thus failed completely; CH240-97L3, which had one of the lowest CS scores of 3.831, failed in all the species except cattle, in which it was universally centromeric; and CH240-305N4 (CS = 4.95) only worked in Artiodactyls. Examples of cross species probe hybridisations can be seen in Figure 3-4.

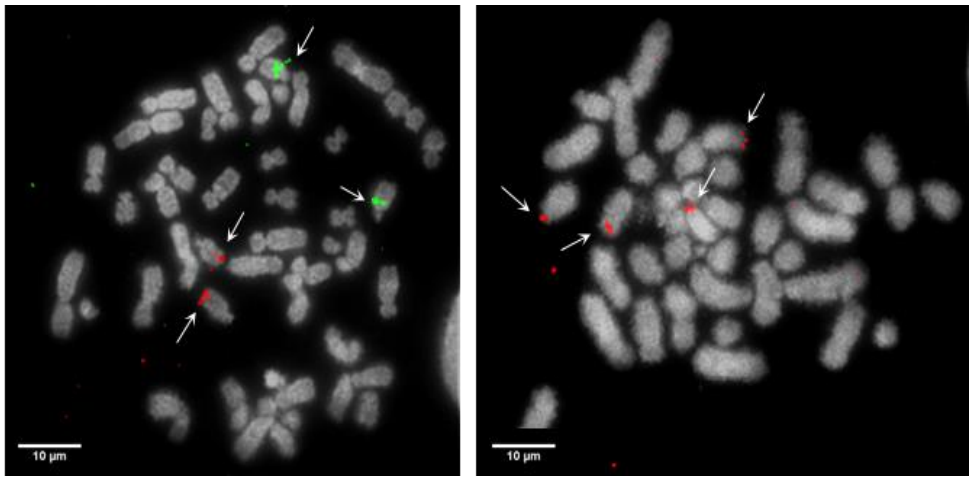


Figure 3-4 Cattle BACs CH240-380F23 (FITC) and CH240-344K23 (Texas Red) hybridising to human chromosomes; and cattle BAC clone CH240-226K16 (Texas Red) hybridising to possum chromosomes (Magnification x1000) (Image source: Own images).

The overall mean CS for the cattle BACs was 6.2, with a standard deviation of 3.606, median of 5.218, minimum of 3.73 and maximum value of 22.674. Interestingly, 10 probes worked better in non-Artiodactyls than in Artiodactyls (CS average = 8.066). 12 BAC clones worked in 9 or more species (CS = 7.667) and 6 probes worked in 10 or more species (CS = 6.126). A low positive correlation between individual BAC CS values and each individual hybridization success rate was found of 0.121. A similar Pearson correlation coefficient was found for the Artiodactyl group and almost no correlation identified for the non-Artiodactyls, of 0.244 and -0.078, respectively. According to PhastCons , the higher the CS of the BAC, the more conserved it is, but this was not seen in our results.

To ascertain if the accuracy of BAC hybridisation is Order specific and thus divergence related, the results were compared between and within orders (Figure 3-5 and Table 3-7). As previously mentioned, there were 5 Artiodactyl species, 2 Rodentia and only one species from the orders of Carnivora, Perissodactyla, Diprotodontia and Primates. One or two species per order would not give a reliable or accurate estimation of conservation of the BACs within that group and further testing in this regard would be required. Regardless, the Artiodactyl average was calculated to be 76.3% and non-Artiodactyl, 49.3%, which was highly significantly different when compared using a t-test of p = 0.036. This is to be expected due to the close relation between certain species to cattle. Non-Artiodactyl species had a broader interquartile range (IQR) than the other groups and an almost 34% lower median than the Artiodactyls. As an interesting observation, if the pig value was removed from the Artiodactyl group (due to the poor hybridisation seen, possibly because of a poor sample quality due to increased cellular debris or sample age, and therefore denaturation of DNA because of acetic acid disruption), the overall order hybridisation success average increases to 86% from 76.3%. If the clade of Laurasiatheria was investigated as a wider opportunity to use the BACs, the average was found

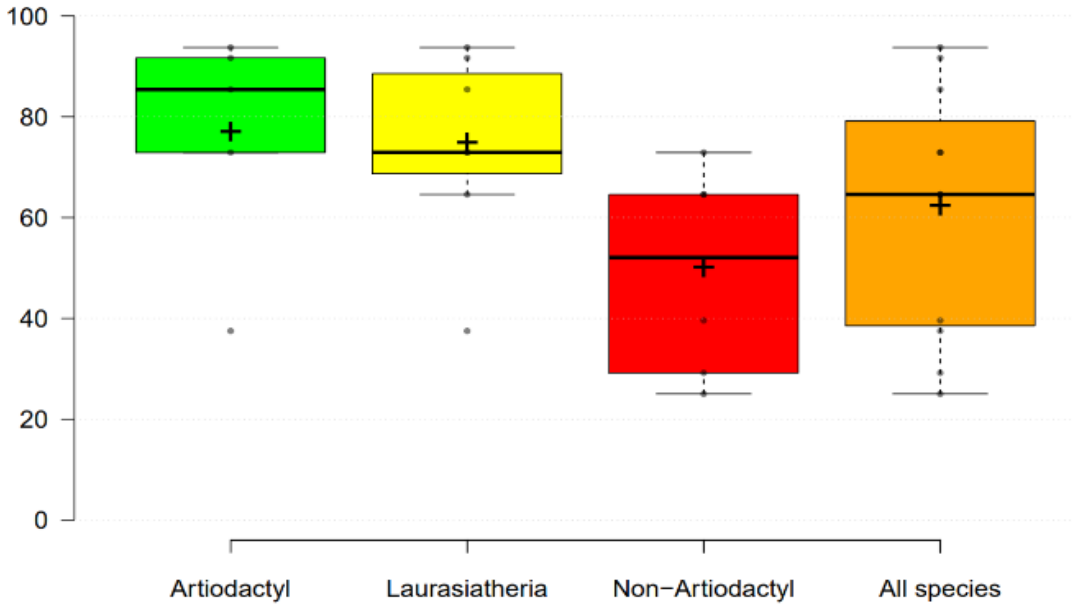


Figure 3-5 Box plots of the different groups of species studied (x axis) versus their cattle BAC clone hybridisation success rates (y axis). Mean (+), minimum, maximum, median, interquartile ranges, and data points such as outliers, are all shown for each box plot. For specific values, refer to Table 3-7.

to be 74.1%. This should be studied in more detail. Further within-branch, as well as within-mammal analysis should also be conducted to establish if there is potential to use this set of highly conserved BACs and a universally hybridising set of clones. If further investigation is done into genetic variation within individual species, it could also determine why certain probes do not hybridise well when compared to other species.

	Artiodactyl	Laurasiatheria	Non-Artiodactyl	All species
Upper whisker	93.75	93.75	72.92	93.75
3rd quartile	91.67	88.54	64.58	79.17
Median	85.42	72.92	52.08	64.58
1st quartile	72.92	68.75	29.17	38.54
Lower whisker	72.92	64.58	25.00	25.00
Nr. of data points	5	7	6	11
Mean	76.25	74.11	49.31	61.55

Table 3-7 Statistics relating to Figure 3-4 are indicated in this table, including the maximum (upper whisker), 3rd and 1st quartiles, median, minimum (lower whisker), number of data points and mean for Artiodactyl, Laurasiatheria, Non-Artiodactyl and overall species hybridisation success rates (%).

3.2.3 Specific aim 1c. To create and map ideograms of the evolutionary changes between the chromosomes of the species studied determine the FLPter measurements of each probe on the individual chromosomes

Figure 3-6 shows an Ideogram of CH240-256G3 on each of the 11 different species. The probe location on the chromosome is based on the fractional length relative to the p-terminus (FLPter) value calculated for each species. Each chromosome in the diagram is based on a calculated ratio of the chromosome full size and proximal and distal BAC sequence location from the centromere of the individual species, yet the chromosomal diagrams displayed are not in ratio to the other species' chromosomes. Only the three chromosomes for possum are in ratio to one another regarding length.

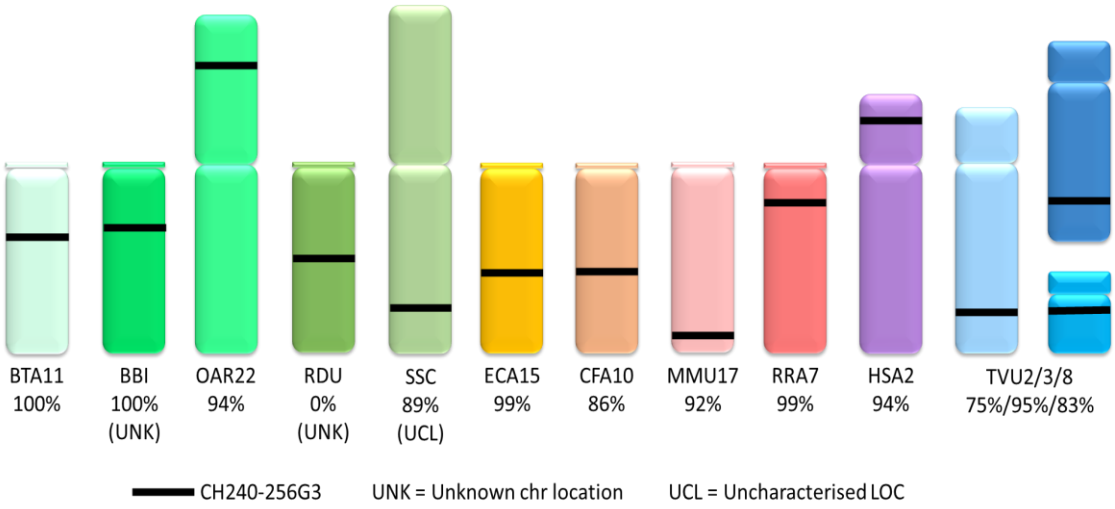


Figure 3-6 Ideogram of the most highly conserved cattle BAC, CH240-256G3, on all 11 different species with chromosomal location identified with FLPter values, and their respective National Center for Biotechnology Information (NCBI) BLASTn CH240-256G3 sequence identity (%) match, underneath each species acronym. The chromosomal sizes are not relative to one another unless there are multiple such as in the case of the possum.

Based only on FLPter values, bison and barasingha were analysed using the probes (Table 3-8). Bison had 9 statistically different ($p < 0.01$) clone placements from cattle, whereas barasingha had 10 statistically different probes for $p < 0.01$ and one for $p < 0.05$ when tested using an unpaired t test. Probes CH240-475L23, CH240-465O11, CH240-474H7, CH240-288K11, CH240-297K13, CH240-18F3, CH240-256G3, CH240-402O18, CH240-305N4, CH240-224G7 and CH240-29N7 were the only probes which were in comparable locations in all three species. CH240-386C22, CH240-341J24, CH240-380F23 and CH240-344K23 were statistically similar between cattle and bison, while CH240-258M12 and CH240-457O14 were the only two probes which had corresponding results in barasingha and cattle.

BTA chr	BAC ID	Cattle	Bison	Barasingha
1	CH240-475L23	57.464	59.980	68.453
1	CH240-377G11	73.250	77.047	89.526**
2	CH240-386C22	36.749	38.808	-
2	CH240-514B6	75.604	-	54.686**
3	CH240-465O11	79.969	85.167	79.492
3	CH240-474H7	55.999	59.183	46.705
3	CH240-288K11	74.053	77.090	76.035
3	CH240-297K13	36.723	41.865	52.680
3	CH240-379P12	46.030	-	64.230*
5	CH240-339P15	29.558	82.506**	-
6	CH240-124I9	17.031	-	39.509**
8	CH240-88P10	47.615	28.233	31.403**
8	CH240-18F3	53.154	57.909	66.616
9	CH240-341J24	43.527	53.645	-
11	CH240-256G3	36.165	31.775	47.984
11	CH240-258M12	73.755	-	69.671
14	CH240-402O18	53.042	48.535	61.291
14	CH240-396P6	64.028	83.003**	80.640**
17	CH240-26E21	23.278	48.457**	-
18	CH240-339M3	94.146	57.510**	-
19	CH240-333I3	64.146	35.480**	54.497
19	CH240-171A7	26.253	54.197**	54.497**
21	CH240-380F23	63.694	68.989	-
21	CH240-344K23	34.139	32.049	-
24	CH240-33I13	85.356	81.310*	70.041**
24	CH240-305N4	77.847	71.840	74.547
25	CH240-451P4	48.965	86.304**	48.668
26	CH240-368N15	56.827	54.696	77.651**
26	CH240-224G7	66.510	57.269	65.406
27	CH240-457O14	43.614	-	54.965
28	CH240-236P3	45.809	64.571**	54.369
X	CH240-128C9	46.080	5.405*	31.710
X	CH240-29N7	30.830	17.904	36.899
X	CH240-359O3	90.863	79.983**	53.614**
X	CH240-48F6	39.221	28.528	66.855**

Table 3-8 FLPter values calculated for each of the probes on cattle, bison and barasingha. Blank cells are due to either there being poor quality images or no signals, and thus no values were calculated for those probe-chromosome combinations. Probes removed from the table were due to there being no FLPter values for those probes or on the bison/barasingha chromosomes. *Statistically significantly different to cattle FLPter mean value at $p < 0.05$. **Statistically significantly different to cattle FLPter mean value at $p < 0.01$.

Chromosome numbers are based on BLASTn searches of the CH240-256G3 FASTA sequence against each individual species' genome, and the identity percentage is displayed under the species and chromosome number. If there was a sequence match, such as in bison (100%), yet no placed scaffold or chromosomal location, "UNK '' or unknown was stated. As in the case for barasingha, no sequence matches were found, yet as a signal was seen, this sequence should be found within the genome. Currently, there is no sequenced genome assembly for a barasingha and at the time of conducting this study, only unplaced genomic scaffold (UGS) for bison; and therefore, it is to be expected for there to be no matching BLASTn results. However, these animals were chosen not only for their phylogenetically close relationship to cattle, as well as for the conservation efforts involved in both species, and not on how complete their genomes were. For pigs, a sequence matching 89% of the clone nucleotide sequence was identified as an uncharacterised locus. Through karyotyping, it was established this location was on chromosome 4.

Using the NCBI BLASTn program, all cattle clone sequences were searched for firstly in cattle and then each of the individual species. Any identity hits above 95% were recorded or for species which did not have any hits above 95%, the next highest hit was recorded. As bison and barasingha both have no structured assemblies, they were excluded from this search. Clones on cattle chromosome 2, 11 and X were specifically searched, due to the number of conserved BACs used in this study on each of those chromosomes, more so than other chromosomes, therefore potentially conserved synteny could potentially be mapped (Figure 3-6). CH240-514B6 and CH240-258M12 FASTA sequences could not be identified, and thus were excluded from this search.

Two of the three cattle chromosome 2 probes (CH240-420D19 and CH240-244I9) were conserved on chromosome 2 specifically, in possum, as well as sheep, humans and mice, potentially indicating where these sequences were situated in an original ancestor species. CH240-244I9 and CH240-386C22 were in similar locations on cattle chromosome 2, and

consequently seem to have been inherited through evolution as a highly conserved block, as they were paired in all 9 species on different chromosomes, relatively the same distance away from one another. For example, they were found in sheep chromosome 2, pig chromosome 15, dog chromosome 36, horse chromosome 18, human chromosome 2, mouse chromosome 2, rat chromosome 5 and possum chromosome 2. CH240-256G3 and CH240-288F24 (cattle chromosome 11) were inherited together on sheep chromosome 2 and pig chromosome 1, even though they were not located close to one another, and they were found on possum chromosome 3, inverted and closely situated to one another and to the centromere. A novel finding was the placement of the highly conserved BAC CH240-256G3 in equines. There was almost a 100% identity of this clone on an UGS when using BLASTn; and through karyotyping, it was found to be located on chromosome 15, approximately 57% of the total length of the distal probe's distance from the centromere. This will need further investigation such as repeated FISH with equine specific probes to be confirmed, but it could aid bioinformaticians who are constructing and updating the equine assembly by physically mapping the location of this UGS.

All X chromosome probes (Figure 3-7) were conserved on the X sex chromosome in all Eutherian species and was found on autosomes (2, 3, 5 and 8) in the only marsupial species studied, the possum, potentially demonstrating how the sex chromosome has changed through evolution when these clades first branched out. In horses, the cattle X chromosome CH240-128C9 sequence was found on three different chromosomes, horse chromosomes 7, 12 and X. As seen in Figure 3-6, CH240-359O3 was found subtelomerically, yet on equine chromosome X, it was found in an inverted location, closer to the centromere, while CH240-128C9 and CH240-29N7 moved subtelomerically. CH240-48F6 remained in a similar location.

As certain BAC clone regions were highly conserved, which could also be related to their sequence length, further investigation into the genes within these regions as well as their purpose would subsequently follow, potentially in future studies with sequencing techniques

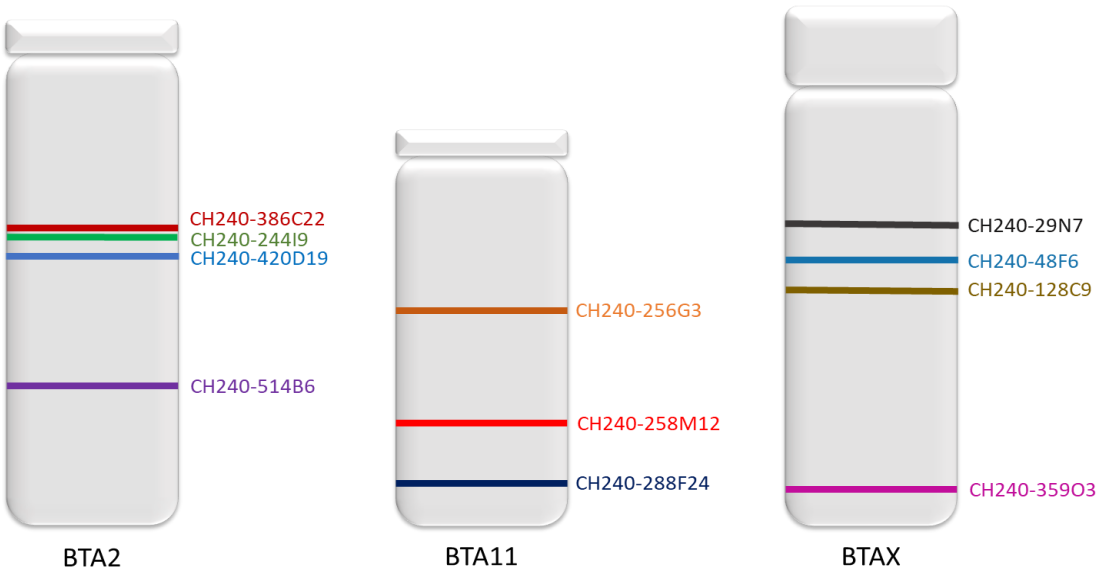


Figure 3-7 Cattle chromosomes 2, 11 and X are displayed with their relative probe FLPter values and locations. Chromosomes are in relative size to one another, and clones are shown by their different colours.

in both the DNA (for BAC length and sequence confirmation), and RNA (for further gene expression analysis). This type of comparison would further allow haplotype or homologous blocks which were previously in unknown locations or in predicted regions, to be further proven to be found in certain highly conserved areas, thereby aiding assembly efforts. Additionally, it would allow an insight into specifically what type of genes are found in mammalian species routing back to the Eutherian clade.

3.2.4 Specific aim 1d. To investigate whether common genes of economic and agricultural importance are found within the highly conserved areas of the 48 BAC probes through bioinformatic alignment and ideogram evaluation in cattle

Many of the BACs had highly conserved genes in their sequences which were found in each of the different species. N = 49 genes were found to be in highly conserved clone regions through the National Center for Biotechnology Information (NCBI) BLASTn research, and these were

found to have orthologues in sheep, horses, dogs, humans, and mice. Consequently, these species were studied further for this aim.

Certain clones including CH240-288F24 (n = 4), CH240-26E21 (n = 2), CH240-244D2 (n = 2), CH240-236P3 (n = 7), CH240-226K16 (n = 2), CH240-128C9 (n = 2), CH240-359O3 (n = 2) and CH240-48F6 (n = 2) had multiple genes situated within their sequence regions (Table 3-9). Table 3-9 also demonstrates orthologous genes found in the various species based on their specific BAC clone sequence hybridisation. This information was confirmed with BLAST searches and fractional length relative to the p-terminus (FLPter) as well RefSeq information. Further testing would involve sequencing and/or karyotyping to confirm these results. Genes were found within the BAC regions of CH240-420D19 (Zinc finger E-box-binding homeobox 2, ZEB2, cattle chromosome 2), CH240-339P15 (Constitutive photomorphogenesis 9 Signalosome Subunit 7A, COPS7A, cattle chromosome 5), CH240-329H2 (Protocadherin 9, PCDH9, cattle chromosome 12), CH240-368N15 (Paired box 2, PAX2, cattle chromosome 26) and CH240-128C9 (Non-receptor tyrosine kinase homology 2 domain containing 1A, SH2D1A and Teneurin transmembrane protein 1, TENM1, cattle chromosome X) in cattle, sheep, horse, dog, human and mouse genomes as orthologues and were confirmed with FLPter locations (<20% difference between the gene chromosomal location and the measurement) and the gene specified chromosome matched the BLAST chromosome; or if the gene specific chromosome matched the BLAST chromosome and there is no/unclear FLPter value (Top section of Table 3-9). 15 genes were identified in the regions of CH240-386C22, CH240-88P10, CH240-256G3, CH240-258M12, CH240-339M3, CH240-380F23, CH240-244K23, CH240-33I13 and CH240-236P3 when the gene chromosomes matched the BLAST chromosome searches and there was no/unclear FLPter value or if the FLPter value has a difference of <30% between the chromosome the gene was on and the measurement. These are in the lower part of Table 3-9. Chromosome X was the most conserved chromosome specifically regarding gene locations in Eutherian mammals, while chromosome 2 was also highly conserved regarding gene locations.

Clone	Gene	Cattle	Sheep	Horse	Dog	Human	Mouse
CH240-420D19	ZEB2	2	2	18	19	2	2
CH240-339P15	COPS7A	5	3	6	27	12	6
CH240-329H2	PCDH9	12	10	17	22	13	14
CH240-368N15	PAX2	26	22	1	28	10	19
CH240-128C9	SH2D1A	X	X	X	X	X	X
	TENM1	X	X	X	X	X	X
CH240-386C22	PSMD14	2	2	18	36	2	2
CH240-88P10	BNC2	8	2	23	11	9	4
CH240-256G3	CAMKMT	11	3	15	10	2	17
CH240-258M12	EHBP1	11	3	15	10	2	11
CH240-339M3	SALL1	18	14	3	2	16	8
CH240-380F23	NPAS3	21	18	1	8	14	12
CH240-344K23	NR2F2	21	18	1	3	15	7
CH24033I13	ZNF521	24	23	8	7	18	18
CH240-236P3	USP54	28	25	1	4	10	14
	SEC24C						
	MYOZ1						
	CAMK2G						
	NDST2						
	CHCHD1						
	FUT11						

Table 3-9 Orthologous genes found in various species based on specific BAC clone sequence hybridisation. Chromosomal gene placement is based on the location given by the National Center for Biotechnology Information (NCBI) search results and not the BLASTn results. The first 6 genes were found in all species, and the last part of the table are genes which were significantly different from the cattle results based on FLPter values, or due to the BLAST sequence chromosomal output not matching the NCBI gene chromosomal result for a specific species.

When certain genes were found in species as orthologues, they were indicated to be on a specific chromosome. When the clone associated with that sequence was placed through

BLASTn results, on a few occasions, they did not match. For example, the EH domain binding protein (EHBP1) which plays a role in endocytic trafficking, specifically causing an aggressive form of prostate cancer, is a model gene found on chromosome 15 in horses, yet according to BLASTn results, the cattle clone containing this gene in cattle, CH240-258M12, is found on equine chromosome 25. This could indicate either a possibly incorrect placement of the gene when sequence assembly was conducted, especially if the gene is only a model or predicted gene, or the clone does not contain this gene in that species. This could potentially be work for future studies. Even though certain BAC sequences were selected for this study, it does not mean they have been placed in the cattle assembly correctly, which is further reasoning for physical mapping of the sequences and studying the genes related to these sequences. Four genes were RefSeq "Validated" in cattle (SLC6A17, RAPBAP1, COPS7A and ZEB2), none in horse, one in sheep (HDAC9), one in dogs (CCS), 11 in humans (TLE4, RC3H2, ZBTB6, RABGAP1, TTC29, ZZNF521, USP54, CHCHD1, FUT11, RBM4 and KLHL34) and almost all (n = 47) in mouse (excluding PCDH9 and FUT11). Only humans (n = 38) and mice (n = 11) orthologs have been "Reviewed" in total. If a gene has been validated or especially reviewed by the NCBI for a particular species, and there was no FLPter value (due to either no signal or unclear signals), the BLAST gene location of the clone is definite. If the FLPter location and the reviewed gene BLAST location are not in the same place, it can be assumed there was an error in labelling the clone or in its hybridisation in the experiment.

Table 3-10 demonstrates how many of the studied genes found in the cattle BAC clone locations, were either provisional, model, inferred, reviewed, or validated. Provisional genes have not undergone individual review and have been provided by outside references or collaborators (Pruitt *et al.*, 2020). The majority of cattle genes are only provisional genes. Model genes are provided by the NCBI Genome Annotation pipeline and have not undergone individual review or alteration between different annotation repeats (Pruitt *et al.*, 2020). Sheep, horses, and dogs all have primarily model genes to reference. Inferred genes have been predicted by various sequencing analyses yet have no experimental data to support them

(Pruitt *et al*., 2020) and only two of these genes were found in cattle BAC clone sequences. Reviewed genes have been assessed by a member of NCBI staff or collaborators, thereby have been subject to extensive literature and data research (Pruitt *et al*., 2020). 38 human and 11 mice orthologs have undergone this process. Lastly, validated genes have undergone an initial review, yet have not been finalised at the time functional information was provided (Pruitt *et al*., 2020); and mice have 47 validated orthologous genes to the ones identified in the cattle BAC sequences.

Species	Provisional	Model	Inferred	Reviewed	Validated
Cattle	**28**	15	2	0	4
Sheep	2	**46**	0	0	1
Horse	1	**48**	0	0	0
Dog	1	**47**	0	0	1
Human	0	0	0	**38**	11
Mouse	1	0	0	1	**47**

Table 3-10 Orthologous genes found in the cattle BACs were investigated in various species. The number of provisional, model, inferred, reviewed, and validated orthologous genes per species are shown.

When the genes were further investigated, trends of specific gene functions began to arise. The genes found in the cattle BAC clone regions are primarily transcriptional regulator/repressor protein genes (n = 20), such as the paired box 2 (PAX2, BTA26) and zinc finger E-box binding homeobox 2 (ZEB2, cattle chromosome 2). The other four highly conserved genes found, include the constitutive photomorphogenic homolog subunit 7A (COPS7A) which is part of the COP9 Signalosome Complex (BTA5), teneurin transmembrane protein 1 (TENM1, cattle chromosome X), protocadherin 9 (PCDH9, cattle chromosome 12) and SH2 domain containing 1A (SH2D1A, cattle chromosome X). Five of these genes are involved in neuronal development (PAX2, ZEB2, COPS9, TENM1 and PCDH9) while SH2D1A is a causative recessive gene for X-linked lymphoproliferative disease type 1 (XLP1) which is known to cause brain lesions in patients.

3.3 Discussion

3.3.1 Evolutionary changes tracked with cattle BACs

The results found in this section of work were based on 11 species, tested with 48 highly conserved cattle BACs, 5 species of which were part of the Artiodactyl order, 2 were Rodentia and one from each of the Perissodactyla, Carnivora, Primate and Diprotodontia orders. The overall average hybridisation rate was 61.6%, compared to the Artiodactyl order and non-Artiodactyl orders means, of 76.3% and 49.3%, respectively. As cattle form part of the Artiodactyla order, it was an expected result to have a high hybridisation success rate from within order species. If dogs (Carnivora) and horses (Perissodactyla) were included in this grouping, as a Laurasiatheria clade category, the hybridisation success rate only lowered to 74.1%, yet this was not significant, indicating a potential avenue to further analyse the cattle BAC set. Within the Artiodactyls however, the pig had a poor hybridisation rate (37.5%) and as $p < 0.0001$ which could indicate a sample issue, potentially due to suboptimal chromosome preparations, which caused poor hybridisation which is a problem often experienced with culture failure (Schaeffer *et al*., 2004). Ideally samples of poor quality should be repeated with a different sample from the same species, or even better the same animal, yet this was not conducted due to COVID-19 disruptions, such as the domestic pig, or the lack of a second sample for certain species such as the House mouse, Black rat and Common Brushtail Possum. Many of the samples were over 10 years old, and consequently had been stored in old fixative, therefore increasing the possibility for DNA denaturation due to methanol evaporation and acetic acid exposure. Four of the 5 Artiodactyl species fall under the Pecora suborder, while pigs do not, and this could potentially be another reason for the poor success rate. However, if this was the case, horses and dogs would also have had a poor hybridisation rate which was not the case. If the pig sample had been removed from analyses, the hybridisation success mean rate would have increased from 76.3% to 86.0% for Artiodactyla, and up to 77.3% for

Laurasiatheria. This species would need to be repeated with another sample in a future experiment.

Mouse, rat, and possum samples were also of poor quality prior to analyses, and this could have inadvertently lowered the hybridisation rate, similar to the pig. All samples were checked for their mitotic index prior to analysis, yet the quality of the chromosomes is not the only indication of a good sample, as there could have been a lot of debris surrounding the metaphases (Wolff *et al*., 2007). Therefore, the highly conserved BACs could not attach themselves to sequences. Another potential reason for the reduced success of the probes in rats and mice, other than the high divergence time of 96 million years apart (MYA) (Kumar *et al*., 2022), could be due to the increased rate of genome rearrangements and reshuffling which occurred in Rodentia evolution, specifically surround evolutionary breakpoint regions (EBRs) (Capilla *et al*., 2016). The high recombination rates, and restructuring of gene content, especially protein-coding genes (Capilla *et al*., 2016), could have disrupted the ability for the probes to hybridise to a sequence fully, hence producing "busy" image visualisations with a lot of excess probe surrounding the metaphases, yet not on them. Possums are 196 MYA diverged from cattle (Kumar *et al*., 2022) and thus their lack of hybridisation could purely be to do with their high divergence time and completely different evolutionary clade. Considering this, a 25% hybridisation success rate between these mammals is, according to our knowledge, the only reported value between these two species regarding cattle BAC comparative studies. The Bovine Genome Sequencing and Analysis Consortium (2009) reported 91.5% of orthologs found in the cattle genome are detected in marsupials, thus there is one of many potential avenues for further investigation.

When investigating the correlation between hybridisation success rate and divergence time within Artiodactyls, there was a strong, significant, negative correlation of $R = -0.963$ ($p = 0.008$) which is expected based on the molecular clock hypothesis (Ho, 2008; Ho, 2020). It states that protein and DNA sequences change at a relatively constant rate over time between

organisms and the divergence time between any two species' common ancestor is directly proportional to the genetic differences noted between them (Ho, 2020). As the correlation coefficient is very high, and the probes are highly conserved around EBRs which are gene and protein rich, this could be a significant example of the relationship between divergence time and conservation of genes (Larkin *et al*., 2009). Similarly, in non-Artiodactyls, the correlation between divergence time and BAC hybridisation rate was still considered high (R = -0.724) yet not significant (p = 0.104). Based on the results, the overall hybridisation rate was highly correlated with divergence time (R = -0.84, p = 0.001) validating the molecular clock hypothesis (Ho, 2008; Ho, 2020). Contrary to predictions, individual BAC CS values were lowly positively correlated to hybridisation of these probes (R = 0.121) which is not what was anticipated when compared to previous results achieved in this research. This indicates there is another factor involved with the potential of a BAC clone to hybridise to homologous sequences in different species, unrelated to conservation score, such as gene or sequence variation (Warnefors and Kaessmann, 2013) as divergence increases.

Another interesting finding, which has opposing views in the literature, is the significant positive moderate correlation between chromosome diploid number (2n) and hybridisation success (R = 0.691, p = 0.019). Graphodatsky, Trifonov & Stanyon (2011) found a negative moderate correlation between these two variables (R = -0.404) specifically when looking at conserved segments between different species and humans. As there are fewer chromosomes in certain mammals compared to others, we assume this statistic could purely be a there are a reduced number of chances for the probes to hybridise to sequences. However, it is important to note, that the quantity of DNA does not determine the conservation of the genes or specific sequences, which can be seen due to the large number of inactive regions of the genome, no matter its size. Nevertheless, the probes would not hybridise if the sequences were not present, hence the conserved nature of these specific probes could also be the reasoning behind the variation in correlations coefficients. Certain chromosomes are also similar and more conserved in mammals with lower divergence rates, and thus protein and gene

expression affect the regulation of the sequences of species with increased divergence (Ferguson-Smith & Trifonov, 2007).

By using the probes, it creates an opportunity to visualise how certain evolutionary changes have occurred between cattle and other distinct species in a similar manner to Larkin *et al*. (2006) when they studied mink chromosomes with cattle BACs. A higher resolution is found with these BACs than standard karyotypic identification, where only macro-regions and gross chromosomal changes can be identified (Kubiak *et al*., 2020). Our results indicate that not only small, but also gross genomic rearrangements can be identified with single locus cattle BACs selected for their place around EBRs and thus highly gene dense sequences (Larkin *et al*., 2009). Adaptation through genome reshuffling around these EBRs with increased divergence is simple to study with the use of these BACs.

3.3.2 Cattle CH240-256G3 Probe

One probe, CH240-256G3 (cattle chromosome 11) not only worked across all 11 species, but also its placement in horses was novel as the current location is an UGS fold for this specific sequence. Combined with karyotyping, it was identified to be on chromosome 15. In cattle, the Calmodulin-Lysine N-Methyltransferase (CAMKMT) gene was found within the sequence of the CH240-256G3, and its equine ortholog is mapped to a chromosomal region of equine chromosome 15 which matches the fractional length relative to the p-terminus (FLPter) value calculated from ideograms, further validating our results. As the FLPter values measures the fractional location of the probe from the p terminus along the chromosome, these results aid in validating the genome sequence data (Sakamoto, *et al*., 1995). However, it is important to note the subjectivity of this measurement which relies not only on user consistency and repeatability, but also on image resolution to be done accurately. Nevertheless, this type of information could aid bioinformaticians who are constructing and updating the equine assembly, by physically mapping the location of this unplaced genomic scaffold (UGS),

indicating the need for not only physical mapping, but specifically with locus specific BACs (Damas *et al*., 2016). In a similar set of circumstances for CH240-256G3, a sequence matching 89% of the clone nucleotide sequence was identified on an UGS in pigs. Through karyotyping, it was established this location was on chromosome 4. The CAMKMT gene has currently not been placed in the pig genome, once again identifying the need for highly conserved BACs for sequence identification.

3.3.3 X chromosome conservation

Looking at the X chromosome specifically, a noticeable difference was indicated between the location of cattle clone CH240-128C9 (cattle chromosome X) as the sequence was identified with BLASTn not only on horse chromosome X, but also on chromosomes 7 and 12, with over 95% identity accuracy. This sequence duplication could be one of the many gross changes which have contributed to reproductive isolation and divergence of these two species millions of years ago, purely due to the highly conserved nature of the probe. However, this observation could also purely be related to the repetitive elements or gene clusters in the genome thereby indicating multiple locations of the sequence. Further investigation into this BAC clone's sequence features and gene content could provide further information on the genomic conservation of this region. Similarly, cattle BAC CH240-359O3 (cattle chromosome X) was found subtelomerically in cattle, while in horses, this distal region was inverted. Raudsepp *et al*. (2004) demonstrated the high conservation between horse and human X chromosome maps, and this was confirmed in our results as both cattle BACs CH240-128C9 and CH240-29N7 were found in almost the same locations on equine Xq and human Xq. Based on FLPter calculations, the relative lengths of the p and q arms of horses and humans also seems to be almost exactly the same.

The cattle X probes which were conserved in the other species, were all found on the respective species' X chromosomes, except for the possum (the only Marsupial in this study) in

which the probes were found on autosomes which is the same as what was assumed in a study by Graves (2016). Conversely, both cattle X genes non-receptor tyrosine kinase homology 2 domain containing 1A (SH2D1A) and Teneurin transmembrane protein 1 (TENM1) (found within the CH240-129C9 region), which were highly conserved in the Eutherian mammals, were also found on possum chromosome X, indicating the highly conserved BAC region and the genes which were found within it. According to Graves (1991), the short arm or p-arm of the Eutherian X chromosome is found autosomally in marsupial species, which validates our findings here. There is a possibility that the genes on these two BAC clones have acquired different functions in marsupials than they do in Eutherians if they have diverged from this genomic region from when they had a common ancestor, which is an expected occurrence due to the difference in chromosome arrangements between the two clades (Koina, Fong & Graves, 2006).

3.3.4 Genes

The regions surrounding evolutionary breakpoint regions (EBRs) are highly gene rich and thus it is no surprise that 49 different genes were identified within the cattle BAC clone regions. On further investigation, many of these genes were found in the same sequences within sheep, horses, dogs, humans, and mice. 6 genes specifically were found within the BAC regions of CH240-420D19, CH240-339P15, CH240-329H2, CH240-368N15 and CH240-128C9 in all the species listed above, both within fractional length relative to the p-terminus (FLPter) mapped regions and through BLASTn searches (PAX2, ZEB2, COPS9, TENM1, SH2D1A and PCDH9). As expected, the underlying functioning of these genes was similar across the species, and all but one (SH2D1A) were involved in neuronal development, but the main function in adult mammals was often quite different (Dressler & Woolf, 1999; Yan *et al*., 2003; Bruining *et al*., 2015; Alkelai *et al*., 2016; Epifanova *et al*., 2019). This once again reassures the molecular clock

hypothesis, whereby genes have changed their function over time to the point where they are unable to create hybrids, causing reproductive isolation (Ho, 2020; Graves, 2016).

Two of the genes found in the cattle BAC sequences, which are involved in not only neuronal development, but also have an influence in fertility are protocadherin 9 (PCDH9) and teneurin-1 (TENM1). PCDHs are involved in cell-to-cell adhesion and mediate the calcium dependency of these unions in normal brain functioning (Strehl *et al*., 1998; Asahina *et al*., 2012; Bruining *et al*., 2015; Serranito *et al*., 2021; Zhou *et al*., 2022). They are regulated developmentally and are found in the nervous system (Asahina *et al*., 2012), on human chromosome 13, mouse chromosome 14 (Strehl *et al*., 1998; Lee *et al*., 2004; Zhou *et al*., 2022), horse chromosome 17 (Lee *et al*., 2004), sheep chromosome 10 (Mastrangelo *et al*., 2019; Serranito *et al*., 2021) and cattle 12 (Lu *et al*., 2021). While PCDH9 is associated with social cues and behavioural components in dogs (Bruining *et al*., 2015) and microcephaly in humans, it is a gene which influences loin strength (LS), a reproductive trait associated with calving ability and uterine robustness in cattle (Lu *et al*., 2021). Cattle with weak LS often have issues with calving ease and other reproductive system diseases due to uterine secretions (Lu *et al*., 2021).

Interestingly, TENM1 is also associated with social behaviour in dogs, specifically canine compulsive disorder (Dodman *et al*., 2016) and various reproductive traits in different mammals. It is a protein which regulates cell adhesion in the nervous system too (Dodman *et al*., 2016), and is highly conserved in mammals whereby human and mouse TENM1 have 97% sequence similarity (Alkelai *et al*., 2016) which further validates our results. Both Arishima *et al*. (2017) and Li *et al*. (2020) found goat and sheep litter sizes to be associated with TENM1, while Carvalho *et al*. (2019) demonstrated how this gene contributes the largest genetic effect to age at first calving in cows. Moreover, TENM1 participates in sex differentiation and embryo development in the foetus, specifically formation of the Müllerian duct in females, and hence can contribute dramatically to the fertility status of a woman (Carvalho *et al*., 2019). By identifying genes within highly conserved chromosomal regions such as the ones surrounding

EBRs, comparative mapping can be used to identify X-linked or other chromosomal fertility related issues in mammals, even if they have slightly different functions. In practice, the identification of these types of issues in other mammals, such as humans or mice, and inferring them in the cattle genome, can help farmers and breeders make informed decisions on cattle breeding or gene identification in their own herd. This allows for breeding out of a population a potentially deleterious allele or quantitative trait locus.

3.3.5 Limitations and future work

As previously discussed, some of the samples used for analyses were of poor quality which could have added to the results seen, especially in closely related species such as pigs. Ideally, fresh samples which have been cultured and fixed specifically for a study such as this one should be used, yet this is often not possible with rare species, or ones which are not located in the United Kingdom (UK), making this a limitation of our study. There were also a few problems regarding certain BACs which did not hybridise to any cattle chromosomes during the quality control, such as CH240-196L19, CH240-182G15 and CH240-288F24. CH240-196L19 did not hybridise to any species and thus it failed completely and there must have been an error during BAC processing.

Selection of BACs from the cattle genome can be a difficult task, as three different breeds have been used to create the current assembly and three different BAC libraries have been developed over the years (Zhou *et al*., 2015). Identification of BACs which were associated with one library and not the others, might cause ambiguity due to the genetic linkage and potentially breed or haplotype specific variability connected to a specific genome (Zhou *et al*., 2015). This could also be the reason two BACs were initially not used as their sequences could not be found on the National Center for Biotechnology Information (NCBI) or their archives. This type of development needs to be taken into consideration for future selection of cattle BACs. This was also demonstrated when certain genes were found in species as orthologues,

where they were indicated to be on a specific chromosome, yet when a BLASTn search was done, the chromosomal location did not match. This could stipulate either a possibly incorrect placement of the gene when sequence assembly was conducted, especially if the gene is only a model or predicted gene, or the clone sequence does not contain this gene in that species (Zimin *et al*., 2009). This could potentially be work for future studies.

In addition to BACs not working, one of the clones was found to be centromeric universal in cattle and it did not work in any other species (CH240-97L3). Alpha satellite DNA, which is found at the centromeres of most mammals, have strong signals when hybridising due to their hundreds to thousands of copies of monomers (Wolff, 2013). Although for this study, these types of probes are not useful, they can be utilised to detect aneuploidy in interphase and metaphase cells (Wolff, 2013) and therefore even though there was one less probe to use for the study, it could potentially be useful for other types of work, especially in fertility of cattle.

When specifically looking at the different orders, the majority of Perissodactyla are either vulnerable, near threatened, endangered, or critically endangered (Steiner & Ryder, 2011) and thus research into their reproductive isolation could be crucial for conservation of these species. Thus, the BAC clones could be very helpful in this regard to map homology between these different species. Additionally, only a small number of comparative molecular studies have been conducted within the Perissodactyla themselves (Steiner & Ryder, 2011), offering another avenue of future evolutionary research with BAC clones. As only one Perissodactyla was mapped (horse), as with the carnivores, primates and Diprotodontia, these specific orders would need further investigation to test the usefulness of the cattle BACs in identifying conserved regions and evolutionary traits within the orders. Primates in particular would be interesting to map, mainly because the cattle BAC clones within humans had a 64.6% success rate, identical to that of the dogs, while the two rodents had poor success rates, ruling out the Euarchontoglires clade for investigation. The clade of Laurasiatheria would be an ideal first step in future research due to the high hybridization average of 74.1% found.

Another technique whereby cattle BACs are used, has been described in section 1.3.1. BACs which are in subtelomeric locations can be used for cryptic translocation screening (O'Connor *et al*., 2017; Jennings, O'Connor & Griffin, 2020). The dairy and beef industries are large contributors to the UK economy and thus the fertility of the cattle involved is vital to the production of the next generation of animals (Lewis *et al*., 2022). By using subtelomeric clones, translocations such as the 1:29 can be identified and these animals can be removed from the population which will save not only money, but also will be beneficial for the environmental and welfare costs of the animal (Jennings, O'Connor & Griffin, 2020; Lewis *et al*., 2022). Some of the clones used in this study could be useful for improving upon the device designed by Jennings, O'Connor & Griffin (2020), such as CH240-339M3 (BTA18) and CH240-359O3 (BTAX). Similarly, sex chromosomes have the highest rate of abnormalities in horses, and improved tools to study these chromosome aberrations need to be further developed. Karyotyping is primarily used for structural and numerical chromosomal screening in horses, yet this can be subjective and labour intensive (Pauciullo *et al*., 2014). By using equine BAC clones in a similar manner to this study with FISH and hybridisation success, tracking and identifying these abnormalities could be useful for future commercial application in an industry where this technology is currently not being used. This is further discussed in section 4.

3.3.6 Conclusion

Mammals have different genomic architectures when compared to bird species, with regards to the abundance of highly conserved micro chromosomes only found in birds and reptiles, while mammals have higher chromosomal rearrangement rates than birds, thus the potential for our results to mimic those previously found with the avian universally hybridising BAC set is limited. However, our findings have significantly improved upon the previous mammalian work done in this laboratory with the selection of BACs as well as the potential to study genetically diverse species. Current findings suggest an Artiodactyl order or Laurasiatheria clade device

would be the most suitable for future research and an overall mammalian device would not currently be feasible due to the reduced accuracies noted on divergent species. However, using the BACs across the mammals we studied, we were able to identify five orthologous genes involved in neuronal development, two of which also have major functions in fertility: validating the 'molecular clock hypotheses' of gene functional changes causing speciation and reproductive isolation. The X chromosome of poorly studied species barasingha and bison, which don't have fully assembled genomes were also mapped to a chromosome-level, which could aid bioinformaticians with further work in placing genomic scaffolds, potentially supporting conservation efforts, and understanding how they have changed over time.

4. Specific aim 2 – To assess equine fertility using cytogenetic analysis and a novel technique for accurate detection of chromosome rearrangements

4.1 Background

4.1.1 A brief history of equine genetics

Domestication in the horse occurred about 5000 to 6000 years ago from the Eurasian Steppes for the purposes of agriculture, food, transport, and warfare (Petersen *et al*., 2013). Since then, selection for traits of importance within closed breeding herds and the process of breed specificity, slowly became more defined (Petersen *et al*., 2013). Performance (gait, strength, speed, and endurance), temperament and appearance (conformation, colour, and size) traits were selected for, and, over the years, more than 400 different horse breeds have been created, all for different purposes (Raudsepp *et al*., 2019). In general, the overall function of the horse in human lives has changed dramatically from work to pleasure, company, and sport, especially with the advent of the industrial revolution and use of machinery replacing horses.

Cytogenetic analysis of the horse began in 1912, when the first diploid number of the horse was reported by Kirillow, with approximately 20 to 34 chromosomes in male horses (Bowling *et al*., 1997). In the years that followed, equine chromosomal abnormalities started to be studied, such as X monosomy, sex chromosome mosaicism and XX and XY sex reversal eventually leading to better identification of gross chromosomal rearrangements (Brosnahan, Brooks & Antczak, 2010). The first standardized karyotype for the domestic horse was published in 1989 at The Second International Conference for Standardization of Domestic Animal Karyotypes, where the diploid number established definitively to be 2n = 64 with chromosomes 1 to 13 being metacentric autosomes, and the last 18 pairs, acrocentric autosomes, with one pair of sex chromosomes (Richer *et al*., 1990; Bowling *et al*., 1997).

During the 1980s and 1990s, genetic causes for different equine specific diseases such as hyperkalemic periodic paralysis (HYPP) and overo-lethal white foal syndrome (OLWS) or lavender foal syndrome were being investigated. With the advent of the Horse Genome Project (HGP) in 1995, these genetic causes were clarified (Brosnahan, Brooks & Antczak, 2010). Scientists from 12 countries and 22 various laboratories played a role in the HGP, developing large scale linkage and physical maps of the autosomes as well as the X and Y chromosomes (Brosnahan, Brooks & Antczak, 2010). Shortly after the creation of the HGP, Breen *et al.* (1997) first described the use of X whole chromosome painting probes to identify X chromosome abnormalities (section 1.3.3) (Bugno *et al.,* 2009) and in 2009, the second improved horse reference assembly, EquCab2.0 was developed from a Thoroughbred mare *Twilight* (Wade, 2009; Stock *et al.,* 2016; Raudsepp *et al.,* 2019). This assembly provided the opportunity for scientists to study inherited diseases of the horse and design new genome-wide tools for research, such as a comprehensive BAC library with over 315 000 clones (CHORI-241) (Brosnahan, Brooks & Antczak, 2010; Stock *et al.,* 2016; Raudsepp *et al.*, 2019). In 2018, the new and improved reference assembly EquCab3.0 was released off the success of EquCab2.0 with the same horse, *Twilight*, filling in numerous gaps relating to copy number variation (CNV) sites and segmental duplications, and including BAC end sequences and physical maps (Kalbfleisch *et al.,* 2018; Raudsepp *et al.,* 2019). EquCab3.0 had higher contiguity (40x increase) and all chromosomes were based on one scaffold bar one, equine chromosome 6.

4.1.2 The current state of equine genetics

According to the British Equestrian Trade Association (2019), there were around 850 000 horses in the United Kingdom (UK) and almost 375 000 households that owned horses. This is small in comparison to the 58 million predicted globally and less than a 10th of what the United States of America (USA) has (Petersen *et al.,* 2013). According to the Food and

Agriculture Organization (FAO) of the United Nations crop and livestock statistics, the UK ranks 24th in the world with respect to numbers of horse stocks, with the majority of horses found in the Americas. Due to the wide range of horses found and the new and improved genome assembly, more than 130 hereditary traits have been identified as well as multiple chromosomal disorders (Raudsepp *et al.,* 2019).

In the late 1990s, Breen *et al.* (1997) reported greater than 90% of chromosomal aberrations that were reported in horses, involved the sex chromosomes and a large proportion of these were specifically related to 63, X karyotypes. Horses with the second sex chromosome missing are generally infertile and often have small angles between their various back leg joints (Breen *et al.,* 1997; Millon & Penedo, 2009), yet many of these horses have also been reported to be phenotypically normal, due to a mosaic genotype, adding to the issue of "hidden" fertility and making identification difficult (Millon & Penedo, 2009). In general, true 63, X horses are normally female with underdeveloped ovaries and often have an anatomically small clitoris (Breen *et al.,* 1997; Millon & Penedo, 2009).

By the release of the EquCab2.0, the most frequent rearrangement was found to be sex chromosome aneuploidy with XX/XY chimerism and XY sex-reversal in mares according to Bugno *et al*. (2009), yet Millon & Penedo (2009) reported 63, X was still the most common problem in horses and sex-reversal, second. Regardless of the type of sex chromosomal abnormality, both causes deformed reproductive systems and behavioural issues (Millon & Penedo, 2009). 65, XXX and 65, XXY aneuploidies have also been reported, with some of these aneuploid horses producing viable offspring which have chromosomal rearrangements. This is rare and ill-advised as chromosomal translocations can be inherited thus further exacerbating the issue (Millon & Penedo, 2009). Similar to Breen *et al.* (1997), Bugno *et al*. (2009), Brosnahan, Brooks & Antczak (2010) and most comprehensively, Bugno-Poniewierska & Raudsepp (2021), reported a very low number of any autosomal abnormalities occurring in

horses, yet trisomies of some chromosomes have been identified through various techniques (section 1.3.3).

In one of the larger studies done by Bugno *et al.* (2009), a panel of FISH probes covering all the various chromosomal pairs was used, as well as standard karyotyping, to study 35 horses with either infertility, subfertility, or developmental abnormalities. Nearly 40% of the horses had abnormalities (Bugno *et al.,* 2009). However, 100-300 of the metaphase spreads needed to be analysed for each animal due to the number of mosaic forms present, yet by using specific FISH probes, this was an easy task due to the rapid method of assessment available when compared to karyotyping (Bugno *et al.,* 2009).

4.1.3 Rationale for chromosomal screening

4.1.3.1 Misdiagnosis

One of the major problems with horse chromosomal disorder identification is the assumption a horse is normal for most of its life until it is time to breed it. Phenotypically normal horses often go through rigorous rounds of hormone induced treatment prior to any cytogenetic analysis, being misdiagnosed as chronic anoestrous (McCue, 1998). Many of these horses, mares in particular, will fall pregnant yet continuously lose the foetus through resorption or abortion, and once again undergo hormone therapy, causing not only distress to the owner, but to the horse. The last resort, only after research, is to karyotype the mare and/or use FISH to identify the issue, and this is often when the misdiagnosis becomes evident. If molecular cytogenetics were to be used instead, the issue becomes a lack of commercially available FISH probes hybridizing to each of the various chromosomes (Lear & Layton, 2002; Bugno *et al.,* 2009; Demyda-Peyrás *et al.,* 2014), and hence karyotyping is normally advised.

As previously discussed in section 1.1.2, karyotyping requires extensive training specific to a species and many translocations cannot be identified due to their cryptic nature (Lear &

Layton, 2002). Additionally, not many laboratories are available to perform these tests due to lack of training and because of the complexity of the equine karyotype (Demyda-Peyrás *et al.,* 2014). Due to this, it is possible that horses are misdiagnosed with incorrect karyotyping, especially if they have mosaic or chimeric cells, and even more so if the horse has a normal phenotype, thus the frequency of potential aberrations could be higher than thought (Demyda-Peyrás *et al.,* 2014). Furthermore, the need for commercially available technology to objectively identify chromosomal rearrangements or rectify misdiagnoses is critical in a world where assisted reproductive technologies are used on a regular basis and ever increasingly high prices of horses are being seen.

4.1.3.2 Frequency

There is a large discrepancy in the reported frequency of chromosomal abnormalities found in the equine worldwide population. For example, Blue, Bruère & Dewes (1978) suggested a value as high as 68% of infertile mares have chromosomal abnormalities (Breen *et al.,* 1997). Considering the average first cycle pregnancy rate, per cycle pregnancy rate and average pregnancy rate per cycle (frozen semen) for horses is approximately 49%, 62% and 30-40%, respectively, there is a significant proportion of mares which are not falling pregnant, and this could be to do with chromosomal abnormalities (Samper, 2001). Of course, environmental factors and breed play a major role in conception rates, yet a proportion this high must have a genetic influence. In a study by Nie, Momont & Buoen (1993), an occurrence of 1.5% of sex chromosome abnormalities was found in a cohort of 204 mares, and they predicted the prevalence of these abnormalities to be around 3% of the population. This value was used as a baseline in literature for the last 30 years, even with the improvement in cytogenetics, gene mapping and sequencing, as well as the improved reproductive technologies such as AI and embryo transfer (ET).

Another value was developed and approximated by Bugno-Poniewierska & Raudsepp (2021) based on large-scale surveys, to be around 2-5% of horses with chromosomal abnormalities

affecting fertility. Additionally, the 68% of infertile mares (only) with chromosomal problems, stated by Blue *et al.* (1978), is now predicted to be around 30% of horses (male and female) with reproductive or developmental issues having chromosomal rearrangements (Bugno-Poniewierska & Raudsepp, 2021). Even though there has been an improvement in identification of chromosomal abnormalities, this is still a large proportion of horses with issues. Based on the total population of horses globally (58 million) and a 5% chromosomal rearrangement rate, that equates to 2.9 million horses worldwide with potential chromosomal aberrations, where fertility could be reduced anywhere from 5% to complete sterility (Bugno *et al.,* 2009). When looking at mares of breeding age (4-20 years old) in particular, almost 1.5 million mares could potentially have chromosomal abnormalities globally, yet there is still no fixed way of identifying these commercially. Additionally, there is still no frequency given for the number of horses with autosomal rearrangements which is another problem which needs rectifying. This means the overall number of chromosomal rearrangements (autosomal and sex chromosomal) could be higher than the maximum of 5% stated by Bugno-Poniewierska & Raudsepp (2021).

4.1.3.3 Assisted reproduction techniques (ART)

One of the major problems in equine breeding is the heavy line breeding in some breeds such as thoroughbreds, standardbreds and warmbloods. Line breeding is the practice whereby parents which have a common ancestor are mated. However, the parental ancestors are not close enough related to be considered inbred. By using this breeding technique, a situation arises where there is heavy use of a small number of high genetic merit stallions for both natural covering and artificial insemination (AI), decreasing the genetic variation in a population of horses, potentially to the point of a genetic bottleneck situation. Horses are selected for breeding predominantly on pedigree and sporting record second, not on any specific genetic tests or genomic breeding values as with other livestock species. Similar to dairy milk industry breeding schemes, by doing this type of breeding pair selection on pedigree

and highly rated traits, there is an increased risk of inbreeding causing more reproductive issues and risk of *de novo* and inherited chromosomal abnormalities (Brosnahan, Brooks & Antczak, 2010). Furthermore, the use of AI and embryo transfer (ET) has greatly aided the ability to distribute genetics of high genetic merit animals around the world, but concurrently caused the reduction of the number of stallions being used (Brosnahan, Brooks & Antczak, 2010). Due to this, the distribution of potentially deleterious alleles can occur, reducing genetic diversity in a population and causing an increase in recessive disease heterozygosity in a stallion if careful genetic testing is not done prior to AI (Brosnahan, Brooks & Antczak, 2010).

Similarly, a mare of recommended pedigree yet carrying a recessive deleterious allele, can, through the use of embryo transfer, distribute this to a large proportion of offspring, once again decreasing genetic diversity and increasing the chance of reproductive issues and *de novo* translocations (Brosnahan, Brooks & Antczak, 2010). By using assisted reproduction technologies (ART) such as AI, ET and now, cloning, there is a much higher rate of backcrossing, line breeding and inbreeding. For example, horses which were once alive can have their genetics “revived” through the use of cloning. In this process, the DNA from a somatic cell of a high genetic merit stallion or mare which has passed away, is transferred into an oocyte which has had its nucleus, and consequently DNA, removed. If these animals are born fertile and live to breeding age, they can be used for breeding, potentially to their own granddaughters and grandsons (Brosnahan, Brooks & Antczak, 2010). As exciting and optimistic the use of ART is, it needs to be used carefully so as to not increase the number of chromosomal abnormalities and decrease the genetic variation occurring in the population.

4.1.3.4 High costs

ART technologies can also be of high cost to horse owners if their animals have low reproductive efficiency, as horses are known to have only one foal a year (much less that most other livestock species), making it an expensive loss if the mare were not to fall pregnant or if she were to lose her foal (Bugno *et al.,* 2009). As horses only come of breeding age between

three and four years of age, a lot of money is invested into them for their needs and welfare as well as the cost of buying them, prior to potentially breeding them (Bugno *et al.,* 2009). Only prior to breeding, horses undergo their first assessments of breeding fitness by a veterinary, yet not cytogenetic, inspection (Bugno *et al.,* 2009). It is at this stage where initial pathology of a problem can often be seen (Breen *et al.,* 1997), yet as previously mentioned, often there are no phenotypic characteristics of a chromosomal abnormality in mosaic horses and therefore horses go through rounds of hormone treatment to solve their anoestrus, further increasing the costs and decreasing the welfare of these horses (Bugno *et al.,* 2009). Many of the phenotypically normal horses go to auction and performance testing where they are frequently sold for hundreds of thousands of pounds, with there being no knowledge of an underlying fertility issue due to a chromosomal abnormality (Bugno *et al.,* 2009). Identifying a chromosomal abnormality at the foal stage of a horse's life, could save years of money and time spent on a horse, as well as give the horse improved welfare by reducing the rigorous hormone procedures it may undergo for conception (Bugno *et al.,* 2009; Millon & Penedo, 2009) and put it on a different pathway than breeding, such as companionship or sports disciplines. By taking these horses out of the breeding horse population, it could improve the overall fertility of the horse population (Pienkowska-Schelling, Kaul & Schelling, 2020).

By creating a commercially available novel FISH device to detect chromosomal abnormalities in horses at birth, could save the chance of misdiagnosis through other subjective cytogenetic tests such as karyotyping, reduce the frequency of these abnormalities and the distribution of potentially harmful genetics in the general horse population; and save time and costs involved with raising and owning horses, improving the overall welfare of the horses. It will also allow owners and veterinarians to make informed and thoughtful decisions with regards to clinical application and breeding.

4.1.4 Specific aims

With the above in mind, **specific aim 2** of this thesis was to develop a novel approach for accurate detection of chromosomal rearrangements affecting equine fertility. In particular:

- **Specific aim 2a.** To summarize the findings of a basic cytogenetic screening service in horses in the laboratory and to test the hypothesis that there are hitherto undiscovered chromosome rearrangements in horses
- **Specific aim 2b.** To provide an audit of 128 putative subtelomeric BACs in terms signal strength, specificity, and location
- **Specific aim 2c**. To design a multiprobe screening device based on the above probes with a view to improving on the basic cytogenetic service for example spotting cryptic translocations
- **Specific aim 2d.** To revisit the screening service in 2a to test the hypotheses that the use of sub-telomeric probes increases the detection capability of cytogenetic screening

4.2 Results

4.2.1 Specific aim 2a. To summarize the findings of a basic cytogenetic screening service in horses in the laboratory and to test the hypothesis that there are hitherto undiscovered chromosome rearrangements in horses

Over the course of two years, 19 different horses were karyotyped in the laboratory. Based on karyotyping alone, seven horses were found to have no abnormalities (D, E, G, J, M, P and R), three were mosaic animals (B, O and Q), five had sex chromosomes abnormalities (A, F, H, I and N), three had assumed autosomal translocations (K, L and S) and one horse was found to be karyotypically male, and not female, which was originally assumed based on phenotype (C)

(Table 4-1). All the horses which had samples sent in, were assumed to be mares (64, XX), yet had had previous issues falling pregnant or phenotypic problems such as smaller uterine horns, ovaries, or abnormal external genitalia. For example, Horse H had recurrent infertility even with hormonal treatment, while Horse I was a filly foal with physical deformities. Horse J had symptoms of Turner syndrome such as small body size, angular deformities, and small ovaries. Horse K had a soft cervix with minimal tone, with no ovarian tissue visible on either side, with very narrow uterine horns, as well as a uterine oedema. Horse M was found to have a very abnormal hormone profile (<10ng/mL) and lastly, horses N and P, had small/underdeveloped ovaries and uterus. Most of the mares (n = 12) were of known breed. Seven thoroughbreds (E,

Horse	**Breed**	**Karyotype**	**Result**
A	Unknown	64, XX with t(X)	X rearrangement
B	Unknown	64, XX/ 63, X	Mosaic with normal female and monosomic karyotypes
C	Unknown	64, XY	Karyotypically male, meant to be female
D	Unknown	64, XX	Normal female
E	TB	64, XX	Normal female
F	TB	63, X	Monosomic animal
G	Unknown	64, XX	Normal female
H	Unknown	64, X,inv(X)(p;q) or X,t(Xp;Xq)	Xp and Xq inversion/translocation
I	TB	64, XX or X,t(Xp;Xq) or X,inv(X)(p;q)	X inversion/translocation, unclear
J	TB	64, XX	Normal female
K	Suffolk	64, XX with t(20p)	20p translocation
L	TB	64, XX with t(5p; Xp)	5p and Xp translocation
M	Crossbreed (“sport”) Horse	64, XX	Normal female
N	Suffolk	64, XX abnormal	Abnormal X chromosome, unclear
O	OLD	64, XX/ 64, XY/ 63, X	Mosaic with normal female, male and monosomic karyotypes
P	Suffolk	64, XX	Normal female
Q	Arabian	64, XX/ 63, X	Mosaic with normal female and monosomic karyotypes
R	TB	64, XX	Normal female
S	TB	64, XX with t(20p)	20p translocation

Table 4-1 The horses karyotyped are listed, along with their breed and karyotyping results. t: translocation; inv: inversion. Equine chromosome 20p is the chromosomal area proximal to the centromere as there is no p-arm in acrocentric chromosomes. TB: Thoroughbred. Suffolk: Suffolk Punch.

F, I, J, L, R and S), three Suffolk Punches (K, N and P), one crossbreed ("sport") horse (M), one Arabian (Q) and one Oldenburg Warmblood (O) and 6 unknown horses (A, B, C, D, G and H) were recorded. There were no pathological and/or sterility/sub-fertility trends found within specific breeds. For example, no specific chromosomal arrangement causing infertility or sub-fertility was found in one specific breed.

To confirm these karyotypes and to check for the potential of a device to works, both individual Texas Red and FITC probes needed to be selected, labelled, and tested.

4.2.2 Specific aim 2b. To provide an audit of 128 putative subtelomeric BACs in terms signal strength, specificity, and location

Sub-telomeric BACs have been used to great effect in the laboratory for both pigs (O'Connor *et al.*, 2018; Lewis *et al.*, 2021) and cattle (Jennings *et al.*, 2020; Lewis *et al.*, 2022) screening for chromosome translocations. Currently it is estimated that the FISH-based approach detects up to twice as many abnormalities in pigs and over three times as many in cattle. Obviating the need for specialist knowledge of chromosome banding patterns is a great advantage over karyotyping and thus, an obvious next step is to apply this approach to horses. A total of 128 horse BACs selected from the CHORI-241 clone placement library based on the most recent EquCab3.0 genome (Kalbfleisch *et al.*, 2018) were selected based on concordance, which is if the clone orientation and size is as expected; unique positioning where the clone is mapped only to one location in the genome (in this case to the appropriate sub-telomeric region); 130-190 kb length based on literature research; and end-sequence/subtelomeric placing to pinpoint potential cryptic translocations. P- and q-arm subtelomeric probes were selected for equine chromosomes 1-13 and X as they are metacentric, while a probe proximal to the centromere (as close as possible) and another subtelomeric/distal to the centromere on the q-

arm were selected for the acrocentric equine chromosomes 14-31 chromosomes. Four subtelomeric BACs were selected for each chromosome, two on the p-arms/proximal and two on the q-arms/distal. The p-arm or proximal to the centromere BACs (n = 64) were labelled with FITC, and the q-arm or distal to the centromere BACs (n = 64) were labelled with Texas Red.

The probes were categorized into five different groupings, depending on how they hybridised to the chromosomes based on location, strength, and specificity. Equine clones were considered either (1) correct, (2) incorrect, to have (3) multiple signals (more than one chromosome pair signal), (4) unclear (could not karyotype) or have (5) no signal at all. Both the Texas Red probes and the FITC probes had 92.2% of their respective clones hybridise to a specific location on the chromosomes. When further investigating their location and specificity by karyotyping each probe, only 64.1% of all the Texas Red, and 56.3% of all the FITC probes, respectively, hybridised to the correct locations. Signal strength varied between the probes. The majority of the 128 horse BAC clones had strong signals (5/5 signal strength) even if they hybridised to the incorrect location. The Texas Red probes had a slightly higher overall signal strength ($\mu = 4.7/5$ signal strength) when compared to the FITC probes ($\mu = 4.6/5$), but this was not a significant difference ($p > 0.05$).

One p-arm probe (CH241-292C11) and four q-arm/distal probes (CH241-69N15, CH241-145C12, CH241-206P14, CH241-158E14) had unclear locations, while conversely four FITC (CH241-118J1, CH241-254K24, CH241-162E13, CH241-294K9) and one Texas Red probe/s (CH241-444C19), respectively, had no signals whatsoever. The p-/proximal and q-arm/distal probes both had 5 horse BAC probes incorrectly hybridise to the wrong chromosomes and/or to the incorrect locations. CH241-20G11 (equine chromosome Xq) hybridised to Xp and not Xq as previously expected, based on the selected National Center for Biotechnology Information (NCBI) genome sequence location. 20.3% of the Texas Red probes and 28.1% of the FITC probes, had more than one chromosome pair signal when imaging, thus being considered “multiple”. An example of a single chromosomal pair signal versus a multiple chromosomal

pair signal can be found in Figure 4-1. When comparing the FITC and Texas Red probes, equine chromosomes 14, 21 and 24, which only had Texas Red signals for them. 1 of the FITC probes (CH241-226M19) was subtelomeric universal and 6.3% of the FITC probes were centromeric universal which could be due to a nick translation labelling issue (CH241-470E7, CH241-228K14, CH241-206A7 and CH241-56C22), while none of the Texas Red probes had this issue.

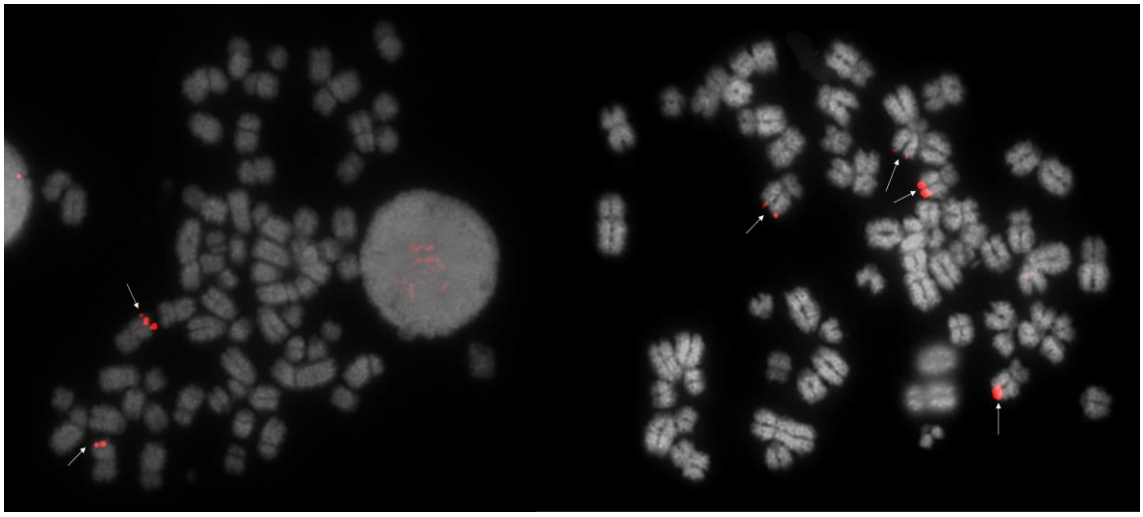


Figure 4-1 CH241-205O16 (equine chromosome 5) indicating a single pair of chromosomal subtelomeric q-arm signals (left); and CH241-470F5 (equine chromosome 9) demonstrating two different pairs of Texas Red subtelomeric signals (right) (Magnification x1000) (Image source: Own images).

When comparing the hybridisation strength, specificity, and location between the metacentric and acrocentric chromosomes, the FITC probe signal strength dropped as the chromosome number increased, from $\mu = 76.9\%$ in the metaphase chromosomes ($n = 20$), to $\mu = 42.1\%$ for the acrocentric and sex chromosomes ($n = 16$). Conversely, the Texas Red probes had a higher number of probes correctly hybridising to the acrocentric chromosomes ($n = 22$) than the metacentric ($n = 19$). Similar results for both the FITC and Texas Red probes was seen in the average signal strength for the metaphase and acrocentric chromosomes. The average p-arm probe signal strength (out of 5) for the metacentric chromosomes was $\mu = 4.9$, while the q-arm probes had a lower value of $\mu = 4.8$. Both the strength values dropped for the acrocentric chromosomes. Yet, when comparing the difference between the signal strength of the metacentric versus acrocentric probes, the FITC clone signal strength difference was significantly lower ($\mu = 4.3$) than the Texas Red probe signal strength difference ($\mu = 4.7$). The

overall probe results can be found in Table 4-2 (following three pages). This could be due to the FITC probes being manufactured first, and consequently a validated protocol had been established when the Texas Red probes were produced.

To not only confirm whether probes hybridise to the same/correct chromosome, but also to check for overlapping signals, probes were paired. The best combination of probes for each chromosome for a potential device were then tested in section 4.2.3. This was the next step in the process of equine BAC device production. Following this, FITC and Texas Red probes for each chromosome were selected to be paired and tested with one another on a multiprobe (equine chromosomes 1-24) and octochrome (equine chromosomes 25-X) device. Probes with multiple signals, no signal, universal signals or incorrect karyotypic locations were excluded (28 p-arm/proximal and 23 q-arm/distal probes, respectively).

4.2.3 Specific aim 2c. To design a multiprobe screening device with a view to improving on the basic cytogenetic service for example identifying cryptic translocations

4.2.3.1 Combination 1

Of the clones which worked from section 4.2.2, 36 p-arm/proximal probes labelled in FITC, and 41 q-arm/distal probes labelled in Texas Red, were tested in combinations and pairs to find the optimal probes for a potential device. For the first combination of probes (Table 4-3) tested on the multiprobe (equine chromosomes 1-24) and octochrome (equine chromosomes 25-X) devices, both 50% of the multiprobe and 50% of the octochrome pair combinations worked and the other half didn't. It was assumed these probes did not hybridise due to the poor metaphase preparations used in this experiment. Further investigation of the octochrome and multiprobe device results as a comparison are discussed in section 4.2.4.

Chr	FITC Probe	Location	Signal	Chr C/I		Chr	Texas Red Probe	Location	Signal	Chr C/I
1	CH241-155D20	Subtelomeric	5	Correct		1	CH241-444C19	No signal	-	No signal
1	CH231-218A20	Subtelomeric	5	Correct		1	CH241-198L13	Subtelomeric	5	Correct
2	CH241-60K15	Multiple signals	-	Multiple		2	CH241-38N11	Multiple signals	-	Multiple
2	CH241-135D13	Subtelomeric	5	Correct		2	CH241-171D17	Multiple signals	-	Multiple
3	CH241-140E5	Multiple signals	-	Multiple		3	CH241-157L23	Subtelomeric	5	Correct
3	CH241-16O19	Subtelomeric	5	Correct		3	CH241-268H12	Subtelomeric	-	Incorrect
4	CH241-226H13	Subtelomeric	-	Incorrect		4	CH241-238E5	Subtelomeric	4	Correct
4	CH241-58P11	Multiple signals	-	Multiple		4	CH241-69N15	Subtelomeric	-	Unclear
5	CH241-326B6	Subtelomeric	5	Correct		5	CH241-278A14	Subtelomeric	5	Correct
5	CH241-284D9	Multiple signals	-	Multiple		5	CH241-205O16	Subtelomeric	5	Correct
6	CH241-243H2	Subtelomeric	3	Correct		6	CH241-238O1	Subtelomeric	5	Correct
6	CH241-118P7	Subtelomeric	5	Correct		6	CH241-182N17	Subtelomeric	5	Correct
7	CH241-445A22	Subtelomeric	5	Correct		7	CH241-194A19	Subtelomeric	4	Correct
7	CH241-73B22	Subtelomeric	5	Correct		7	CH241-466E8	Subtelomeric	5	Correct
8	CH241-138N6	Subtelomeric	5	Correct		8	CH241-207B3	Subtelomeric	5	Correct
8	CH241-150C15	Subtelomeric	5	Correct		8	CH241-447E16	Subtelomeric	5	Correct
9	CH241-78C23	Subtelomeric	5	Correct		9	CH241-147D16	Subtelomeric	5	Correct
9	CH241-238A5	Subtelomeric	5	Correct		9	CH241-214B6	Subtelomeric	5	Correct
10	CH241-292C11	Subtelomeric	-	Unclear		10	CH241-470F5	Multiple signals	-	Multiple
10	CH241-213H3	Subtelomeric	5	Correct		10	CH241-452L20	Subtelomeric	5	Correct
11	CH241-89I23	Subtelomeric	5	Correct		11	CH241-162E1	Subtelomeric	3	Correct
11	CH241-243E9	Subtelomeric	5	Correct		11	CH241-346F11	Multiple signals	-	Multiple
12	CH241-338I24	Metacentric	5	Correct		12	CH241-441F18	Subtelomeric	5	Correct
12	CH241-332I23	Subtelomeric	5	Correct		12	CH241-189J7	Subtelomeric	5	Correct

13	CH241-462D15	Subtelomeric	5	Correct		13	CH241-463O14	Subtelomeric	5	Correct
13	CH241-205B14	Subtelomeric	5	Correct		13	CH241-7M5	Subtelomeric	5	Correct
14	CH241-226M19	Universal subtelomeric	-	Multiple		14	CH241-10G24	Multiple signals	-	Multiple
14	CH241-118J1	No signal	-	No signal		14	CH241-145C12	Subtelomeric	-	Unclear
15	CH241-75N19	Subtelomeric	5	Correct		15	CH241-206P14	Subtelomeric	-	Unclear
15	CH241-224I12	Subtelomeric	5	Correct		15	CH241-52I17	Subtelomeric	5	Correct
16	CH241-326J20	Subtelomeric	5	Correct		16	CH241-462M22	Multiple signals	-	Multiple
16	CH241-222K9	Subtelomeric	5	Correct		16	CH241-385L23	Subtelomeric	5	Correct
17	CH241-464K16	Subtelomeric	5	Correct		17	CH241-163E4	Multiple signals	-	Multiple
17	CH241-360G6	Multiple signals	-	Multiple		17	CH241-404G13	Subtelomeric	5	Correct
18	CH241-293H3	Multiple signals	-	Multiple		18	CH241-187P4	Multiple signals	-	Multiple
18	CH241-157L16	Subtelomeric	5	Correct		18	CH241-95G13	Subtelomeric	5	Correct
19	CH241-457F24	Subtelomeric	5	Correct		19	CH241-452B11	Subtelomeric	4	Correct
19	CH241-70D6	Multiple signals	-	Multiple		19	CH241-82J19	Subtelomeric	5	Correct
20	CH241-470E7	Universal centromeric	-	Multiple		20	CH241-219C9	Multiple signals	-	Multiple
20	CH241-71D19	Acrocentric	-	Incorrect		20	CH241-449F5	Subtelomeric	5	Correct
21	CH241-117G15	Subtelomeric	-	Incorrect		21	CH241-46M8	Subtelomeric	4	Correct
21	CH241-464H3	Multiple signals		Multiple		21	CH241-466D12	Subtelomeric	5	Correct
22	CH241-186N17	Multiple signals	-	Multiple		22	CH241-231A16	Subtelomeric	5	Correct
22	CH241-6L7	Subtelomeric	5	Correct		22	CH241-158E14	Subtelomeric	-	Unclear
23	CH241-46E18	Subtelomeric	-	Incorrect		23	CH241-90I9	Multiple signals	-	Multiple
23	CH241-27L9	Subtelomeric	5	Correct		23	CH241-343M4	Subtelomeric	-	Incorrect
24	CH241-46J10	Multiple signals	-	Multiple		24	CH241-50M21	Subtelomeric	5	Correct
24	CH241-228K14	Universal centromeric	-	Multiple		24	CH241-464J16	Subtelomeric	4	Correct
25	CH241-200A1	Subtelomeric	5	Correct		25	CH241-464F20	Subtelomeric	5	Correct

25	CH241-195H18	Subtelomeric	3	Correct		25	CH241-188F8	Subtelomeric	5	Correct
26	CH241-254K24	No signal	-	No signal		26	CH241-265D2	Subtelomeric	5	Correct
26	CH241-281A21	Subtelomeric	1	Correct		26	CH241-224I24	Subtelomeric	5	Correct
27	CH241-206A7	Centromeric universal	-	Multiple		27	CH241-222E18	Multiple signals	-	Multiple
27	CH241-162E13	No signal	-	No signal		27	CH241-92B21	Subtelomeric	5	Correct
28	CH241-223A13	Subtelomeric	2	Correct		28	CH241-271I18	Subtelomeric	4	Correct
28	CH241-294K9	No signal	-	No signal		28	CH241-244K4	Subtelomeric	5	Correct
29	CH241-219A13	Multiple signals	-	Multiple		29	CH241-292C13	Subtelomeric	-	Incorrect
29	CH241-56C22	Centromeric universal	-	Multiple		29	CH241-136O18	Subtelomeric	3	Correct
30	CH241-17B6	Subtelomeric	5	Correct		30	CH241-294F5	Subtelomeric	-	Incorrect
30	CH241-411P18	Subtelomeric	3	Correct		30	CH241-161P14	Acrocentric	4	Correct
31	CH241-381J7	Subtelomeric	5	Correct		31	CH241-336J2	Multiple signals	-	Multiple
31	CH241-24L6	Metacentric	-	Incorrect		31	CH241-170N2	Multiple signals	-	Multiple
X	CH241-469J20	Multiple signals	-	Multiple		X	CH241-20G11	Centromeric	-	Incorrect
X	CH241-159K1	Multiple signals	-	Multiple		X	CH241-457E2	Subtelomeric	5	Correct

Table 4-2 The results of each "FITC Probe" and "Texas Red Probe". Chr: The chromosome the probes were selected to be on, based on BAC choice on the National Center for Biotechnology Information (NCBI) prior to labelling and testing. Location: the probe chromosomal location based on imaging (Subtelomeric, Multiple signals, No signal, Universal subtelomeric/centromeric). Signal: the signal strength of the probe (out of 5, 5 being the highest and 0 being the lowest). Chr C/I: if the probe was found on the correct (C) or incorrect (I) chromosome based on karyotyping. If multiple signals were seen, no signal strength value was given and "Multiple signals" was stated under location, as well as "Multiple" in the Chr C/I. If the probe was karyotyped and found to be on the incorrect chromosome, no signal strength value was given, even if the "Location" was subtelomeric.

When investigating the individual probes on the multiprobe device, the equine chromosome 17d (distal) probes had poor signals (CH241-464K16 and CH241-404G13) and only the FITC BAC for equine chromosome 19 (CH241-457F24) correctly hybridised to its respective chromosome. As expected, due to the results in section 4.4.2, equine chromosome 21p (proximal) probe CH241-117G15 did not hybridise to the same chromosome as CH241-466D12, as CH241-117G15 is found based on karyotyping on equine chromosome 14p. However, the chromosome pair combinations which did hybridise to their correct locations, had good signal strengths and specificities, making them promising probe combinations for the final screening device/prototype.

Similarly, for the octochrome device, there were no metaphases on the slides for 3 out of the 8 BAC pairs. These probe combinations were repeated on single microscope slides to assess if this was due to poor slide preparation or poor probe hybridisation. For equine chromosome 26 and 29, only the distal (subtelomeric) probes indicated signals, CH241-224I24 and CH241-136O18, respectively. Once again, as expected due to karyotyping results in 2.4.2, the Texas Red probe, CH241-294F5 (equine chromosome 30) was confirmed to chromosome 26d (subtelomeric location on q-arm of equine chromosome 26) as it did not hybridise on the same chromosome as CH241-17B6, the equine chromosome 30 FITC probe which had previously been karyotyped to this chromosome.

A different sample was selected for testing of the second combination of probes to improve the metaphase preparations. Additionally, these were tested on individual slides, instead of on the multiprobe and octochrome devices, to allow for more metaphases to be tested in a larger area.

Combination 1				
Multiprobe				
Chr	**p-arm/proximal probe**	**q-arm/distal probe**	**Correct chr**	**Signal**
1	CH241-218A20	CH241-198L13	No metaphases	-
2	CH241-135D13	NONE	No metaphases	-
3	CH241-16O19	CH241-157L23	No metaphases	-
4	CH241-226H13	CH241-238E5	No metaphases	-
5	CH241-326B6	CH241-278A14	Yes	Good
6	CH241-118P7	CH241-182N17	No metaphases	-
7	CH241-445A22	CH241-466E8	No metaphases	-
8	CH241-138N6	CH241-447E16	No metaphases	-
9	CH241-78C23	CH241-147D16	Yes	Good
10	CH241-213H3	CH241-452L20	Yes	Good
11	CH241-243E9	CH241-162E1	Yes	Good
12	CH241-332I23	CH241-441F18	Yes	Good
13	CH241-205B14	CH241-463O14	Yes	Good
14	NONE	CH241-145C12	Yes	Good
15	CH241-75N19	CH241-52I17	Yes	Good
15	CH241-75N19	CH241-52I17	Yes	Good
16	CH241-326J20	CH241-385L23	Yes	Good
17	CH241-464K16	CH241-404G13	Yes	Poor
18	CH241-157L16	CH241-95G13	Yes	Good
19	CH241-457F24	CH241-82J19	FITC only	-
20	CH241-71D19	CH241-449F5	No metaphases	-
21	CH241-117G15	CH241-466D12	None	Good
22	CH241-6L7	CH241-231A16	No metaphases	-
23	CH241-27L9	CH241-343M4	No metaphases	-
24	CH241-46J10	CH241-50M21	No metaphases	-
Octochrome				
25	CH241-200A1	CH241-188F8	Yes	Good
26	CH241-281A21	CH241-224I24	TxR only	-
27	CH241-206A7	CH241-92B21	No metaphases	-
28	CH241-223A13	CH241-244K4	No metaphases	-
29	CH241-56C22	CH241-136O18	TxR only	Poor
30	CH241-17B6	CH241-294F5	None	Good
31	CH241-381J7	CH241-336J2	No metaphases	-
X	CH241-159K1	CH241-457E2	Yes	Good

Table 4-3 The first combination of p-arm/proximal and q-arm/distal probes tested in pairs. Chr: The chromosome they are assumed to be on based on National Center for Biotechnology Information (NCBI) selection. Correct chr: The outcome of the experiment. Signal: The quality of the signals seen. None: The probes did not work, even with metaphase spreads; TxR: Texas Red probe. Signals were classified as either “Good”, “Average” or “Poor”.

4.2.3.2 Combinations 2 and 3

18 different combinations of probes were tested, as shown in Table 4-4, all the p-arm/proximal and q-arm/distal probe combination hybridised to the expected chromosome pair locations with accurate specificity and location (except CH241-117G15). It was assumed this was due to the increased number of metaphases and better sample quality. As with the previous pair combination, to confirm CH241-117G15 did not hybridize to the correct location, it was paired with a different Texas Red probe (CH241-46M8) which confirmed its incorrect sequence hybridisation.

61.1% of the probe pairs had good signal strength, yet in ⅓ of the pairs, the FITC probe was considered poor. The Texas Red probes worked in all the combinations. The BAC pair on equine chromosome 22 (CH241-6L7 and CH241-158E14) were found to have overlapping signals. Additionally, the FITC signal (CH241-6L7) was not as strong as the Texas Red (CH241-158E14). Ideally a new probe should be identified for this chromosome which is a further distance from the Texas Red probe.

Combination 2				
Individual slides				
Chr	**p-arm/proximal probe**	**q-arm/distal probe**	**Correct chr**	**Signal**
1	CH241-155D20	CH241-198L13	Yes	Good
6	CH241-118P7	CH241-238O1	Yes	Good
7	CH241-445A22	CH241-194A19	Yes	Good
7	CH241-73B22	CH241-194A19	Yes	Good
7	CH241-73B22	CH241-466E8	Yes	Good
8	CH241-138N6	CH241-207B3	Yes	Good
8	CH241-150C15	CH241-207B3	Yes	Good
8	CH241-150C15	CH241-447E16	Yes	Good
19	CH241-457F24	CH241-452B11	Yes	Good
21	CH241-117G15	CH241-46M8	FITC incorrect	TxR Good
22	CH241-6L7	CH241-158E14	Yes	FITC poor; overlapping with TxR
23	CH241-46E18	CH241-343M4	Yes	FITC poor
24	CH241-46J10	CH241-464J16	Yes	FITC poor
26	CH241-281A21	CH241-265D2	Yes	FITC poor
28	CH241-223A13	CH241-271I18	Yes	Good

30	CH241-17B6	CH241-161P14	Yes	Good
30	CH241-411P18	CH241-161P14	Yes	FITC poor
30	CH241-411P18	CH241-294F5	None	FITC poor

Table 4-4 The second combination of p-arm/proximal and q-arm/distal probes tested in pairs. Chr: The chromosome they are assumed to be on based on National Center for Biotechnology Information (NCBI) selection. Correct chr: The outcome of the experiment. Signal: The quality of the signals seen. None: The probes did not work, even with metaphase spreads; TxR: Texas Red probe. Signals were classified as either "Good", "Average" or "Poor".

Combination 3 - Final Horse BAC combinations					
Multiprobe			**Octochrome**		
Chr	**p-arm/proximal**	**q-arm/distal**	**Chr**	**p-arm/proximal**	**q-arm/distal**
1	CH241-218A20	CH241-198L13	25	CH241-200A1	CH241-188F8
2	CH241-135D13	NONE	26	CH241-281A21	CH241-265D2
3	CH241-16O19	CH241-157L23	27	CH241-206A7	CH241-92B21
4	CH241-226H13	CH241-238E5	28	CH241-223A13	CH241-271I18
5	CH241-326B6	CH241-278A14	29	NONE	CH241-136O18
6	CH241-118P7	CH241-182N17	30	CH241-17B6	CH241-161P14
7	CH241-445A22	CH241-466E8	31	CH241-381J7	CH241-336J2
8	CH241-138N6	CH241-447E16	X	CH241-159K1	CH241-457E2
9	CH241-78C23	CH241-147D16			
10	CH241-213H3	CH241-452L20			
11	CH241-243E9	CH241-162E1			
12	CH241-332I23	CH241-441F18			
13	CH241-205B14	CH241-463O14			
14	NONE	CH241-145C12			
15	CH241-75N19	CH241-52I17			
16	CH241-326J20	CH241-385L23			
17	CH241-464K16	CH241-404G13			
18	CH241-157L16	CH241-95G13			
19	CH241-457F24	CH241-452B11			
20	CH241-71D19	CH241-449F5			
21	NONE	CH241-46M8			
22	CH241-6L7	CH241-231A16			
23	CH241-27L9	CH241-343M4			
24	NONE	CH241-464J16			

Table 4-5 The final equine BAC combinations for each of the chromosomes which were used to test the 19 different horses. The table contains the p-arm/proximal and q-arm/distal probes and the chromosome they are assumed to be on, based on karyotyping. Chr: Chromosome.

It was confirmed that there were no FITC probe signals for equine chromosomes 14, 21, 24. Therefore, for the selection of the final probe combinations available for the potential screening device, the p-arm BACs were excluded for these chromosomes. The final probe combinations selected can be seen in Table 4-5.

4.2.4 Specific aim 2d. To revisit the screening service in 2a to test the hypothesis that the use of sub-telomeric probes increases the detection capability of cytogenetic screening

4.2.4.1 Probe hybridisation success as a device

With the success of the different probes, each probe pair was investigated based on hybridisation across the 19 horses from section 4.2.1. 11 of the 19 horses which had previously been karyotyped through the commercial cytogenetic company described in section 4.2.1 were tested with the screening device developed in section 4.2.3. These 11 horses had enough metaphases to be used for the analysis (C-H, J-K, O, Q-R), and the other 8 mares had poor samples and thus could not be analysed with the device and the octochrome (A-B, I, L-N, P, S).

From the 11 animals screened with the probes from Table 4-5 in section 4.2.3, the mean success rate for the FITC probes with regards to their hybridisation success to the 11 horses, was found to be $\mu = 57.8\%$ and for Texas Red, $\mu = 75.9\%$, which was significantly different ($p = 0.0289$). Even though CH241-71D19, an equine chromosome 20p (proximal to centromere) probe, had previously been tested, it did not have FITC signals appear on any of the individual horses. Concurrently, equine chromosome 30 and X both had a 100% success rate for their probe combinations, and 18 of the 32 probe combinations had a mean success rate of $\mu = 70\%$. Eight of the probe pairs had an average hybridization rate of $\mu = 50$-69%, three combinations, $\mu = 30$-49%, and three had poor overall success rates below $\mu = 29\%$. These results can be seen

in Figure 4-2 (on page 185). Inconsistencies in hybridisation success rates per probe could be due to poor sample quality or poor probe processing and quality control prior to testing.

Specifically looking at the multiprobe and octochrome devices, when the p-arm/proximal multiprobe hybridization rate of $\mu = 49.4\%$ was compared with the q-arm/distal multiprobe hybridization rate of $\mu = 70.1\%$, there was a significantly higher rate for the Texas Red probes ($p = 0.019$). This significant result was not found for the FITC ($\mu = 83.1\%$) and Texas Red ($\mu = 93.2\%$) probes on the octochrome ($p = 0.617$) which was expected due to the high hybridization success rates for both probes. When the overall octochrome results for both FITC and Texas Red were compared with the multiprobe, there was a significantly greater success rate with the probes on the octochrome ($\mu = 87.5\%$) when compared with the multiprobe device ($\mu = 62.9\%$) ($p = 0.005$). This could be to do with the increased surface area for each probe, and thus more metaphases and chance of hybridisation for those probes. This type of result between the multiprobe and octochrome success rates, was also witnessed when looking at the individual p-arm/proximal and q-arm/distal clones. Overall, there was an improved rate of hybridisation for the q-arm/distal probes, regardless of the type of device used for testing.

4.2.4.2 Karyotyping results when compared to FISH results

To understand the difference between the FISH results and the karyotyping results, the horses were individually analysed to further establish if any errors were made with karyotyping. Eight horses (A, B, I, L, M, N, P and S) had metaphase preparations which did not work with the probes. As these 8 horses could not be tested, they were set up on individual slides to assess the condition of their X chromosomes only. As previously mentioned, the 11 horses were tested with the multiprobe and octochrome devices. From the eight horses which were tested with the X chromosome probes, half of the horses had their karyotypic conclusions confirmed. These results are summarised in Table 4-6.

Horse	Karyotype	FISH	Overall result
A	64, XX with t(X)	64, XX, t(X) mosaic	Mosaic X chromosome translocation Confirmed with FISH
B	64, XX/ 63, X	64, XX/63, X	Mosaic normal and monosomic Confirmed with FISH
I	64, XX or X,t(Xp;Xq) or X,inv(X)(p;q)	64, XX	Normal sex chromosomes Incorrect karyotype
L	64, XX with t(5p; Xp)	64, XX	Normal sex chromosomes Incorrect karyotype
M	64, XX	64, XX	Provisionally normal Confirmed with FISH
N	64, XX abnormal	64, XX	Normal sex chromosomes Incorrect karyotype
P	64, XX	64, XX	Normal sex chromosomes Confirmed with FISH
S	64, XX with t(20p)	64, XX/63, X	Mosaic normal and monosomic Mosaicism identified with FISH Could not identify the 20p translocation

Table 4-6 The results of the 8 horses tested with the X chromosome probes. The overall results based on karyotyping and FISH probe results available can be seen in the last column. Karyotypes confirmed with FISH are stated as “confirmed with FISH”. Karyotyping mistakes found with FISH are stated as “incorrect karyotype”.

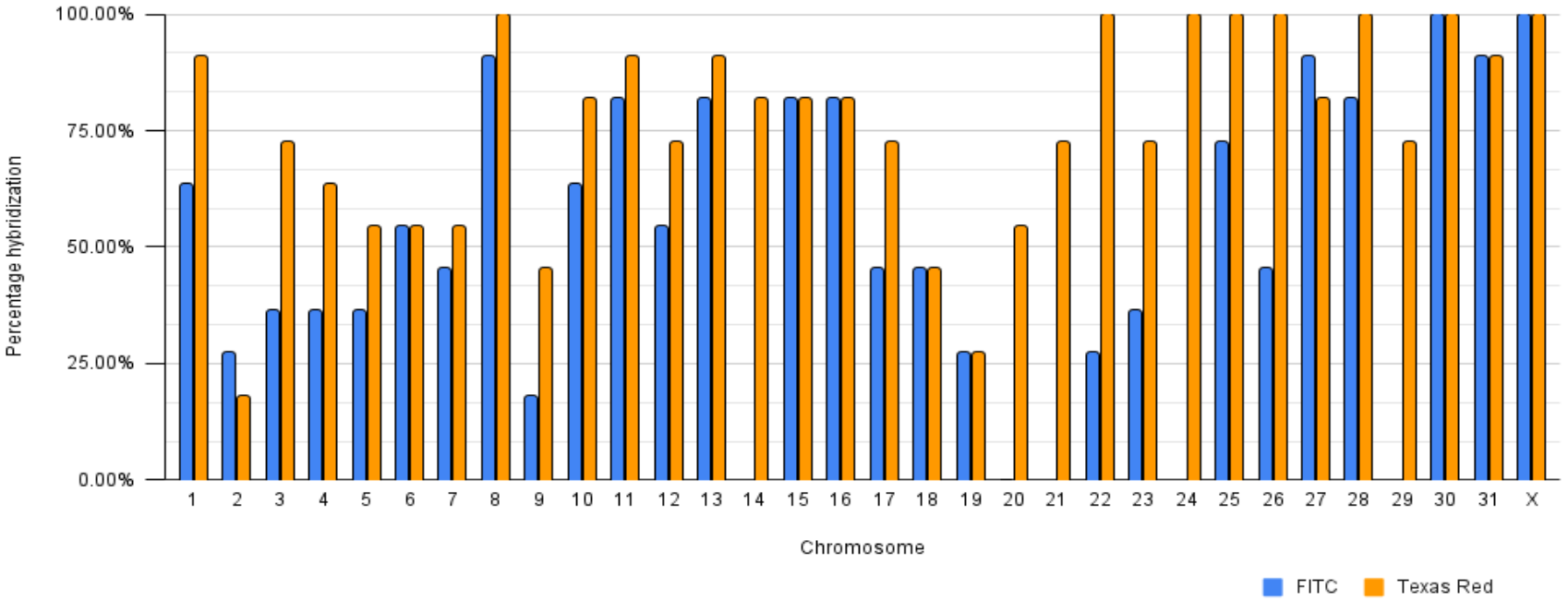


Figure 4-2 The FITC (blue) and Texas Red (orange) probe multiprobe (equine chromosomes 1-24) and octochrome (equine chromosomes 25-X) hybridisation success rate based on the probes per chromosome in Table 4-5. Chromosomes are listed on the x axis and the average percent hybridisation success rate, is on the y axis. Chromosomes 14, 21, 24 and 29 did not have appropriate proximal probes (FITC) and thus were excluded, while the distal probes (Texas Red) for those chromosomes were included.

Horse A was confirmed to have an X chromosome translocation, but it is unclear with which chromosome it was with due to the lack of probes. This translocation was also seen to be mosaic based on the FISH images, as the probes were found subtelomerically in some of the images and not in others (Figure 4-3). The red and green artefacts (additional spots on the different chromosomes) seen in the images, such as in Figure 4-3, could be due to repeats which were not blocked with an agent such as an equine hybloc. This sample should be repeated with a repetitive sequencer blocker to further establish the diagnosis for this horse to ensure reduced subjectivity and improve repeatability. Horse B was confirmed to be a mosaic normal and monosomic animal mare and mares M and P were confirmed to be normal (64, XX) (Table 4-6). The statements "normal sex chromosomes" are used, as it is unclear if there are any autosomal rearrangements, however, based only on the horses' sex chromosomes, they are considered "normal". Nevertheless, horse I, which was thought to have a chromosome X translocation, was incorrectly karyotyped as identified through FISH, as it had normal sex chromosomes (64, XX). Similarly, horses L and N, which were both assumed to have chromosome X abnormalities through karyotyping, were found to be normal 64, XX mares with the FISH probes, thus having been incorrectly karyotyped as well. Although these three horses do not have sex chromosome abnormalities based on the FISH probes used, it does not

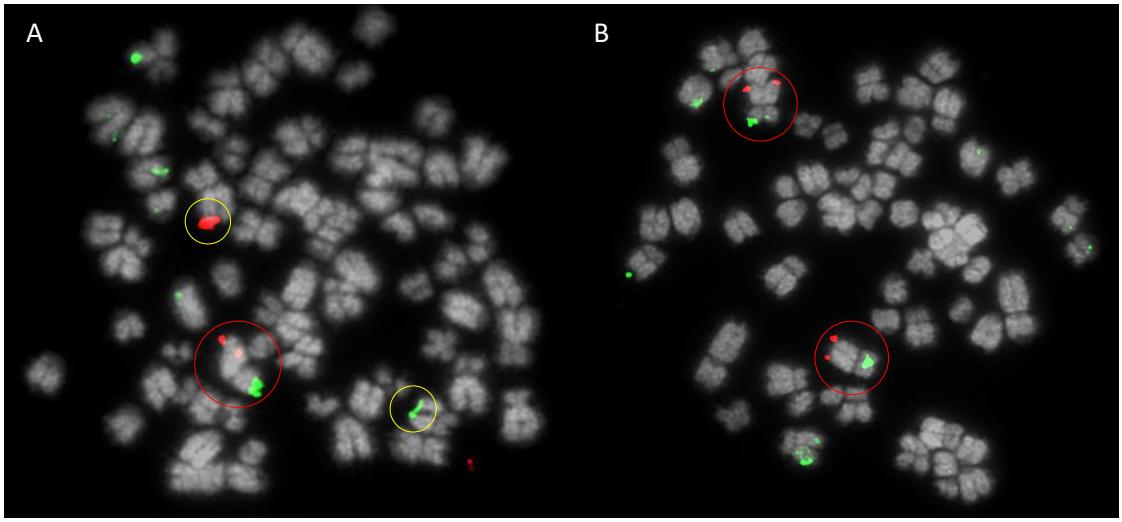


Figure 4-3 Horse A, with an X chromosome translocation (A); and a normal X chromosome arrangement (B). The red circles indicate normal X chromosomes, while the yellow circles show the translocations. Thus, this horse was characterised as having a mosaic X chromosome translocation (Magnification x1000) (Image source: Own images).

exclude the possibility of other autosomal rearrangements which were causing their infertility issues, but when purely investigating the sex chromosomes, no aberrations were found, indicating the importance of FISH screening and the mistakes which can easily be made with karyotyping. Lastly, horse S which was assumed to have a translocation in chromosome 20, was found to be a mosaic animal with both normal and monosomic metaphase spreads. This had previously been missed with karyotyping. There is still a possibility that this horse has a 20p translocation, yet without further research this cannot be confirmed.

Horse	Karyotype	FISH	HSR	Result
C	64, XY	64, XY	Good	64, XY male Confirmed with FISH
D	64, XX	64, XX	Poor	No detected abnormality Confirmed with FISH
E	64, XX	64, XX	Average	No detected abnormality Confirmed with FISH
F	63, X	63, X, t(20d)	Good	Monosomic X with a 20d (distal) translocation Unclear with FISH which autosome the translocation occurred with. **Novel translocation** with FISH.
G	64, XX	64, XX	Good	No detected abnormality Confirmed with FISH
H	64, X,inv(X)(p;q) or X,t(Xp;Xq)	64, X, t(10p; Xp) and X, t(10q; Xq)	Average	10p translocation with Xp and 10q translocation with Xq Abnormality identified with karyotyping. **Novel translocation** confirmed with FISH.
J	64, XX	64, XX	Good	No detected abnormality Confirmed with FISH
K	64, XX with t(20p)	64, XX	Poor	Sex chromosomes were normal. Could not confirm the 20p translocation due to only 9 probe pairs hybridising.
O	64, XX/ 64, XY/ 63, X	64, XX, t(15q; 28q)/ 63, X	Good	15q translocation with 28q and mosaic monosomic and normal sex chromosomes **Novel translocation** with FISH
Q	64, XX/ 63, X	64, XX/ 63, X	Good	Mosaic normal and monosomic sex chromosomes Confirmed with FISH
R	64, XX	64, XX/ 63, X	Good	Mosaic normal and monosomic sex chromosomes Incorrect karyotype

Table 4-7 The results of the 11 horses tested with the multiprobe and octochrome devices. The hybridisation success rate (HSR) is based on the number of probes which hybridised successfully to the chromosomes available. HSR: Hybridisation success rate (%) is considered poor: 1-20; Average: 21-25; Good: 26-32 based on the number of probe or probe combinations which were successful. The overall results based on karyotyping and FISH probe results available can be seen in the last column.

Table 4-7 indicates the outcomes of both karyotyping and FISH for the 11 horses with clear metaphase spreads. Horse C which was karyotypically a normal male (64, XY), and phenotypically female, was confirmed to be 64, XY with FISH and image analysis. With the lack of a Y FISH probe, any horses with mosaic karyotypes or Y chromosomes found through karyotyping, had their FISH images individually counted for chromosome number and investigated for a Y chromosome. Mares D, E, G and J were confirmed to be provisionally normal (64, XX) based on the probes which did work. Novel translocations previously not identified with karyotyping were found for horses F and O with the FISH probes. Horse F was confirmed to be 63, X with FISH, and a translocation on chromosome 20q was identified, while horse O, was found to have a translocation between equine chromosomes 15q and 28q, with mosaic normal and monosomic sex chromosomes.

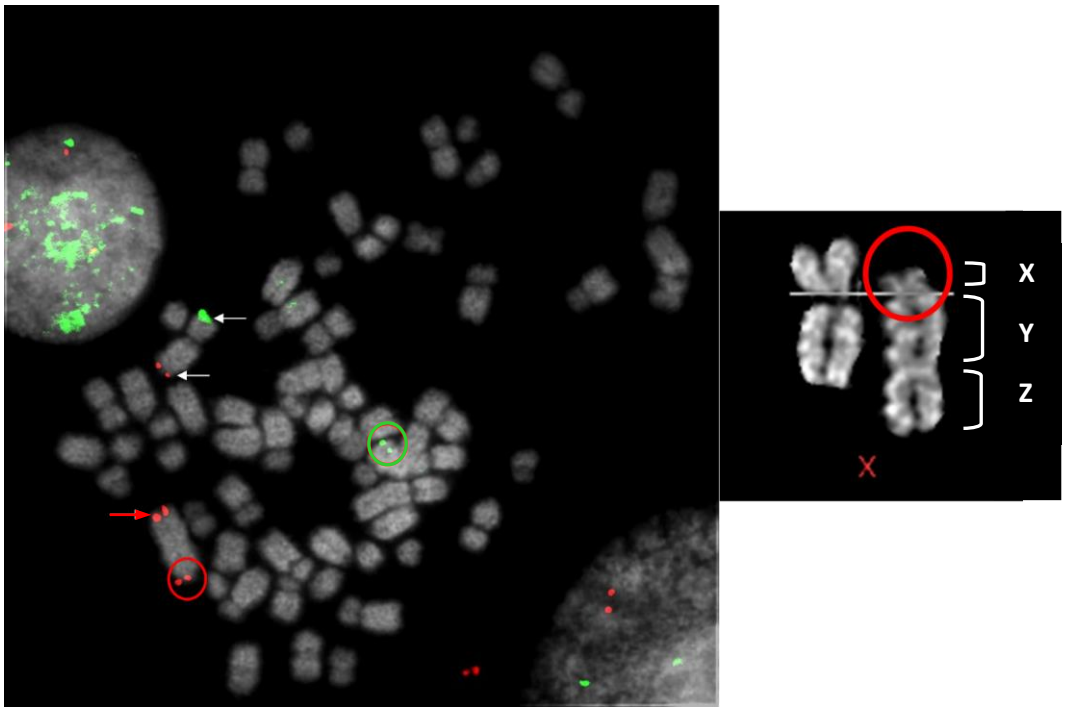


Figure 4-4 Images of horse H with the equine chromosome X FISH probes. A normal chromosome is indicated with white arrows pointing to the Xp (FITC) and Xq (Texas Red) probes (left). Based on the FISH results in the left image, chromosome Xp (green circle) has swapped with 10q, which has attached itself to the centromere of chromosome X. Part of Xq has attached itself to the end of 10q (red arrow), while the other part of Xq has rearranged to the opposite end of the centromere (red circle). The overall sex chromosomal translocation originally identified with karyotyping prior to use of BAC clones, is demonstrated on the right. In this image, "X" indicates part of Xq, "Y" shows 10q and "Z" is the other part of Xq.

The most interesting result was found with horse H. When karyotyping, there was an obvious problem with one of the X chromosomes, yet the nature of this issue was unclear. Through FISH results, it was found this mare had a novel translocation with two different chromosomal rearrangements. The Xp arm was found rearranged to the 10p arm while part of the Xq arm was attached to the end of 10q (which was translocated to the X chromosome centromere), and a small part of Xq was rearranged to the opposite end of the centromere attached to 10q (Figure 4-4), indicating a possible Robertsonian translocation.

Unfortunately, an equine chromosome 20p translocation identified in horse K through karyotyping could not be confirmed with the equine chromosome 20 FITC and Texas Red probes, as they did not work. This will need further investigation. Horse R had its karyotype results corrected with regards to its sex chromosomes. Q was karyotyped and confirmed with FISH as a mosaic normal and monosomic mare (no XX and X monosomic ratios were calculated for this study), while mare R which had a 64, XX karyotype, was found to be a mosaic animal with both 64, XX and 63, X metaphase and interphase spreads with the FISH probes. As previously mentioned with other results, as not all the probes worked on all the horses (not a 100% hybridisation success rate on any of the animals), these FISH results are provisional and are not completely accurate until a full panel of BACs is achieved.

Overall, FISH confirmed 10 of the karyotyping results, while 4 of the karyotypes were incorrect and were corrected with the FISH probes. 3 of the 19 horses had unclear results and therefore more probes are needed to confirm or disprove the karyotyping results. As horse F had a novel translocation yet more probes were needed to confirm the other autosome involved in this translocation, it fell under two categories in Table 4-8. Three novel translocations were identified with FISH. Table 4-8 summarises these results. This not only shows the need for more probes to continue this work, but also the subjectivity of karyotyping, and the promise the FISH probes show for cryptic abnormality identification.

	Result (n = 20)
Correct karyotype confirmed with FISH	10
Incorrect karyotype corrected with FISH	4
Unclear/need more probes	3
Translocations identified/confirmed with FISH	3

Table 4-8 Overall karyotype and FISH results for the 19 horses. Horse F fell under both a novel translocation and unclear/need more probes categories, hence there being a sample number of 20.

4.3 Discussion

4.3.1 Chromosome translocation detection using FISH

Cytogenetic screening and translocation detection is possible with the correct tools such as subtelomeric probes, yet selection of the correct probes can significantly affect the outcome of results (Bugno *et al.,* 2009). Although 92.2% of both the Texas and FITC probes hybridised to a chromosomal location, not all of them were subtelomeric or had singular signals, as only 56.3% of the p-arm/proximal probes, and 64.1% of the q-arm/distal BACs, had successful hybridizations to sub-telomeric locations of correct specificity and location with adequate strength. However, this presented the opportunity to use probes that were originally intended for a certain chromosome, to be inherently placed where another may have failed. For example, CH241-268H12, an equine chromosome 3q probe, was karyotyped to 31q, and as neither of the equine chromosome 31d BACs hybridized to the correct location, this probe could be used in their place, saving money and time. This will need to be confirmed with another equine chromosome 31d BAC, yet it is worth investigating to save money on future BAC purchases due to the high cost of long-sequence clones (Bejjani *et al.*, 2005). Nevertheless, this type of sequence assembly error needs to be addressed, as results have been found in a similar study by O'Connor *et al.* (2017) with regards to the pig genome and significant issues with the assembly of it. Moreover, 5 p-arm and 5 q-arm probes were found to be in the incorrect locations based on karyotyping and confirmed with respective FISH probes, further necessitating the need for physical mapping of the equine genome with BACs which have not previously been tested (O'Connor *et al.*, 2017). Y chromosome equine BAC

probes are also not available to visualize from the National Center for Biotechnology Information (NCBI), however they are available in various equine BAC libraries.

As most chromosomal related equine fertility issues are due to changes in the sex chromosomes (Bugno-Poniewierska & Raudsepp, 2021), it could be worth investigating a device purely focused on the X chromosome. When investigating eight of the 19 horses, 50% of them had their karyotypic results confirmed (A, B, M and P) while the other half (I, J and N) were either incorrectly karyotyped or their results could not be identified (S). Horse S was karyotyped to have a translocation on equine chromosome 20, yet due to the lack of metaphases, this sample could not be checked with the screening device and only its sex chromosomes could be checked. Chromosome 5, which was assumed to have a translocation with chromosome X for horse L, has been shown in the literature to be relatively unstable, as it has been involved in Robertsonian translocations resulting in two acrocentric chromosomes in the close relative, the Przewalski's horse (Huang *et al.*, 2014) and in a translocation with chromosome 16 (Bugno-Poniewierska & Raudsepp, 2021), so this is a potential translocation to look out for. Conversely, there is currently no literature substantiating the involvement of horse chromosome 20 in any structural or numerical aberrations in horses, so if this could be determined, a potential novel translocation would be identified. However, human chromosome 6 which is of evolutionary synteny to equine chromosome 20, has been known to be involved in balanced translocations affecting reproductive failure in humans and could thus be further investigated in horses (Carbone *et al.,* 2006).

On the contrary, without the use of the FISH BAC panel, two novel translocations in horses F (63, X, t(20q)) and O (64, XX, t(15q;28q)) would not have been discovered. This is especially true, as horse F had no phenotypic traits, other than one small ovary, to indicate any sort of reason for infertility or subfertility. Horse O however, indicated abnormal reproductive anatomy and intermittent endometriosis with caudal reproductive pain, which are common phenotypic indications in equines of translocations or numerical abnormalities (Lear & Layton,

2002; Bugno-Poniewierska & Raudsepp, 2021). In the literature, autosomes 15 and 28 are known to be part of other translocations and abnormalities such as X;15 and equine chromosome 28 trisomy found by Power (1987), Lear & Layton (2002) and Brito *et al.* (2008), so the involvement of these chromosomes in translocations is not novel. Unfortunately, there still is not enough probe information yet to substantiate a full screening device which was seen with the horses which were assumed to be normal based only on the probes which hybridised in them. For example, mares D, E, G and J were stated as provisionally normal (64, XX), but as only a portion of the probes hybridised to the samples, some autosomes had no results, of which there could be cryptic translocations.

However, even though a full screening device is not yet fully available, using them allowed the exact nature of a gross abnormality identified with karyotyping, to be confirmed with FISH in horse H. This was also a novel finding which would have not been possible without the FISH probes. Originally horse H was assumed to have a pericentric inversion, yet the FISH probes identified a translocation with equine chromosome 10p as well as a breakage of equine chromosome Xq (64, X, t(10p; Xp) and t(10q; Xq)). Additionally, five horses had X numerical losses as 63, X monosomy or mosaicism, which is expected as it is the most common equine aberration as identified in the literature (Bugno-Poniewierska & Raudsepp, 2021). Novel translocations are unique, as only 15 have currently been cited in the literature, most commonly with chromosomes 1, 4, 13 and 16 (Bugno-Poniewierska & Raudsepp, 2021), none of which have been found in our results. This further enhances the need to expand on this work for a device which could look for these novel translocations with increased accuracy. Although there was a lot of uncertainty in the results, there is potential for a translocation screening device to be developed with further probe testing and confirmation, especially as there are currently no commercial FISH translocation testing systems for horses available.

4.3.2 Conventional karyotyping versus FISH analysis

It has been seen in previous work by Lewis *et al.* (2021) and Lewis *et al.* (2022), karyotyping can't identify cryptic translocations which can ultimately cost millions to a company or breeder, based on one breeding animal being missed by a chromosomal abnormality. Horses F and O are both breeds which are sold for incredibly high prices for individuals, the Thoroughbred (TB) and Oldenburger warmblood (OLD). At the recent Goffs UK Premier Yearling sale in August 2022, the average price of an individual two-year-old horse was £44 043 with a 5-year average of £45 610, while prize winning fillies and colts can sell for £1.2 million (Burton, 2022). Similarly, Oldenburger Warmblood foals (under 1 year of age) can also sell for up to £200 000 and prized competition horses £396 627 as found by the recent Oldenburger Horse Society's Elite Auctions (Eurodressage, 2022). Mares and stallions which have cryptic abnormalities such as the ones found in F and O, could cost the owners a large sum of money if unnoticed by karyotyping, like these ones seen with boars in the pig breeding industry (O'Connor *et al*., 2017; Lewis *et al.,* 2021) and cattle in the dairy and beef industry (Jennings, Griffin & O'Connor, 2020; Lewis *et al.*, 2022).

The majority of the horses tested in this study were assumed to be karyotyped correctly, prior to FISH analysis, due to special training and supervisory input, but this is time consuming, costly, and labour intensive (O'Connor *et al.,* 2017; Jennings, Griffin & O'Connor, 2020). Interpretation of horse karyotypes is known to be prone to human error due to the complexity of their banding patterns on the small acrocentric chromosomes (Brito *et al.*, 2008) causing difficulty for people to identify and distinguish certain chromosomes from one another. Therefore, making a multiple hybridization experiment based in a single assay, a potential significant contribution to the commercial breeding of equines, similar to ones developed for cattle by Jennings, Griffin & O'Connor (2020) and pigs by O'Connor *et al.* (2017). One could argue the cost of FISH is overall much higher with regards to reagents and initial input for BACs, yet the overall time investment, skill development and pay needed for a person to

conduct the karyotyping analysis, far outweighs the costs of a singular FISH experiment. Especially as once BACs have been tested and confirmed, future BACs can purely be analysed against the ones currently in stock and thus won't need to be individually karyotyped. If karyotyping is conducted by a trained professional, it can be very accurate, however measurements such as single target BAC analysis are more efficient and less error prone (O'Connor *et al.*, 2017). For example, horse R had its result corrected when the FISH BAC panel was used. Yet, 10 different probes didn't hybridize to the correct locations, and these would not have been identified without karyotyping. Overall, karyotyping has an important place in identification of potential gross chromosomal aberrations, yet for more cryptic or smaller structural changes, which affect fertility, a method such as FISH or even in the future, sequencing, are important (O'Connor *et al.*, 2017).

4.3.3 Limitations and future recommendations

The development of novel screening techniques with a new set of BACs has its challenges and limitations, but equally provides ample opportunity for improvement. One of the main problems encountered was the significant difference between the Texas Red and the FITC probes. Following testing on 11 of the 19 different horses, the Texas Red probes had an overall significantly higher average hybridization success rate of $\mu = 75.9\%$ compared to the FITC probes of $\mu = 57.8\%$. The superior results found in the Texas Red probes could be due to the sensitivity of FITC to pH changes and particularly photo bleaching, while Texas Red does not have this problem (Chen *et al.,* 2008). This type of difference could also be due to a labelling issue. All the FITC probes were labelled first, and equine probes had never been labelled by the company before. Because of this, there is a possibility for error to have been introduced early, prior to testing the probes in the laboratory. The Texas Red probes were labelled second and thus the protocol had been established by the suppliers, further demonstrating the possibility for the better fluorescence of one probe colour over the other. Luckily, as the protocol has

been established, if this was a problem, it won't be one for future research. However, probes which did not work at all, or were universal, will need to be excluded and new BAC clones will need to replace them in order to get a fully functioning set of probes to a point of commercial availability.

Similarly, there was no hybloc used for this research due to the lack of one commercially available (even though regular genomic DNA can be used to block repeats, it was not used in this study). Hybloc is needed for repetitive sequences as it suppresses cross hybridization of repetitive DNA therefore decreasing background noise and increasing signal intensity (Applied Genetics Laboratories, Inc., 2020). This could be the possible reason for the multiple signals seen with many of the probes. Future work could include the addition of a bovine hybloc into the probe mixture, potentially suppressing the signals of the repeats, yet this would not be ideal due to the vast evolutionary differences between the species of 78 millions of years apart (MYA) to their closest ancestor of the Laurasiatheria. Yet, based on our results from specific aim 1, section 3, this could be a potential avenue for further investigation. Additionally, the horse BAC clone sequences were deemed "concordant" and "unique" with their positioning which were stipulations regarding their positioning yet post experiment BLAST analysis of a number of probes, indicated many of the double or multiple signals seen such as for CH241-38N11 (equine chromosome 2) and CH241-171D17 (equine chromosome 2), were due to the clone sequences matching sequences on other chromosomes. This should be noted for future BAC selection.

Although the probes worked better on the octochrome than the multiprobe device (potentially due to poor slide preparation on the multiprobe devices, the lower concentration of sample metaphases or reduced slide area to conduct the experiment), only once the overall probe selection and labelling process is refined and each chromosome has both a p- and q-arm specific probe which works and hybridizes to the correct location, will a differentiation between the two devices be resolved or be accurately tested. The strength of the probe signal

is also determined by the sample quality as samples which may be of good enough quality for metaphase imaging, may have too much debris for the probes to attach to the metaphases; and thus, equine blood culturing and harvesting should be further investigated and improved, to optimize the method and technique used. This will allow the metaphase spreads to be consistent and more accurate and repeatable results will be obtained for translocation screening and the device overall. This being said, it is known that FISH probes can be used on less-than-ideal samples and these types of samples can normally not be karyotyped due to the lack of distinction between chromosomes (O'Connor *et al*., 2017).

It is unclear why the eight horses did not have good enough sample quality for probes to hybridise to them, but this could be due to the process of heating and cooling, storage, as well as humidified conditions in the methods used in the experiments. It is also possible that certain horse metaphase preparations could not tolerate these environmental factors. Another avenue causing concern could be due to length of time between harvesting and testing the samples, especially if it is over a year (Howe, Umrigar & Tsien, 2014). Within the fixative in which the samples are stored (3:1 Methanol: Acetic acid), the acetic acid causes cells to swell, which is ideal for imaging, but it unfortunately also evaporates quickly, allowing the methanol to shrink the cells over time and decrease the sample quality (Howe, Umrigar & Tsien, 2014).

Future work should begin with selecting, labelling, and testing more potential probes for the screening device, yet other avenues of study could include more in-depth research into the individual probes such as their evolutionary conservation level, CG content related to potential genes of interest and any known evolutionary breakpoint regions (EBRs). This will allow for more precise and informative decisions to be made regarding the probes prior to labelling and thus their location on the chromosomes. Additionally, genes of interest may be investigated, hence allowing for not only cryptic translocation screening, but also for potential gene of interest exploration.

4.3.4 Conclusion and impact

Based on the novel cryptic translocations identified such as 64, X, t(20p) (through a combination of FISH and karyotyping) and 64, XX, t(15q; 28q), the recognition of the specific nature of a translocation 64, X, t(10p, Xp) and t(10q; Xq) previously found by karyotyping, and the correction of a horses with the use of horse BAC probe panel in FISH (normal 64, XX corrected to mosaic 64, XX/63, X), the preliminary results for use of a chromosomal screening device are promising. With the diverse literature basis regarding frequency of autosomal and sex chromosomal abnormalities, the overall high cost of individual horses on the market, plus the impact individual mares and stallions can make in the industry, as well as the significant use of artificial insemination (AI) and assisted reproduction technologies (ART) in equine breeding, there is even further pressure to identify the nature of subfertility in horses with the use of a chromosomal translocation screening device. However, until a full panel of horse probes is available for screening, sex chromosome screening could be achieved with the probes available, and more importantly, karyotyping, is still vital for initial laboratory work and should be used as a basic tool for gross chromosomal aberration identification.

As a result of my efforts, cytogenetic screening of horses, Stallion AI has included it in their directory of services (Figure 4-5).

Figure 4-5 Two photos from the 2023 Stallion AI Ltd. *The Directory: A Breeders Guide*; including a written piece regarding karyotyping as an option for breeders. (Image source: Own images).

5. Specific Aim 3 - Development of a flow cytometric assay for membrane lipid oxidation in human and equine sperm

5.1 Background

5.1.1 Combining cytogenetics and semen analysis in male fertility evaluation

In horses, as mentioned in section 4, chromosomal aberrations are a significant factor determining mare as well as stallion fertility. Although these abnormalities are more commonly identified in mares, many studies have demonstrated the impact of structural changes in stallions as well. For example, a tandem t(1;30) translocation in a Thoroughbred stallion was identified by Long (1996); and a t(12;25) was found in an Arabian stallion by Ghosh *et al*. (2021). Both horses were discovered due to reduced fertility, yet they were phenotypically normal, similar to mares with mosaic normal and monosomic sex chromosomes. Although rare, stallions with translocations can produce viable offspring, such as a case whereby a Warmblood stallion with a t(4;30) translocation produced 9 phenotypically viable progeny, and although four of them were karyotypically normal, four had the same translocation as their sire, and one had a trisomy of equine chromosome 4p (Ghosh *et al.*, 2021). Thus, it is vital to test stallions for these kinds of problems, as not to pass them on to their offspring.

One of the more interesting and relevant studies to this section of work, was done by Ruiz *et al.* (2019), which investigated not only the chromosomal problems of a Friesian stallion potentially causing infertility, but also the semen parameters of the horse. The stallion was found to be azoospermic (absence of spermatozoa in the ejaculate) after repeated attempts at AI and via ejaculate evaluation (Ruiz *et al.*, 2019). He had no phenotypic abnormalities other

than slightly smaller than normal testicular size, and after cytogenetic testing, it was found that the stallion had a translocation between the Y chromosome and equine chromosome 13 (Ruiz *et al.*, 2019). Although azoospermia is rare in stallions, accounting for <1% of ejaculatory problems (Ruiz *et al.*, 2019), it is relatively common in infertile men at 10-15% (Sharma & Leslie, 2023). However, chromosomal rearrangements such as Y-autosome translocations in men, which can cause azoospermia, are incredibly rare (1/2000) when compared to environmental or other genetic factors leading to azoospermia (Alharbi *et al.*, 2022). Yet, if a Y-autosome translocation is present, 80% of the time it is associated to azoospermia (Alharbi *et al.*, 2022). Genetic abnormalities are found in approximately 10-20% of sub-fertile human patients (Kuroda *et al.,* 2020), which further validates the need for not only chromosomal screening, but also semen analysis in these cases to confirm the severity of the sub- or infertility. 60% of men who have balanced translocations have at least one abnormal semen parameter, thus they can be difficult to spot, hence chromosomal screening should become a norm (Kuroda *et al.*, 2020). Additionally, a large portion of infertility in male patients with out of parameter semen characteristics is unexplained and more novel testing protocols are needed to understand these types of problems.

The level of unexplained male infertility is currently unknown in the horse, yet it is approximated to be about 15% in human (Hamada *et al.*, 2012). As spermatozoa and semen quality are greatly affected by environmental factors such as genetic and physical stressors (Fraser, 2004), DNA damage and the impact of reactive oxygen species (ROS) have been studied in recent years as a potential indicator of explained male factor infertility due to their relationship with environmental factors. It is well known that ROS causes DNA damage and chromosomal abnormalities in spermatogenesis (section 1.4.1 and 1.4.4) (Drevet & Aitken, 2020; Aitken, 2017), yet the best way to measure these two parameters has been a matter of debate. Additionally, many andrologists do not agree with the World Health Organization (WHO) guidelines on semen analysis and semen parameter ranges (section 1.5 and section 6), for example, the use of computer assisted semen analysis (CASA) is not recommended (WHO,

2010; WHO, 2022), but is widely used (section 6). In the agricultural industry, CASA is predominantly utilised due to its cost effectiveness and ease of use for sperm counting, pH evaluation, morphology analysis, sperm concentration and other parameters (Waberski, Suarez & Henning, 2022). The efficiency of CASA is debated, as only a small number of sperm are evaluated, therefore not representing the whole sample, some scientists are recommending the use of flow cytometry (FC), in conjunction with CASA, to give not only subpopulation data, but an objective multiparametric representation of the whole sample (Peña, Ferrusola & Muñoz, 2016; Boe-Hansen & Satake, 2019).

5.1.2 Flow cytometry (FC)

The flow cytometer (FC) is an important, and currently commercially underused, piece of equipment in the andrology laboratory. With the right assay, it can provide a more accurate and objective means of semen evaluation, giving a better overall representation of the sample as a whole as it evaluates thousands of sperm in a single run, when compared to other (e.g. microscopy-based) semen analysis tests, which only measure a couple of hundred (Graham, 2001; Squires, 2005; Piasecka *et al.*, 2007; Petrunkina & Harrison, 2011; Peña, Ferrusola & Muñoz, 2016). Objections have been raised with regards to FC testing as individual sperm are not independently observed and investigated (Petrunkina & Harrison, 2011), yet if used in conjunction with other tests such as morphology fixing and staining, it could be beneficial for human and equine andrological analysis. As semen subpopulation data, such as white blood cells, cellular debris, bacteria, and sperm, can be separated using different dyes and templates, it allows a simple, fast and cost-effective means of semen analysis (Graham, 2001; Hossain *et al.*, 2011; Peña, Ferrusola & Muñoz, 2016; Boe-Hansen & Satake, 2019). Additionally, semen is a monocellular suspension and as it takes up dyes and fluorochromes with ease, it further adds to the potential of testing using this technology (Graham, 2001; Hossain *et al.*, 2011; Peña, Ferrusola & Muñoz, 2016; Boe-Hansen & Satake, 2019).

FC was first used for semen analysis in 1968 by Wolfgang Gohde and Partec; later used for sperm DNA analysis in the 1980s and sperm membrane integrity, oxidative stress, membrane fluidity/permeability, mitochondrial membrane potential and lipid peroxidation (LPO) from the 1990s (Evenson, Darzynkiewicz & Melamed, 1980; Hughes *et al.*, 1996; Petrunkina & Harrison, 2011; Peña, Ferrusola & Muñoz, 2016; Boe-Hansen & Satake, 2019) (Section 1.4.3). Flow cytometers use lasers emitted at different wavelengths to identify cells passing through in a single stream of fluid, whereby each filter separates out the cell based on different set parameters such as size, shape, and light scattering properties (Graham, 2001; Hossain *et al.*, 2011; Boe-Hansen & Satake, 2019). By using different fluorochromes, each individual sperm gives a specific absorbance/fluorescence allowing for an overall idea of the sperm population simultaneously (Graham, 2001; Hossain *et al.*, 2011). Recently, FC has become commercially available and used increasingly for reproductive biology research due to its user-friendliness and wide range of tests available (Hossain *et al.*, 2011; Petrunkina & Harrison, 2011; Battut *et al.*, 2016).

5.1.3 Commercial application

It is widely known that lipid peroxidation (LPO), reactive oxygen species (ROS) and DNA damage are all highly correlated with male infertility (30-80% causation) (Dutta, Majzoub & Agarwal, 2019; Nowicka-Bauer & Nixon, 2020). Because of this, many researchers and clinicians are currently attempting to find new and improved ways of testing for these three parameters with user friendly, reliable, repeatable, and cost-effective methods such as the Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) and Sperm Chromatin Structure Assay® (SCSA®) assays (DNA damage) which use a flow cytometer (FC), and MiOXSYS (Reactive oxygen species, ROS) (Sections 1.4.3 and 1.4.4). However, there are no known standardised tests for LPO, which offer a potential gap in the commercial market, especially with the ease of use of FC. “Bench-top” FC availability has allowed more sperm parameters to

be measured, such as DNA integrity, sperm viability, mitochondrial status, and acrosome reaction, thereby giving a predictive value of an individual's fertilising and freezing capacity (Graham, 2001; Squires, 2005; Piasecka *et al.*, 2007; Hossain *et al.*, 2011). Additionally, it has been used to identify ROS and LPO, yet often not very accurately (Hossain *et al.*, 2011; Peña, Ferrusola & Muñoz, 2016). As ROS and LPO can influence DNA damage, thereby causing the cytogenetic framework of a cell to be altered (Bohlander & Kakadia, 2015), it is vital to have accurate ways to measure these factors in a commercial setting for both frozen and fresh semen samples. Thus, developing a LPO test using FC could be a way forward.

FC is increasingly being recognized as a robust tool in veterinary and human andrology for addressing sperm function and quality, for humans as fertility management for couples, and in stallions due to their high semen and reproductive treatment costs. For example, elite Grand Prix show jumping (including Olympic gold medal winning) stallions such as Golden Hawk, Gunner, Big Star, Emerald and Copain du Perchet, all have semen straw costs of over £900 per dose and with an average of three doses needed for a mare to fall pregnant, good stallion fertility is vital to maintaining the equine industry (Stallion AI Services, 2023). As explained in section 1.3.3, the equine industry in the United Kingdom (UK) is worth 4.7 billion pounds (British Equestrian Trade Association, 2022) and reduced fertility can cost owners, breeders, and buyers a significant amount of money if a translocation is present, or a horse has poor semen characteristics, yet the animal has a normal phenotype (Vilar *et al.,* 2017). Similarly, couples who repeatedly try to fall pregnant can be sent to expensive fertility treatments without identifying the underlying cause of the problem, costing thousands of pounds prior to identification of the route of the issue (section 6). Commercial testing for these types of aberrations is crucial for the welfare, cost and emotional stress animals and people alike can go through for fertility.

Another factor to consider, for both stallions and humans, is the freezability and chillability of sperm. Because of the high use of artificial insemination (AI) in stallions and age-related

deterioration of human sperm quality (Harris *et al.*, 2011), sperm freezing is one of the more important aspects of fertility preservation in many mammals. FC can be used to determine freezing quality of sperm, as freezing, and thawing of semen can have a detrimental impact on the quality of the sperm (section 1.4.5) (Squires, 2005) and if the sperm thaws poorly, it will cause a problem for equine breeders such as further increasing costs, specifically for mare owners (Loomis, 2001). Additionally, if a stallion of high value passes away, his sperm can still be used for years after his death, yet if the sperm thaws poorly, these genetics are lost. In humans, if the semen doesn't thaw correctly after freezing, men can potentially lose the opportunity to conceive a child, making freezability of sperm vital to test. Cryoprotectants, which are highly permeable and low molecular weight chemicals (Di Santo *et al.*, 2012) are used to aid in freezing of sperm. Most cryoprotectants for both stallion and human sperm, contain dimethyl sulfoxide (DMSO) and propylene glycol as anti-freezing agents (Bustani & Baiee, 2021) for protection against ice crystal formation in the sperm (Di Santo *et al.*, 2012). In stallions, milk- and egg-based cryoprotectants and extenders are used to try and alleviate the potential for cold damage to occur (personal communication Stallion AI Services, 2022). A complication with cryoprotectants and extenders, is that stallion semen seems to be specific to either one type of protein base. For example, some stallion semen works better with an egg-based extender such as Spervital Red, while others chill better with milk-based extenders such as INRA 96™ or BotuSemen Gold™. Egg-based extenders are more feasible on a price and quality scale, yet some stallions do not chill well with this type of protein source (Bustani & Baiee, 2021). Regardless of whether it occurs as a result of freezing or chilling sperm, DNA damage to some degree does occur during the thawing process (Squires, 2005). Therefore, the ability to test for DNA damage caused by ROS accurately and objectively, and LPO (which is vital for sperm motility), both prior- and post-freezing/chilling, are crucial in commercial settings at an affordable rate (Griffin *et al.,* 2019). With all the above in mind, an accurate, cheap and technically reproducible approach to sperm DNA integrity screening and LPO would be of great benefit both to the horse breeding industry and in the clinic.

5.1.4 Specific aims

In this research, the use of a flow cytometer (FC) to test for both DNA damage using the Sperm Chromatin Structure Assay® (SCSA®), and a novel test for lipid peroxidation (LPO), as well as their relationship to known fertility indicators are therefore investigated.

Specific aim 3 of this thesis was therefore to develop a novel flow cytometric assay for membrane lipid peroxidation (LPO) in human and equine sperm and to the hypothesis that there is a correlation between that, and DNA damage (and, by extension, fertility) in stallions and men. In particular:

- **Specific aim 3a.** To optimise a commercially available cell-based non-gametic flow cytometric lipid peroxidation sensor and kit for sperm testing
- **Specific aim 3b.** To assess the association between lipid peroxidation, DNA damage and standardized semen parameters as a suite of tools for fertility screening in humans
- **Specific aim 3c.** To briefly evaluate lipid peroxidation and DNA damage testing as an effective method of flow cytometric fertility screening in stallions
- **Specific aim 3d.** To investigate the effects of (i) chilled versus frozen semen, (ii) good versus bad pregnancy outcomes, (iii) cryoprotectant type, and (iv) breed, on stallion lipid peroxidation values and DNA damage screening

5.2 Results

5.2.1 Specific aim 3a. To optimise a commercially available cell-based non-gametic flow cytometric lipid peroxidation sensor and kit for sperm testing

As a commercially available non-gametic lipid peroxidation sensor (LPS) to identify 4-hydroxynonenal (4HNE) was available for use, specifically for cells growing in a culture-based environment, it needed to be altered and optimised to be used on sperm. The following

subsections investigate various parameters for not only sperm lipid peroxidation (LPO) testing, but also for optimisation of a positive control using hydrogen peroxide to ensure quality control of variables which could impact the experiment. As human sperm was readily available due to prior consent from the patients at The Doctors Laboratory (TDL) Andrology (for more information please see section 2.6.2), these tests were conducted on human sperm. As mentioned in section 2.8.4, LPO is the ratio of red to green cells, whereby green cells are damaged and red cells are healthy. Therefore, a lower ratio should indicate a more damaged sample with regards to LPO (specifically 4HNE damage) in human patients.

5.2.1.1 Hydrogen Peroxide Concentration

First, the concentration of hydrogen peroxide (H_2O_2) on sperm as a positive control was investigated. To see the effect of an increasing hydrogen peroxide level on the sperm, untreated patient lipid peroxidation (LPO) ratio values were tested using the standard parameters stipulated in section 2.8.4 (no addition of H_2O_2 to untreated samples). These same untreated samples were then subjected to 250 µM (n = 9) (Figure 5-1B), 500 µM (n = 29) (Figure 5-1C), 1 mM (n = 4) (Figure 5-1D), 10 mM (n = 2) (Figure 5-1E), 100 mM (n = 2) (Figure 5-1F) and 1 M (n = 5) (Figure 5-1G) H_2O_2 to gain an understanding of the effect of the H_2O_2 on the LPO value.

A mean increase was seen in the LPO value from µ = 1.132 in the untreated samples, to µ = 1.223 (Standard error = 0.165) in the 250 µM, followed by a decrease in the LPO ratio to µ = 0.989 (Standard error = 0.306) with treatment of 500 µM H_2O_2. As the H_2O_2 concentration increased on the untreated cells, a further increase in the mean LPO damage was seen with 1 mM (µ = 1.624; standard error = 0.115), after which the positive control LPO mean once again dropped with the treatment of 10 mM H_2O_2(µ = 1.422). Lastly, the mean H_2O_2 positive control further decreased with addition of 1 M H_2O_2 (µ = 1.258; standard error = 0.103). The results

are summarized in Table 5-1. These were unexpected results, as a general increasing or decreasing trend with the positive control was expected to be seen (Figure 5-2).

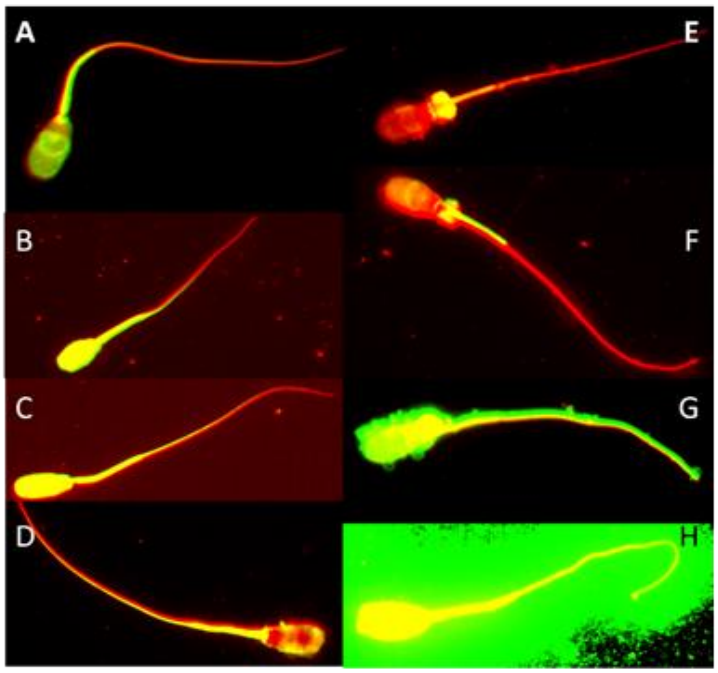


Figure 5-1 Increasing hydrogen peroxide (H_2O_2) concentration on human sperm. (A) No H_2O_2/Untreated sample; (B) 250 µM H_2O_2; (C) 500 µM H_2O_2; (D) 1 mM H_2O_2; (E) 10 mM H_2O_2; (F) 100 mM H_2O_2; (G) 1 M H_2O_2; and (H) 9.8 M H_2O_2. Settings for capturing images: Tritc exposure 2.00 seconds, Gain 6.612, no BGR auto or enhance; FITC exposure 2.00 seconds, Gain 3.718, no BGR or enhance. DAPI exposure 1.825 seconds, Gain 1.00, no BGR and enhance on. Auto was off for all images. (BGR – blue green, red) (Magnification x1000) (Image source: Own images).

Sample/Positive control	µ	SE	σ	IQR	R	N
Untreated samples	1.132	-	0.395	0.374	-	32
250 µM H_2O_2	1.223	0.165	0.484	0.210	0.940	9
500 µM H_2O_2	0.989	0.306	0.379	0.304	0.545	29
1 mM H_2O_2	1.624	0.115	0.944	0.557	0.949	4
10 mM H_2O_2	1.422	-	0.374	0.265	-	2
100 mM H_2O_2	1.473	-	0.049	0.035	-	2
1 M H_2O_2	1.258	0.103	0.696	0.376	0.972	5

Table 5-1 The untreated human optimisation samples and relative positive control statistics are shown in the table indicating the mean (µ), standard error (SE), standard deviation (σ) and interquartile range (IQR). The correlation (R) between the untreated sample and the hydrogen peroxide treatment on those samples, is also shown. Where no value is indicated, there were not enough samples to form a reliable correlation. The number of samples tested for a specific positive control (or untreated) is found in the last column (N).

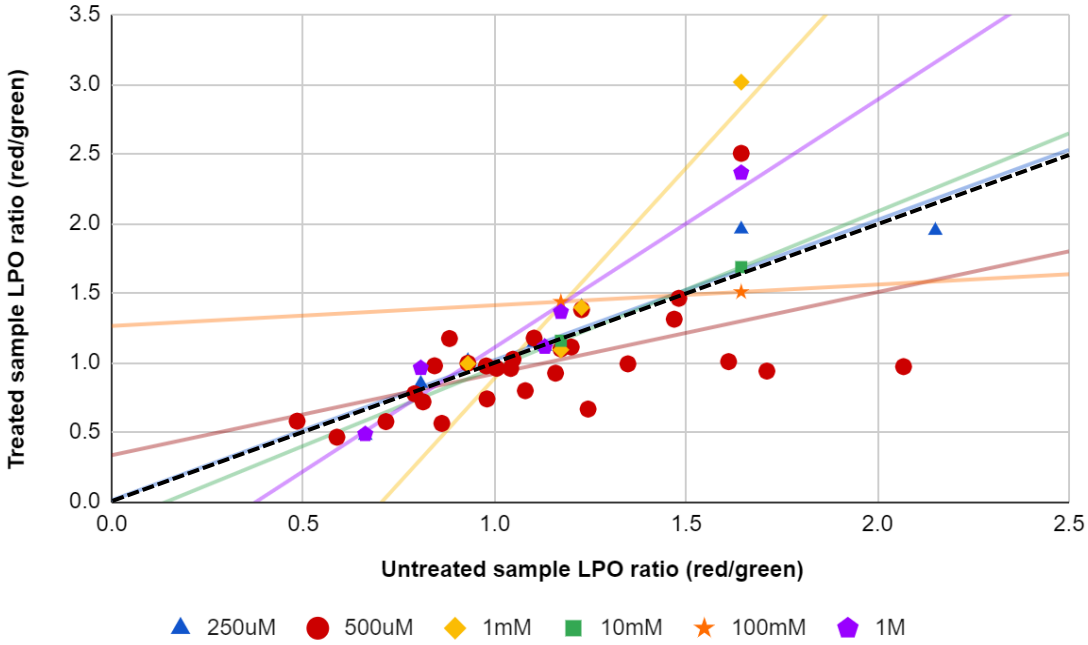


Figure 5-2 Scatter plot of the untreated human sample lipid peroxidation (LPO) ratios (red/green) (x axis) versus the LPO ratio of the same samples treated with various concentrations of hydrogen peroxide (H_2O_2) (y axis). Trend lines for each of the concentrations are shown, relative to the colour of each H_2O_2 concentration. The black dotted line is a threshold, below which, treated sperm should have a reduced LPO ratio value than their respective untreated sperm, and thus, is doing what we expect, and damage is being induced. uM: µM.

However, working on the assumption from the literature, that a decrease in the LPO value (due to an increase in green fluorescence), is expected with more LPO damage; it is assumed that with increasing H_2O_2, the LPO value should decrease (an inverse relationship). Healthy sperm should have a higher LPO ratio than the treated samples. It seems there was not enough H_2O_2 to induce damage with the 250 µM due to the overall scattering of the sample values in Figure 5-2, while the 500 µM H_2O_2 clearly induced damage to almost all of the untreated samples, as most of the values were below the 1 LPO ratio threshold (Figure 5-2). No other H_2O_2 concentrations had the majority of their treated sample values below this threshold. It is assumed that as increasing H_2O_2 concentrations did not increase damage, due to the sperm being already highly damaged and consequently cannot be pushed further past this point, or the contents of the sperm cell (as seen in Figure 5-1) is lost. Because of this, 500 µM was selected as the optimal H_2O_2 positive control. It is assumed that past 500 µM H_2O_2 there is an excess of hydrogen peroxide.

Additionally, the lowest positive correlation was found between the 500 µM and the sample LPO values (R = 0.545). If the correlation is too high between the LPO value and the positive control, it is difficult to determine how damaged the sample is. Overall, a decrease in the LPO ratio is assumed to be associated with more damaged sperm and 500 µM H_2O_2 was selected as the optimal positive control for further testing.

An analysis of variance (ANOVA) was conducted between the LPO sample values (untreated samples), 250 µM, 500 µM and 1 M concentrations, as these were the only samples with 5 or more data points for each variable; yet no significant difference was found for the groups, making the results non-significant at a p-value < 0.05.

5.2.1.2 Sperm concentration and lipid peroxidation sensor concentration

The next parameters to investigate, were the sperm concentration of the sample being put into the flow cytometer (FC), and the lipid peroxidation sensor (LPS) concentration added to the sperm (prior to washing and subsequent testing). When investigating the optimal concentration of sperm for lipid peroxidation (LPO) testing, the flow rate of the sperm through the FC was more important than the actual sperm concentration. For example, it didn't matter what concentration the sperm was, as long as the concentration was creating a flow rate through the FC at a maximum rate of 300 cells/second. This was further validated, as there was no significant difference between the LPO ratio means of the sperm concentrations (µ = 0.876 for 2 million/mL; µ = 0.836 for 5 million/mL; and µ = 0.869 for 10 million/mL). Any sperm concentration which caused a flow rate higher than 300 cells/second created unreliable results i.e., scattering of cells, clogging of the FC and multiple populations of cells. These samples were diluted and retested to ensure accuracy in the results.

When investigating the LPS concentration added to the sperm sample for incubation, to identify the ideal length of time for LPS to saturate the sperm cells, there were nonsignificant differences seen between the various LPS concentration standard deviations of 0.5x (σ =

0.855), 0.2x (σ = 0.115) and 0.1x (σ = 0.468). The concentrations all had standard deviations higher than the LPO sample values tested at the 1x (σ = 0.066) LPS concentration (recommended by the Abcam kit) which means they showed increased variation amongst the samples and their respective repeats. The 0.1x (IQR = 0.391), 0.2x (IQR = 0.081) and 0.5x (IQR = 0.739) LPS concentration interquartile ranges were also slightly higher than the 1x (IQR = 0.054); further indicating a lower variation in the results seen when the patient samples were run with the recommended LPS concentration. Based on the results and commercial protocol recommendations, a 1x LPS solution was decided to be used for future testing. These results are summarised in Table 5-2 below.

Test	Mean	S.d. (σ)	IQR	R	N
1x LPS (recommended)	0.876	0.066	0.054	-	3
0.5x LPS	1.304	0.855	0.739	0.809	3
0.2x LPS	0.947	0.115	0.081	-	2
0.1x LPS	1.054	0.468	0.391	0.915	3

Table 5-2 The lipid peroxidation (LPO) sensor concentration test statistics are shown in the table indicating the mean, standard deviation (s.d.) and interquartile range (IQR). The correlation (R) between the 1x lipid peroxidation sensor (LPS) test and the other concentrations (0.5x, 0.2x and 0.1x), is also shown. Where no value is indicated, there were not enough samples to form a reliable correlation. The number of samples tested for a specific test is found in the last column (N).

5.2.1.3 Incubation time and length of time to lipid peroxidation sperm testing

The Abcam protocol booklet stated a maximum of two hours should be adhered to between addition of the lipid peroxidation sensor (LPS) and testing; as well as the addition of the hydrogen peroxide to testing the positive control. According to the recommended protocol, the incubation time of the hydrogen peroxide for the positive control, is 30 minutes, followed by the addition of the LPS to both the positive control and the sample, which is an additional 30 minutes (total of 1 hour for the positive control and only 30 minutes for the sample). After this, the samples still need to be washed (~15-20 minutes). Therefore, the length of time to testing (including the incubation time and washing time) is short and seems to be a crucial

component of the protocol. This was therefore investigated. Further information on testing can be found in sections 2.8.3 and 2.8.4.

The mean incubation time of the hydrogen peroxide (H_2O_2) with the cells did not have a significant effect on the mean lipid peroxidation (LPO) ratios (Figure 5-3). A small variation in the results was seen, whereby the standardized run samples (normal incubation and washing procedure) had a mean LPO value of 1.281 (n = 3), when compared to the 30-minute (μ = 1.208; n = 2), 1 hour (μ = 1.105, n = 3; standard error = 0.156) and 2-hour (μ = 1.349; n = 2) sperm. A decreasing standard deviation was seen with the increase in time length in H_2O_2 incubation (inverse correlation), with the highest dispersion seen with the untreated samples (σ = 0.381) and the lowest with 2-hours (σ = 0.024). This shows that as time increases, less of a difference between the samples is seen, causing potential differentiation between patients to be difficult. Thus, the recommended 30 minute H_2O_2 incubation time was used for further experiments.

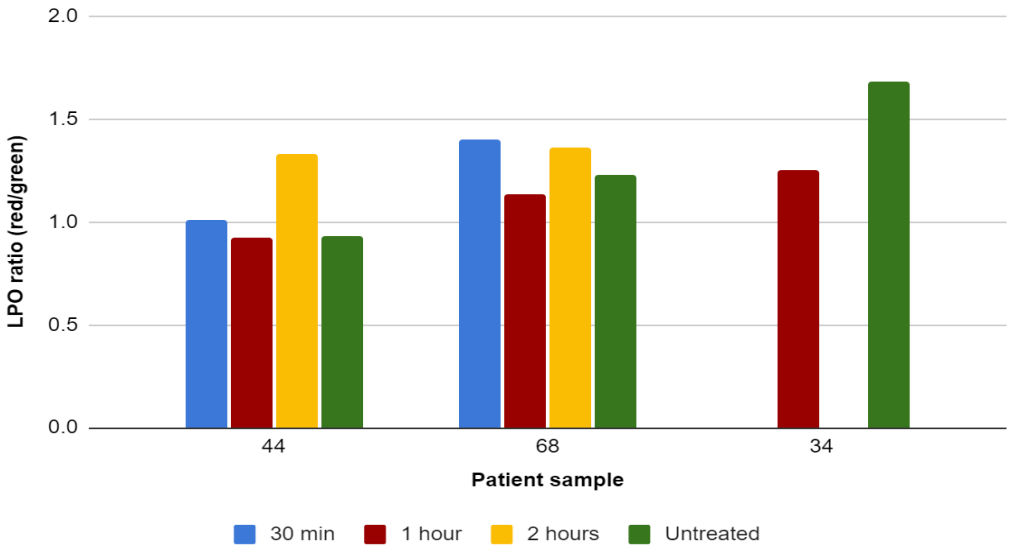


Figure 5-3 The effect of the incubation time of the hydrogen peroxide (H_2O_2) (bars in chart) on patient sperm sample (x axis) lipid peroxidation (LPO) ratios (y axis).

A large discrepancy in the results was found when investigating the length of time after washing the LPS from the samples, to testing (Figure 5-4 and Figure 5-5). The shorter the time frame before testing, the more reliable and repeatable the results were. The cells which were

immediately washed and tested had the lowest mean LPO value of μ = 0.973 (n = 7). The samples which were tested after 30 minutes had a LPO ratio value increase to μ = 1.199 (n = 2). However, if the cells were untested for an hour, the mean LPO ratio value increased to μ = 66.358 (standard error = 0.285; n = 7; p = 0.157). This could be due to two of the samples which had LPO ratio values of μ = 439.833 and μ = 18.601 (excluded from Figure 5-4 and Figure 5-5). Regardless, even though these values are higher (indicating healthier cells), they were more diverse and not repeatable with an IQR of 29.368 for the 1-hour post-washing samples. The sooner after washing the samples were tested, the more accurate the results would be.

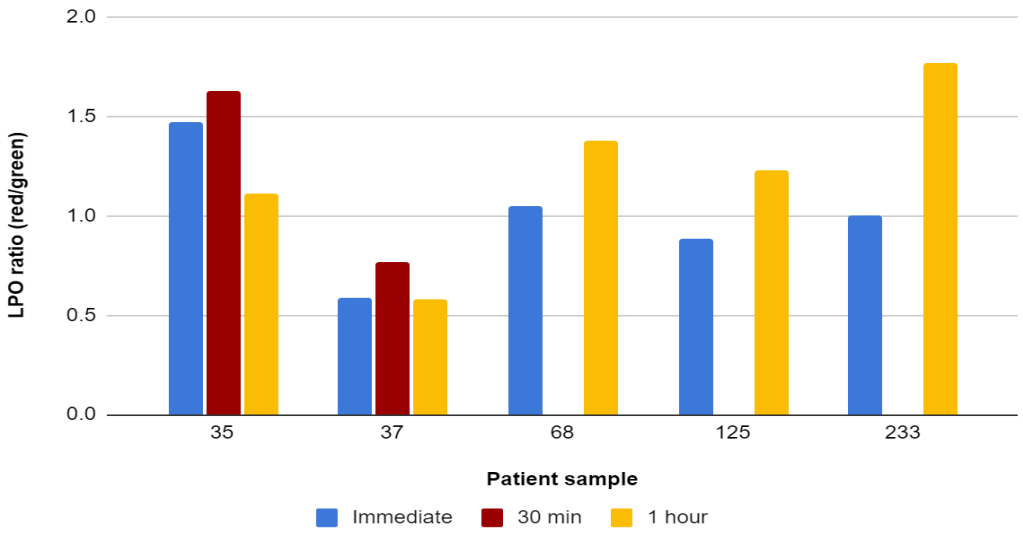


Figure 5-4 The effect of time length between washing the lipid peroxidation sensor (LPS) (bars in chart) on patient sperm sample (x axis) lipid peroxidation (LPO) ratios (y axis).

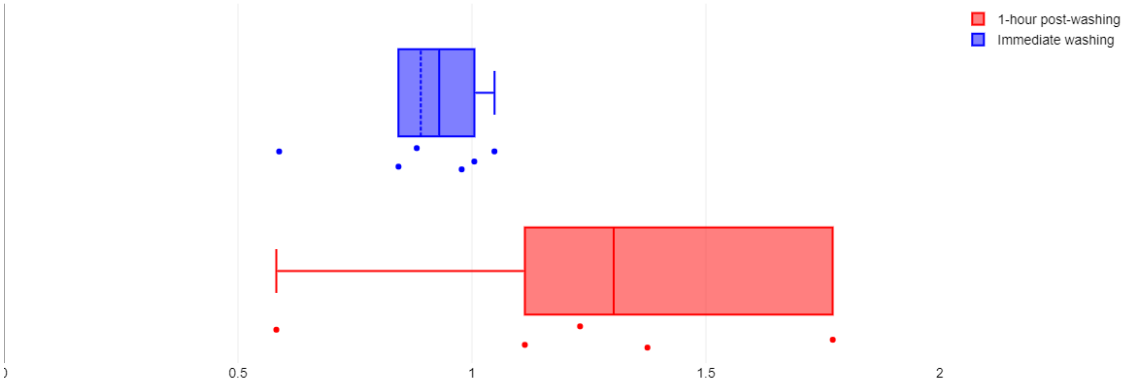


Figure 5-5 Box and whisker plot demonstrating the distribution of the results when testing immediately after washing of the lipid peroxidation sensor (blue) versus testing 1-hour post-washing (red). N = 2 outliers are not demonstrated on the plot for 1-hour post-washing testing. Lipid peroxidation ratio (red to green) values are stated on the x axis.

5.2.2 Specific aim 3b. To assess the association between lipid peroxidation, DNA damage and standardized semen parameters as a suite of tools for fertility screening in humans

Using the testing parameters from section 5.2.1 (500 µM hydrogen peroxide, H_2O_2, positive control for each sample, 1x lipid peroxidation sensor (LPS) concentration, 300 cells/second flow rate, 30 minute H_2O_2 incubation time and immediate sample testing after LPS washing), patient samples were further evaluated. To (1) assess the association between different standardised semen analysis parameters and (2) other sperm tests indicating potential fertility status of a male, as well as the (3) possibility of creating a suite of tools for fertility screening; lipid peroxidation (LPO) (n = 41), thawed sample Sperm Chromatin Structure Assay® (SCSA®) DNA fragmentation index (DFI) (n = 52), oxidation-reduction potential (ORP) (n = 6) and high DNA stainability (HDS) (n = 44) were all tested. A total of 75 patient samples were investigated. Standard World Health Organization (WHO) semen parameter information was available for progressive motility (PR), total motility (TM), concentration, and morphology (n = 15 had all 4 parameters), as well as for fresh sample ORP (n = 6), DFI (n = 67) and HDS (n = 67) from The Doctor's Laboratory (TDL) andrology which allowed for further validation and cross comparison with fresh sample values and the in-house testing. Not all samples had values for all the parameters investigated.

5.2.2.1 Standard semen analysis parameters

The mean testing sample population concentration (n = 75) was identified as 67.238 million/mL which is acceptable according to World Health Organization (WHO) guidelines (WHO, 2010; WHO, 2021). A small population of these samples (n = 15) were tested for progressive, total motility, normal forms, head/midpiece/tail defects, and excess residual cytoplasm. This population of samples primarily had morphology issues including a mean

normal forms value of μ = 2.9%. However, it is known that the threshold for a normal fertile sample is around 4% normal forms (WHO, 2010), thus this value is not unreasonable, yet it is still lower than the general population. Head defects were the most prevalent in these samples with μ = 97.1%, secondly, midpiece defects at μ = 25%, and lastly mean tail defects, μ = 9.5%. Less than 1% of excess residual cytoplasm was seen in the small population of samples (μ = 0.3%). Although morphology was an issue, the progressive and total motility of these patients' normal forms was μ = 60.2% and μ = 68.9%, respectively, which similarly to the concentration of the total population, is in accordance with the WHO guidelines (WHO, 2010). A wide variability of results was seen for the overall patient samples, for example the lowest concentration seen was 0.2 million/mL and the highest, 400 million/mL. The distribution of sample results for concentration, progressive motility (PM) and total motility (TM) are summarised in Figure 5-6 and Table 5-3 (section 5.2.2.2).

5.2.2.2 DNA damage and high DNA stainability

The mean Sperm Chromatin Structure Assay® (SCSA®) DNA fragmentation index (DFI) and high DNA stainability (HDS) values produced in-house were found to be μ = 21.6% and μ = 10%, respectively. The mean DFI was considered to be of good to fair sperm DNA integrity (>15% to <25% DFI), while the HDS was less than the 25% threshold value, above which the number of sperm with immature chromatin and abnormal protein levels will directly negatively affect pregnancy outcomes (Table 5-3, section 5.2.2.3). A high positive correlation was found between The Doctors Laboratory (TDL) and in-house testing results for DFI (R = 0.908) and HDS (R = 0.888). There was no significant difference found between the different means for DFI (p = 0.068) and HDS (p = 0.499) which indicates the in-house testing results could be used for commercial application (Table 5-4, section 5.2.2.3).

5.2.2.3 Oxidation reduction potential and lipid peroxidation

The mean thawed sample oxidation reduction potential (ORP) value was found to be μ = 37.933 mV/mil which is higher, yet not significantly different from the mean fresh sample ORP value of μ = 17.543 mV/mil (p = 0.110) (Table 5-3). Even though there was a larger mean difference for the thawed samples, there was a significant high positive correlation of R = 0.946 (p = 0.001) (Table 5-4). Thus, similar to the DNA damage testing, the thawed ORP values would be able to be tested in-house without prior testing at The Doctors Laboratory (TDL).

Although there was a non-significant negative correlation between ORP and lipid peroxidation (LPO) (R = -0.482), and only three samples were tested, it is promising to find this relationship between the two parameters, as ORP causes oxidative stress (OS) and LPO. Thus, it could potentially be indicative of an inverse relationship (increased ORP shows a decreased LPO ratio) which would need further research. As there is almost no relationship between fresh sample DNA fragmentation index (DFI) and thawed sample LPO (R = -0.060), we can assume LPO and DNA damage are independent of one another.

	C (M/mL)	PR (%)	TM (%)	NF (%)	DH (%)	DM (%)	DT (%)	ERC (%)
μ	67.238	60.2	68.9	2.9	97.1	25.0	9.5	0.3
σ	65.432	15.6	12.7	2.5	2.5	7.9	8.4	0.5
IQR	74.650	14.5	14.5	2.5	2.5	6.0	7.0	1.0
N	75	15	15	15	15	15	15	15

	TDL ORP (mV/mil)	IH ORP (mV/mil)	TDL DFI (%)	IH DFI (%)	TDL HDS (%)	IH HDS (%)	LPO (R/G)
μ	9.443	17.543	17.9	21.6	10.0	10.0	1.595
σ	19.721	26.399	12.6	13.8	7.4	7.8	2.797
IQR	3.931	16.743	16.0	17.8	6.0	5.0	0.587
N	12	6	67	52	67	44	41

Table 5-3 The mean (μ), standard deviation (σ), interquartile range (IQR) and number (N) of the human samples are shown for each of the parameters tested. IH: In-house; TDL: The Doctor's Laboratory; C: Concentration; M/mL: Million sperm/mL; PR: Progressive motility; TM: Total motility; NF: Normal forms; DF: Defects head; DM: Defects midpiece; DT: Defects tail; ERC: Excess residual cytoplasm. ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm. DFI (%): DNA fragmentation index as measured by the Sperm Chromatin Structure Assay® (SCSA®). HDS (%): High DNA stainability as measured by SCSA®. LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio.

The mean LPO ratio was found to be μ = 1.595, with one outlier sample which had a mean LPO value of 18.878. A wide dispersion of the results was seen of σ = 2.797, and the most damaged sample had a LPO ratio value of 0.484. A IQR was seen of 0.587, which means there isn't much variation in the results around the mean, if compared to other test IQR's. Yet, when compared to the mean LPO ratio, it is quite a large value and thus differentiation between patient samples can be identified and therefore, the LPO assay could have potential as a screening test once exact parameters for poor, average, and high LPO are set, and further samples are tested to validate the results (Table 5-3).

		In-house			
		ORP (mV/mil)	**DFI (%)**	**HDS (%)**	**LPO (R/G)**
TDL	**ORP (mV/mil)**	0.946 (6)* p = 0.004	-0.478 (10)	0.656 (10)* p = 0.039	-0.482 (3)
	DFI (%)	-0.780 (4)	0.908 (47)* p < 0.00001	-0.089 (47)	-0.060 (36)
	HDS (%)	0.002 (4)	-0.095 (47)	0.888 (47)* p < 0.00001	-0.072 (36)

Table 5-4 Correlations between the in-house human sperm testing results and TDL testing results. Values in brackets are the number of samples which were tested for both parameters. ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm. DFI (%): DNA fragmentation index as measured by the Sperm Chromatin Structure Assay® (SCSA®). HDS (%): High DNA stainability as measured by SCSA®. LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Significant correlation at p < 0.05.

5.2.2.4 Associations between different parameters tested

Six significant correlations were found between the various parameters. Both in-house DNA fragmentation index (DFI) and high DNA stainability (HDS) were found to be significantly moderately inversely correlated with concentration (DFI: R = 0.304, p = 0.032, n = 50; HDS: R = -0.475, p = 0.0005, n = 50, respectively). In-house HDS had a high negative correlation with the percentage of normal forms (R = -0.707; p = 0.049; n = 8) and was strongly positively correlated with defects of the midpiece (R = 0.790; p = 0.02; n = 8) (Table 5-5). Both DFI and HDS were found to have significant moderate correlations with the lipid peroxidation (LPO) ratio (DFI: R = -0.380, p = 0.042, n = 29; HDS: R = 0.412, p = 0.026; n = 29, respectively) (Table 5-6).

	In-house			
	ORP (mV/mil)	**DFI (%)**	**HDS (%)**	**LPO (R/G)**
Conc (million/mL)	-0.746 (6)	0.304 (50)* p = 0.032	-0.475 (50)* p = 0.0005	-0.133 (40)
PR (%)	-0.966 (3)	0.001 (8)	-0.300 (8)	-0.023 (7)
TM (%)	0.144 (3)	-0.086 (8)	0.034 (8)	-0.206 (7)
Normal forms (%)	-0.875 (3)	0.624 (8)	-0.707 (8)* p = 0.049	-0.207 (7)
Defects mid (%)	-0.053 (3)	-0.184 (8)	0.790 (8)* p = 0.020	0.110 (7)
Defects tail (%)	-0.288 (3)	0.314 (8)	0.636 (8)	0.569 (7)

Table 5-5 Correlations between standard semen parameters and the in-house testing results for the human patient samples for the human samples. Values in brackets are the number of samples which were tested for both parameters to create the respective correlations. ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm. DFI (%): DNA fragmentation index as measured by the Sperm Chromatin Structure Assay® (SCSA®). HDS (%): High DNA stainability as measured by SCSA®. LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Significant correlation at $p < 0.05$.

	ORP (mV/mil)	**LPO (R/G)**	**HDS (%)**
DFI (%)	-0.568 (6)	-0.380 (29)* p = 0.042	-0.077 (44)
HDS (%)	0.412 (6)	0.412 (29)* p = 0.026	-

Table 5-6 Correlations between the in-house testing parameter results. Values in brackets are the number of samples which were tested for both parameters. ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm. DFI (%): DNA fragmentation index as measured by the Sperm Chromatin Structure Assay® (SCSA®). HDS (%): High DNA stainability as measured by SCSA®. LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Significant correlation at $p < 0.05$.

Interestingly, it seems that thawed LPO does have a significant effect on thawed sample DFI based on the results in Table 5-6 ($R = -0.380$, $p = 0.042$) and thus LPO could potentially have an influence on DNA fragmentation during the freeze-thawing process. Even though correlations were high between in-house oxidation reduction potential (ORP), and semen analysis parameters such as concentration ($R = -0.746$), PR ($R = -0.966$) and normal forms ($R = -0.875$), these were not statistically significant due to the low numbers of samples tested. This was also seen with moderate correlations between ORP and DFI ($R = -0.568$), ORP and HDS ($R = 0.412$), DFI and normal forms ($R = 0.624$), HDS and tail defects ($R = 0.636$), and lastly LPO and tail defects ($R = 0.569$).

5.2.3 Specific aim 3c. To briefly evaluate lipid peroxidation and DNA damage testing as an effective method of flow cytometric fertility screening in stallions

Lipid peroxidation (LPO) and DNA damage (DNA fragmentation index, DFI) are not routinely investigated in stallions as a potential source of fertility issues, even though this has extensively been researched in the literature (Section 1.4.5 and 5.1.3). DFI is often examined with either the Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) or Sperm Chromatin Structure Assay® (SCSA®) assays, yet TUNEL identifies actual DNA damage while the SCSA® looks at chromatin integrity (Henkel *et al.*, 2010). However, in andrology they are often used interchangeably. Therefore, we briefly tested LPO, and both DFI assays, and compared them with normal semen analysis parameters such as motility (progressive, non-progressive and total), pH, concentration, and vitality. These standardised parameters are routinely used to define a stallion's fertility potential and thus, direct comparison with them, would give an indication of the effectiveness of DNA damage testing using both the TUNEL and SCSA® assays, and LPO kit.

20 horses were evaluated for different semen parameters with each going through different tests. Seven stallions were investigated for their LPO information (Horses A and O-T). These stallions had 16 different samples available for testing. The mean overall LPO ratio for the 7 horses (n = 12, as four samples were repeats, for example two frozen straws for one collection for horse A, the data per stallion was averaged per collection date, section 2.8.4) was $\mu = 1.086$ and standard deviation, $\sigma = 0.215$. Mean individual stallion LPO ratio values ranged from 0.778 (poor sperm) to 1.455 (healthy sperm), whereby the two highest LPO ratios both had two populations of cells each (horses Q and T). This was unexpected and was not due to the flow rate (which was proven in section 5.2.1 to be a possible reason for doublets), indicating sperm which have potentially already undergone capacitation or acrosome reaction. Another theory

is the samples are broken due to separation of the tail from the head due to centrifugation, fragile sperm, or other external environmental factors, yet this is unclear as was not seen in the FITC vs forward scatter (FSC) graphs (damaged sperm). Based on the use of the LPO kit and lipid peroxidation sensor (LPS) parameters, we assume these populations are still healthy cells, a population which has lower LPO than the other, as the graphs indicating two populations of cells are PE (red fluorescing cells) vs FSC (Figure 5-7).

When investigating the various correlations with LPO ratios, a non-significant high positive correlation was seen between the concentration (million/mL) and the LPO ratio information (R = 0.821, p = 0.179, n = 4) and a non-significant positive moderate correlation between LPO ratio values and pH (R = 0.519) (Table 5-7). As only 4 samples were evaluated for these parameters, it is unlikely any significant correlations would be found, and this needs further investigation.

	ORP (mV/mil)	SCSA® DFI (%)	TUNEL DFI (%)	HDS (%)	LPO (R/G)
Conc (million/mL)	-0.160 (13)	-0.521 (17)* p = 0.032	-0.217 (17)	0.002 (17)	0.821 (4)
PR (%)	-0.295 (17)	-0.470 (17)	-0.330 (17)	0.392 (17)	-0.187 (4)
NP (%)	0.265 (13)	0.394 (13)	0.350 (13)	-0.406 (13)	-
TM (%)	-0.326 (13)	-0.498 (17)* p = 0.042	-0.302 (17)	0.258 (17)	0.216 (4)
pH	0.269 (13)	0.069 (17)	-0.007 (17)	0.073 (17)	0.519 (4)
Vitality (%)	0.290 (13)	-0.762 (25)* p < 0.00001	-0.305 (25)	0.235 (25)	-0.028 (12)

Table 5-7 Correlations between the stallion sperm testing results. Values in brackets are the number of samples which were tested for both parameters to create the respective correlations. Conc: Concentration; PR: Progressive motility; TM: Total motility; NP: Non-progressive motility; ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm; DFI (%): DNA fragmentation index; HDS (%): High DNA stainability as measured by the Sperm Chromatin Structure Assay® (SCSA®); LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Significant correlation at p < 0.05.

DFI and high DNA stainability (HDS) values were collected for all 20 horses (A-T), with a total of 42 samples tested. Samples were combined per collection per stallion, i.e., if there were two straws from one stallion from the same collection, those values were averaged, giving a total of n = 25 data points. SCSA® DFI was significantly highly correlated with both the TUNEL DFI (R = 0.780; p < 0.00001) (Table 5-8) and vitality (R = -0.762; p < 0.0001) (Table 5-7), respectively.

Two significant moderate negative correlations were also found for SCSA® DFI and sperm concentration (R = -0.521, p = 0.032), as well as SCSA® DFI and total motility (R = -0.498, p = 0.042) (Table 5-7).

Almost no correlation was identified between the MiOXSYS oxidation reduction potential (ORP) with SCSA® DFI (R = 0.061) and MiOXSYS ORP with TUNEL DFI (R = -0.079) (Table 5-8). When investigating the association between DFI and LPO; SCSA® DFI had a non-significant low negative correlation with the LPO ratio (R = -0.241), which is similar to what was found between the TUNEL DFI and LPO ratio (R = -0.296), as well as in the humans (section 5.2.2). This indicates there is a connection between lipid oxidative damage and chromatin integrity (DFI), more so than with redox potential.

	ORP (mV/mil) (n = 13)	**LPO (R/G)** (n = 12)		**SCSA® DFI (%)** (n = 25)	**TUNEL DFI (%)** (n = 25)
SCSA® DFI (%)	0.061	-0.241		-	-
TUNEL DFI (%)	-0.079	-0.296		0.780* p < 0.00001	-
HDS (%)	-0.248	0.235		-0.211	-0.054

Table 5-8 Correlations between the stallion sperm testing results. ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm; DFI (%): DNA fragmentation index; HDS (%): High DNA stainability as measured by the Sperm Chromatin Structure Assay® (SCSA®); LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Significant correlation at p < 0.05.

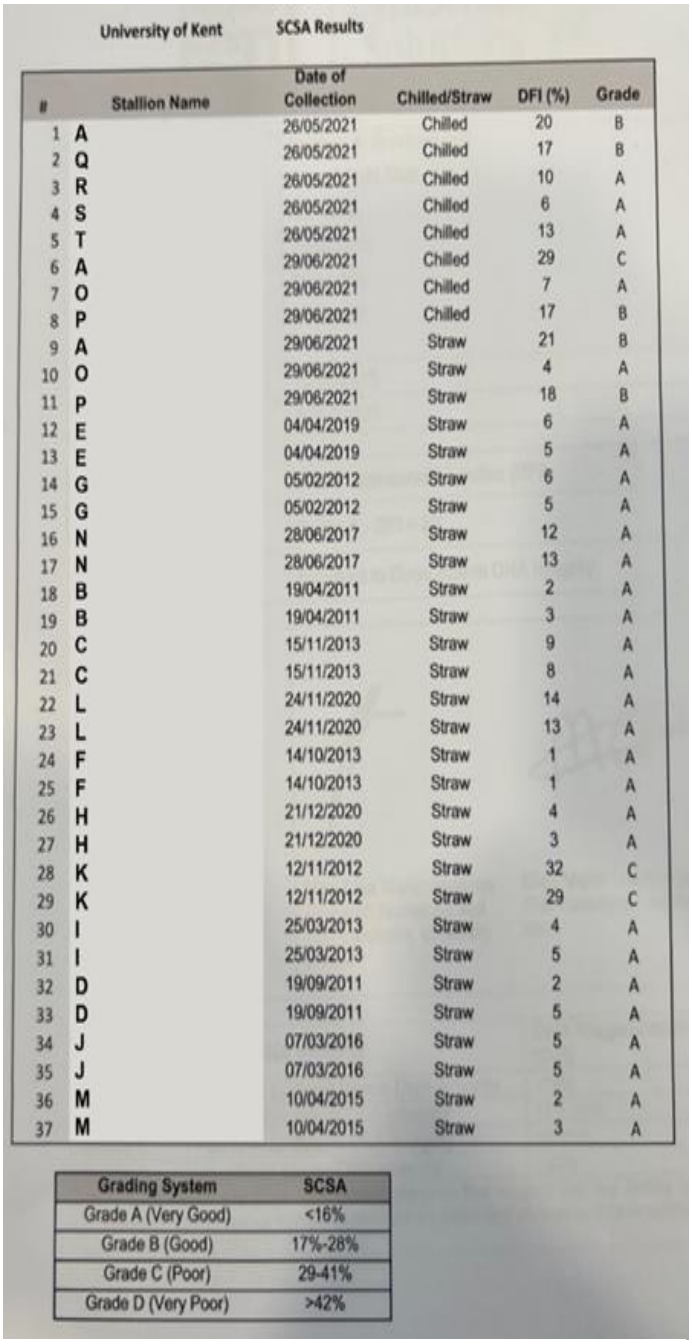
5.2.4 Specific aim 3d. To investigate the effects of (i) chilled versus frozen semen, (ii) good versus bad pregnancy outcomes, (iii) cryoprotectant type, and (iv) breed, on stallion lipid peroxidation values and DNA damage screening

To investigate the effect that different parameters have on the DNA damage values, and lipid peroxidation (LPO) ratios, four different sets of tests were determined. Stallion semen is often sold as either chilled or frozen and thus the effect of these parameters is important to study (Figure 5-8, next page); while it is important to understand the difference between good and bad pregnancy outcome stallion information with regards to determining limits and thresholds in LPO and DNA fragmentation index (DFI) data. It is also known that cryoprotectant type, and breed of stallion, both influence DNA damage, yet not much is known of their effect on LPO, therefore this needed to be studied further.

5.2.4.1 Good versus bad pregnancy outcomes

In this subsection of research, stallions with either “good” or “bad” pregnancy outcomes were tested for standard semen analysis parameters and DFI with both the Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) and Sperm Chromatin Structure Assay® (SCSA®). Six of the stallions had either “Good” (E, F and J; n = 3) or “Bad” (B, G and N; n = 3) pregnancy outcomes, while stallion A is known as having “one of the best fertility rates and pregnancy outcomes” and was used as a model animal in this study (private communication, Stallion AI Ltd.). Stallion A was the only horse in this subsection which had lipid peroxidation (LPO) tested for it, and thus is used as an example, but not included in the group results. As explained in section 2.6.1, bad samples had a poor pregnancy rate which is considered less than 50% success per cycle, and/or less than 75% of pregnancies by the end of

the season, on multiple cycles (per mare). While good samples, have higher success values. This is assuming there were no issues with the mare or external environmental factors.



University of Kent SCSA Results

#	Stallion Name	Date of Collection	Chilled/Straw	DFI (%)	Grade
1	A	26/05/2021	Chilled	20	B
2	Q	26/05/2021	Chilled	17	B
3	R	26/05/2021	Chilled	10	A
4	S	26/05/2021	Chilled	6	A
5	T	26/05/2021	Chilled	13	A
6	A	29/06/2021	Chilled	29	C
7	O	29/06/2021	Chilled	7	A
8	P	29/06/2021	Chilled	17	B
9	A	29/06/2021	Straw	21	B
10	O	29/06/2021	Straw	4	A
11	P	29/06/2021	Straw	18	B
12	E	04/04/2019	Straw	6	A
13	E	04/04/2019	Straw	5	A
14	G	05/02/2012	Straw	6	A
15	G	05/02/2012	Straw	5	A
16	N	28/06/2017	Straw	12	A
17	N	28/06/2017	Straw	13	A
18	B	19/04/2011	Straw	2	A
19	B	19/04/2011	Straw	3	A
20	C	15/11/2013	Straw	9	A
21	C	15/11/2013	Straw	8	A
22	L	24/11/2020	Straw	14	A
23	L	24/11/2020	Straw	13	A
24	F	14/10/2013	Straw	1	A
25	F	14/10/2013	Straw	1	A
26	H	21/12/2020	Straw	4	A
27	H	21/12/2020	Straw	3	A
28	K	12/11/2012	Straw	32	C
29	K	12/11/2012	Straw	29	C
30	I	25/03/2013	Straw	4	A
31	I	25/03/2013	Straw	5	A
32	D	19/09/2011	Straw	2	A
33	D	19/09/2011	Straw	5	A
34	J	07/03/2016	Straw	5	A
35	J	07/03/2016	Straw	5	A
36	M	10/04/2015	Straw	2	A
37	M	10/04/2015	Straw	3	A

Grading System	SCSA
Grade A (Very Good)	<16%
Grade B (Good)	17%-28%
Grade C (Poor)	29-41%
Grade D (Very Poor)	>42%

Figure 5-8 The list of stallions in the specific aim 3 results, with stallion letter allocation (names blanked for privacy reasons), dates of semen collection, whether the samples were chilled/frozen, their DNA fragmentation index (DFI), and grade according to the literature. This list was provided to Stallion AI Ltd as proof of principle. The stallions are all used for commercial breeding purposes (Image source: Own image).

The good samples had a higher mean concentration (μ = 166.616 million/mL), progressive motility (PR) (μ = 90.8%) and total motility (TM) (μ = 96.5%). They also had a lower SCSA® DNA fragmentation index (DFI) (μ = 3.8%), high DNA stainability (HDS) (μ = 1.4%), TUNEL DFI (μ = 0.8%) and marginally lower vitality (μ = 94.1%) mean value than the bad pregnancy outcome horses (Table 5-9). The good samples also had a higher mean oxidation reduction potential (ORP) value (μ = 230.4 mV) which is associated with increased fertility in horses, even though the opposite is found in humans. Stallion A (n = 4 collections) had a slightly below average LPO value in the whole set of horses (section 5.2.3) of μ = 1.009 (overall mean μ = 1.059) yet the highest average SCSA® DFI (μ = 25.497, overall mean μ = 11.683, section 5.2.3) which would normally indicate a poor sample, yet surprisingly it has high success rates with fertilisation and conception. This DFI value is also significantly higher than the good pregnancy outcome horses in this subsection (p = 0.008), making it an outlier and would need further study to understand why it has such as high DFI and average LPO.

PregRate	Conc (M/mL)	PR (%)	TM (%)	ORP (mV)	SCSA® DFI (%)	HDS (%)	TUNEL DFI (%)	Vitality (%)
Good	166.616	90.8	96.5	230.4	3.8	1.4	0.8	94.1
Bad	123.466	82.7	95.3	219.9	6.8	1.9	2.1	94.6

Table 5-9 Stallion pregnancy results and their respective mean semen parameters are indicated above. Conc: Concentration; PR: Progressive motility; TM: Total motility; ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm; DFI (%): DNA fragmentation index; HDS (%): High DNA stainability as measured by the Sperm Chromatin Structure Assay® (SCSA®)

None of the group mean results in this subsection were statistically significant from one another due to the low number of samples tested (n = 3 for good, and n = 3 for bad). Due to this, correlations between the parameters were not calculated and we would need more samples to study and see reliable results. However, based on these preliminary findings, it seems optimistic to relate good pregnancy outcomes with not only improved standard semen analysis parameters, but also lower DFI and increased vitality.

5.2.4.2 Frozen versus chilled stallion semen

We received frozen samples (n = 30) from 16 different horses, as well as chilled samples (n = 12) from 7 different stallions. Both frozen and chilled samples were tested for horses A (2 frozen, 5 chilled samples), O (1 frozen and 1 chilled sample) and P (1 frozen and 1 chilled sample). Each sample had three repeats done on it for each test. As with section 5.2.3, frozen sample values were averaged for each stallion per collection straw when compiling these results (for example, two frozen straws from one collection would end with 6 values per test, all averaged for one stallion), and thus, we ended with n = 16 frozen samples from 16 different stallions. The same was done for the chilled samples, resulting in n = 9 chilled samples from 7 different stallions. The various collection dates for the stallions can be found in Figure 5-8.

When comparing the overall frozen versus chilled sample lipid peroxidation (LPO) ratios, the mean value for the frozen samples (μ = 1.230, n = 16) was higher than the chilled samples (μ = 1.075, n = 9), yet this was not significantly different (p = 0.094). The standard deviation for both the chilled (σ = 0.221) and frozen samples (σ = 0.209) was almost identical which indicates low variation in the data, potentially indicating a high number of fertile stallions in the category. Further work on subsection 5.2.4.1 will allow further differentiation of this data. When further separating the stallions which had both chilled and frozen samples, the mean LPO ratio for frozen and chilled samples was μ = 1.121 and μ = 0.989, respectively (Table 5-10).

	Frozen		**Chilled**	
	All (n = 16)	**Stallions A, O and P (n = 3)**	**All (n = 9)**	**Stallions A, O and P (n = 5)**
Mean	1.230	1.121	1.075	0.989
S.d.	0.209	0.239	0.778	0.094

Table 5-10 Lipid peroxidation (LPO) values for the frozen and chilled samples. Both groups have means and standard deviations (s.d.) for both all the samples in the group, and for stallions A, O and P only. No significant differences were found.

The same number of stallion samples as stated above, were tested for Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) and Sperm Chromatin Structure Assay® (SCSA®) DNA fragmentation index (DFI). The overall mean SCSA® DFI for frozen and chilled samples were μ = 10.5% (n = 16) and μ = 16.8% (n = 9), respectively. The overall TUNEL DFI mean value for frozen samples was similar to the SCSA® at μ = 10.7%, and not significantly

different for the chilled samples either, with μ = 9.5% (Figure 5-9). When looking specifically at stallions A, O and P which had both frozen and chilled samples, both the frozen and chilled samples tested with the SCSA® had higher mean DFI values (frozen μ = 14.4%; chilled μ = 21.1%). The frozen TUNEL DFI mean for these stallions was however lower than the overall mean (μ = 8%), yet the chilled TUNEL DFI mean was higher (μ = 12.6%). As the only DFI value which did not increase, was the TUNEL frozen DFI (specifically when looking at these three stallions), it could indicate there were too few samples in this group to get a reliable result or the horses have healthy sperm. As SCSA® is more sensitive and measures more types of damage, it makes sense the DFI values for this group are slightly higher. Table 5-11 summarizes these results.

	Frozen		**Chilled**	
SCSA®	**All (n = 16)**	**Stallions A, O and P (n = 3)**	**All (n = 9)**	**Stallions A, O and P (n = 5)**
Mean	10.5%	14.4%	16.8%	21.1%
S.d.	9.391	8.940	9.147	9.948
TUNEL				
Mean	10.7%	8%	9.5%	12.6%
S.d.	10.8	6.860	6.6	7.281

Table 5-11 Sperm Chromatin Structure Assay® (SCSA®) and Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) DNA fragmentation index (DFI) values for the frozen and chilled samples. Both groups have means and standard deviations (s.d.) for both all the samples in the group, and for stallions A, O and P only.

The associations between DFI and other parameters was investigated. The SCSA® DFI for the chilled samples, was highly significantly positively correlated with TUNEL DFI (R = 0.916, p = 0.001, n = 9) and vitality (R = -0.934, p = 0.0002, n = 9), and non-significantly strongly negatively correlated with concentration (R = -0.815, p = 0.393, n = 3) and total motility (R = -0.845, p = 0.359, n = 3). Similarly, a high non-significant positive correlation was found between lipid peroxidation LPO and pH (R = 0.941, p = 0.220, n = 3) for these same samples. LPO was also non-significantly strongly negatively correlated progressive motility (R = -0.922, p = 0.253, n = 3) and total motility (R = -0.724, p = 0.485, n = 3), which was an unexpected result as healthy sperm (increased LPO) should have higher motility values. Even though these described correlations were high, they were non-significant due to the low number of samples tested; and thus, these would need further investigation.

However, for the frozen samples, LPO had a strong negative correlation with SCSA® DFI, TUNEL DFI and vitality values (R =-0.970, p = 0.156; R = -0.893, p = 0.297; R = -0.949, p = 0.204, respectively) (n = 16). SCSA® DFI once again had a high significant correlation with TUNEL DFI of R = 0.846 (p = 0.00004), and a significant moderate correlation with vitality, R = -0.521 (p = 0.039). Oxidation reduction potential (ORP) was not tested on the chilled samples, so there is no cross comparison with these results.

For the three horses which had both chilled and frozen samples (A, O and P). In two of the cases (O and P) the frozen sample LPO ratio mean was higher than the chilled samples, μ = 1.377 and μ = 1.082, versus μ = 0.953 and μ = 0.858, respectively, indicating a lower quantity of LPO damage in these cryopreserved samples. One sample however, A, had the inverse occur, possibly implying this sample has more fertilising capacity when it is chilled, and it does not freeze well. This was the case for all three of stallion A's chilled samples which were all collected on different dates. As stallion A has contradictory results with the other samples in section 5.2.4.1, as well as in this section, his sperm should be further investigated on fresh samples to accurately determine the semen parameters of this horse.

5.2.4.3 Cryoprotectant comparison

The effect of the specific protein base (egg or milk) of the cryoprotectant on the standard semen analysis parameters and DNA fragmentation index (DFI) using Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) and Sperm Chromatin Structure Assay® (SCSA®) was briefly investigated. Cryoprotectant data was assumed for 13 stallions (B-O) based on the colour of their sample post-thaw, whereby white would-be milk-based and yellow, egg-based cryoprotectant. As the exact nature of the cryoprotectant is proprietary information, it was not released to us. N = 18 samples were frozen individually in egg-based cryoprotectant and n = 8 individually in the milk-based cryoprotectant. Once again, due to the average per collection per stallion results (sections 2.8.4, 5.2.4.1 and 5.2.4.2), n = 9 egg- and n

= 4 milk-based cryoprotectant values were used in the results. These samples were purposefully selected to be stored in these cryoprotectants based on previous work conducted by Stallion AI Ltd on the individual stallions.

The egg-based samples were found to have better mean results for standard semen analysis parameters, such as increased concentration, progressive motility, total motility, and vitality than the milk-based cryoprotected samples (Table 5-12). However, milk-based cryoprotectants had improved mean results for oxidation reduction potential (ORP), SCSA® DFI, high DNA stainability (HDS) and TUNEL DFI (Table 5-12). These results demonstrate how further sperm testing for parameters (such as DFI, HDS and ORP) could have a major benefit in understanding the fertility status of a stallion. It also shows that even though milk-based cryoprotectants might not demonstrate superior results with basic semen analysis parameter testing when compared to egg-based cryoprotectants, it has better qualities related to protecting the sperm chromatin integrity, increasing the reactive oxygen species (ROS) levels, and reducing DNA damage during thawing. No significant results were found between each of the parameter means when using an unpaired t-test at $p < 0.05$ which is assumed due to the low number of samples.

Cryoprotectant	**Conc (M/mL)**	**PR (%)**	**TM (%)**	**ORP (mV)**	**SCSA® DFI (%)**	**HDS (%)**	**TUNEL DFI (%)**	**Vitality (%)**
Egg (n = 9)	158.85	87.1	95.7	219.2	8.4	1.9	10.0	93.5
Milk (n = 4)	132.15	72.5	90.5	224.7	5.6	1.2	3.3	95.0

Table 5-12 Egg- and milk-based cryoprotectant results. Conc: Concentration in million sperm/mL; PR: Progressive motility (%); TM: Total motility (%); ORP: Oxidation reduction potential (mV); DFI (%): DNA fragmentation index; HDS (%): High DNA stainability as measured by Sperm Chromatin Structure Assay® (SCSA®).

5.2.4.4 Breed Comparison

As with section 5.2.4.3, the effect of the breed as a variable was tested on standard semen analysis parameters and DNA fragmentation index (DFI) using Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) and Sperm Chromatin Structure Assay® (SCSA®). Lipid peroxidation (LPO) was also investigated in this section of work. All 20 horses in the study were included in this subsection. Horses were classed as either Warmbloods (n = 9), crossbreed ("sport") Horses (n = 4), Racing/Endurance breeds, which included one Arabian, one Standardbred and one Thoroughbred (TB) (n = 3), and "Other" as there were no additional samples from those specific breeds in this study. The "Other" breed category included an Appaloosa, Appaloosa Miniature Gypsy, Suffolk Punch and Welsh Section D stallion (n = 4). Samples were averaged per stallion and not per collection as the overall breed effect was studied.

When investigating the semen analysis parameters, sperm concentrations were highest for the crossbreed ("sport") Horses (μ = 162.95 million sperm/mL) and lowest for the Warmbloods (μ = 107.014 million sperm/mL). The Warmbloods also had the lowest percentage of viable sperm in their samples (μ = 72.8%). Vitality of the crossbreed ("sport") Horses (μ = 93.6%) and "Other" breeds (μ = 93.5%) was also significantly higher than the Warmbloods, with p = 0.023 and p = 0.013, respectively. The highest total and progressive motility was seen in the "Other" breeds group (μ = 96.6% and μ = 88.9%, respectively), with a significantly lower mean for progressive motility seen in the Racing/Endurance breeds (μ = 55.5%; p = 0.045).

The Warmbloods also had a significantly higher Sperm Chromatin Structure Assay® (SCSA®) DNA fragmentation index (DFI) (μ = 19.1%) than the "Other" breed category (μ = 3.6%, p = 0.012), as well as the crossbreed ("sport") Horse stallions (μ = 5.1%, p = 0.010). Similarly, Warmbloods had the highest lipid peroxidation (LPO) ratio than all the breed groups (μ = 0.996) and once again, the "Other" breeds had the lowest mean oxidation reduction potential

(ORP) (μ = 217.9mV) and LPO ratio (μ = 0.778) values. These results are summarized in Table 5-13.

Breed	Conc (M/mL)	PR (%)	TM (%)	ORP (mV)	SCSA® DFI (%)	HDS (%)	TUNEL DFI (%)	Vitality (%)	LPO
Warmblood	107.014	76.7	90.4	217.9	19.102	1.290	14.4	72.81	0.996
Crossbreed ("sport") Horse	162.950	87.5	95.5	235.7	5.129* p = 0.010	1.392	4.1	93.91 p = 0.023	0.953
Racing/ Endurance	148.600	55.5	84.5	230.1	8.933	1.011	9.2	93.62	0.778
Other	149.738	88.9** p = 0.045	96.6	208.3	3.596* p = 0.011	1.808	0.7	93.46 p = 0.013	-

Table 5-13 Breed comparisons for various semen parameters and tests. N = number of stallions in that category. Conc: Concentration; PR: Progressive motility; TM: Total motility; ORP (mV/mil): oxidation reduction potential (mV) as measured by MiOXSYS per million sperm; DFI (%): DNA fragmentation index; HDS (%): High DNA stainability as measured by Sperm Chromatin Structure Assay (SCSA®); LPO (R/G): Lipid peroxidation as measure by flow cytometry, red/green ratio. *Indicates mean is statistically different from Warmblood horses using an unpaired t-test at a p-value < 0.05. **Indicates mean is statistically different from Racing/Endurance horses using an unpaired t-test at a p-values < 0.05.

Overall, Warmbloods had the lowest concentration, ORP and vitality, with significantly higher SCSA® DFI than the crossbreed ("sport") horses and other breeds. However, they also had the highest LPO values which are associated with healthy sperm. The racing/endurance horses had the worse semen parameters overall, with significantly lower progressive motility than the other breeds, lower total motility, and low LPO. Due to the relatively high ORP value seen in racing/endurance breeds and the lowest LPO ratio, it could be an indication of prolonged oxidative stress (OS). Even though OS is associated with increased fertility, prolonged OS will cause damage to the sperm. The crossbreed ("sport") horses had arguably the best semen parameters overall, with the highest concentration, high progressive motility and total motility, high ORP (associated with increased fertility in horses) and LPO, significantly lower SCSA® DFI than the warmbloods and the highest vitality. The "Other" breeds category had similar parameter values to the crossbreed ("sport") horses yet had the lowest SCSA® DFI and highest progressive and total motility. Due to the significant results found, further analysis

should be studied regarding the effect of breed on semen parameters as there is a breed effect.

5.3 Discussion

5.3.1 Initial optimisation of a lipid peroxidation test

To test sperm for lipid peroxidation (LPO) accurately, the hydrogen peroxide (H_2O_2) causing the most immediately detectable damage to the sperm was selected at 500 µM. This was the method used by Mesa *et al.* (2016) as low doses of H_2O_2 can cause DNA damage in human sperm due to elevated reactive oxygen species (ROS) and thus 4-hydroxynonenal (4HNE), and its toxicity to sperm has been well reported since 1943 (Williams & Ford, 2005). The pattern demonstrated with increasing H_2O_2 in these results was validated with microscopy images of the sperm which show how LPO gradually increases in the midpiece with increasing hydrogen peroxide, moving to the head, and then from the head to the tail, the midpiece swells and there was an initiation of the acrosome reaction due to the overwhelming level of hydrogen peroxide the sperm has been exposed to. The clear peroxidative damage evident in the midpiece at 500 µM, has been studied in stallions (Ferrusola *et al.*, 2009) and humans (Aitken, 2017).

The concentration of the sample was highly important to the experiment. It is well known that running a sample with a significantly high concentration can cause electronic abortion of an experiment and clogging of the flow cytometer (FC) indicating doublets or false positives (Easthope, 2020). Thus, the concentration of sperm is crucial. Similarly, LPO sensor concentration prior to washing was also important. The lowest IQR, indicating the least variation in the results with that concentration of sensor, was found at the recommended lipid peroxidation sensor (LPS) of 1x (IQR = 0.066). By using a concentration of LPS with higher variation, even if it is slight, means the same sample is demonstrating a wide variation of results which is not ideal for repeatability and consistency (Chon, Dash & Ju, 2009).

Likewise, time sensitivity was highly important in LPO testing. According to the recommended protocol, testing of samples should be conducted within 2 hours of staining the samples. Yet, standard semen analysis should be done within one hour of production (WHO 2010; WHO 2021) and consequently there is a time frame from freeze-thawing whereby the sperm become increasingly damaged by oxidation (Muñoz *et al.*, 2015; Agarwal *et al.*, 2016). This was demonstrated by a decrease in the standard deviation of the samples treated with hydrogen peroxide, as time increased. An inverse relationship in the standard deviation shows the initial testing allowed a differentiation between the samples to be made ($\sigma = 0.381$), whereas after two hours, all the samples were so damaged that there was almost no difference between them ($\sigma = 0.024$). Similarly, the lowest IQR was seen with samples tested immediately (IQR = 0.153) when compared to samples tested an hour after washing (IQR = 29.368), indicating once again more accurate and reliable results with immediate testing (Chon, Dash & Ju, 2009).

5.3.2 Efficiency of lipid peroxidation, oxidation reduction potential and DNA damage screening in humans

According to the World Health Organization (WHO), less than 2 million sperm/mL concentration is considered low (WHO, 2021) while a range of 15-259 million/mL was the WHO 2010 reference range. The mean concentration found for the 75 samples tested was 67.238 million/mL, well within reference range (WHO, 2010). Total motility (60.2%) and progressive motility (68.9%) provided for 15 samples in the population, were also both found to be between the 40-81% and 32-75% WHO 2010 reference ranges for each parameter, respectively (WHO, 2010). Conversely, for the 15 patients with motility information, their morphology was found to have a mean of 2.933 normal forms, which is below the 4% advised by the WHO 2010 guidelines (WHO, 2010). A significant moderate positive correlation was found between LPO and tail morphology ($R = 0.539$, $p = 0.027$) which is a similar finding by Gautam *et al.* (2019) in rat sperm. In their method of lipid peroxidation (LPO) investigation,

they used a Bradford assay which looked for levels of Malondialdeyde (MDA) (Gautam *et al.*, 2019) in sperm. Conversely to tail morphology, LPO did not show any correlations with motility, yet as Rao *et al.* (1998) explains, poor motility is related to membrane fragility in the midpiece and not the tail. This could be the reason there were no strong correlations between the LPO ratio and motility found in these results as the lipid peroxidation sensor (LPS) could be measuring a different parameter than previously thought. Further investigation is needed, as it is unclear what exact parameter of LPO the LPS in this research identifies. Additionally, a LPO ratio IQR of 0.587 and mean of 1.595 were calculated. In the scheme of this test, a large variation of data where the majority of the results are situated, is desired, as it allows the user to differentiate between the samples (Wilcox, 2022).

In addition to LPO, DNA fragmentation index (DFI) is a useful test for evaluating fertility, often debated more so than standard semen parameters (Agarwal *et al.*, 2019; Hamilton & Asssumpção, 2019; Homa *et al.*, 2019; Tanga *et al.*, 2021). The mean DFI in our results (mean = 21.6%) is higher than the average found by Evenson *et al.* (2020) (mean = 17.8%) and less than that of Henkel *et al.* (2010) (mean = 36.4%). As these results were obtained in-house, this is a positive support from the literature. Additionally, the strong positive significant correlation between the DFI of fresh and freeze-thawed semen samples ($R = 0.908$, $p < 0.00001$) further validates the results found in-house and substantiates the literature. DFI is significantly moderately correlated with LPO ($R = -0.380$, $p = 0.046$) which was expected due to the connection between DNA damage and sperm plasma membrane integrity (Piasecka *et al.*, 2007), supplementing the potential for LPO and DFI to be used as fertility indicators in sperm. DFI was also highly correlated with oxidation reduction potential (ORP) ($R = -0.568$), even more so in the small population of men with low morphology parameters ($R = -0.807$). ORP has also previously been defined in the literature as an efficient measure of fertility due to its effect on most parameters including LPO, motility and DFI (Dutta *et al.*, 2019; Homa *et al.*, 2019;

Yurchuk *et al.*, 2021) and has been linked with decreased fertility and poor embryonic development (Williams & Ford, 2005; Gibb & Aitken, 2015).

5.3.3 Stallions

5.3.3.1 DNA fragmentation index (DFI) and Vitality

The mean DNA fragmentation index (DFI) found by Sperm Chromatin Structure Assay® (SCSA®) analysis was $\mu = 11.683$ and $\mu = 8.502$ by Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL), respectively. Both values are within the 4.8-19% range established by Morrell *et al.* (2008) and the mean found by these researchers by SCSA® analysis was $\mu = 11.6\%$, which is the same as in this study, validating our results. SCSA® and TUNEL are both important tests for DNA damage. However, it is debated in the literature whether what component of DNA damage is actually indicated by each test. For example, Evenson (2016) discusses how the high correlation seen between SCSA® and TUNEL DFI results in stallions, indicates how they could actually be measuring the same component of DNA damage, even though in a paper by Henkel *et al.* (2010), they show how the SCSA® measures "potential" DNA damage (susceptibility to damage) while TUNEL measures "real" DNA damage (actual DNA strand breaks). Although the research done by Henkel *et al.* (2010) was conducted on humans, there is a chance that the part of DNA damage being measured, depends on the species investigated, as well as the method used. The results and hypothesis put forward by Evenson (2016) is demonstrated with the results obtained in this research, as a high significant correlation of $R = 0.780$ ($p < 0.00001$) between SCSA® and TUNEL DFI values was found, thus if different parameters were measured, the correlation would surely be lower and less significant. Additionally, the SCSA® is a robust, cost-effective, and accurate test, therefore the high correlation between the two tests presents the opportunity to use only the SCSA® for further testing of stallion spermatozoa DNA damage and DNA integrity as an indicator of fertilising capacity (Wach-Gygax *et al.*, 2017).

A negative moderate, yet significant relationship was found between SCSA® DFI and concentration ($R = -0.521$, $p = 0.032$) as well as with total motility ($R = -0.498$, $p = 0.042$), which is expected as motility decreases as the number of DNA damaged cells increases. Vitality (%) was highly negatively correlated with SCSA® DFI ($R = -0.762$; $p < 0.00001$) which means as the number of live sperm decreases, DFI increases. Thus far, only one study by Varner, Gibb & Aitken (2015) investigated the connection between DFI and vitality specifically, indicating a connection between increased reactive oxygen species (ROS)-induced DNA damage and low vitality values, which is what we found in these results. The concept of ROS and DNA damage has been well studied (Castro, Morales & Parraguez, 2020) and the idea of "live fast, die young" type of notion of stallion sperm was further demonstrated when the chilled and frozen samples were separated (section 1.4.5). As chilled samples still allow the sperm to use up their adenosine triphosphate (ATP) resources when compared to frozen semen samples, where the sperm do not use any energy (Varner, Gibb & Aitken, 2015; Evenson, 2016), it is no surprise that both SCSA® and TUNEL DFI were higher in the chilled samples than the frozen ones. This is however contradictory to what was found by Wach-Gygax *et al.* (2017), where DFI was higher in frozen-thawed samples for specific stallions, which could very well be due to the cryopreservation process which causes DNA damage in equine sperm (Wach-Gygax *et al.*, 2017; Castro, Morales & Parraguez, 2020). This is, however, similar to what was found for stallion A, potentially indicating that this stallion's semen does not freeze well and using chilled or frozen semen is very much stallion dependent for improved fertility outcomes.

5.3.3.2 Oxidation reduction potential, lipid peroxidation, and motility

As previously mentioned, cryopreservation increases the generation of reactive oxygen species (ROS), which causes the production of 4-hydroxynonenal (4HNE), thereby promoting senescence, reduced motility due to mitochondrial degeneration, and DNA fragmentation by inhibiting the functional components of the sperm (Ferrusola *et al.*, 2009; Muñoz *et al.*, 2015; Wach-Gygax *et al.*, 2017). Therefore, it was unexpected for DNA fragmentation index (DFI) and

motility parameters, progressive motility (PR) and total motility (TM), to have low correlations with lipid peroxidation (LPO), of R = -0.241 (p = 0.451), R = -0.187 (p = 0.813) and R = 0.216 (p = 0.784), respectively. However, only a small number of samples were tested for these parameters and therefore, even a low correlation could be indicative of a potential relationship. Additionally, these values were skewed by the opposing correlations found between the chilled (R = -0.038, n = 9) and the frozen (R = -0.970, n = 3) samples for the DFI values. It was expected that as the LPO ratio increases, the DFI decreases, especially in the frozen sperm as their metabolism's have been "frozen". Based on the literature and our findings, we assume that if the LPO ratio is high in fresh sperm (healthy), and oxidative stress (OS) and ROS still have an effect on the sperm, a lower LPO (μ = 0.989) and higher Sperm Chromatin Structure Assay® (SCSA®) and Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) DFI (μ = 21.1% and μ = 12.6%, respectively) can be seen. But, when the sample is thawed after being frozen, there is still very little DNA damage in the sperm (SCSA® DFI μ = 14.4%) due to the lipid membrane still being intact (LPO μ = 1.121).

As ROS had a low correlation with LPO, yet DFI has such as high correlation with LPO, it indicates there is a connection between chromatin integrity and LPO damage, more so than with redox potential. This could indicate DFI is being caused by both LPO and ROS in different ways. Additionally, LPO has been known to cause sublethal damage in stallion sperm, so it is no surprise there is such a high correlation between DFI and LPO with such a low number of samples (Ferrusola *et al.*, 2009).

When looking at the breeds, the highest LPO ratio, therefore the least LPO damage, was found in the warmbloods. This is the same as what was found by Wach-Gygax *et al.* (2017). Yet, the Warmbloods had the highest SCSA® and TUNEL DFI values of all the breeds (μ = 19.1% and μ = 14.4%, respectively) as well as the lowest vitality (μ = 72.8%). All the horses except for the racing/endurance breeds had progressive motility values considered to be excellent (>70%), while the racing/endurance breeds only just managed to be classified as good at 55.5% (55-

69%) which was significantly different to the other breeds category (88.9%; p = 0.045) (Brinsko *et al.*, 2000). Although, when looking at the individual horses, the Arabian categorised as a racing/endurance breed, was found to have an average progressive motility value of only 38% which is classified as poor motility (Brinsko *et al.*, 2000). However, it is normal to find significant breed variations in semen characteristics (Dowsett & Knott, 1996). This further necessitates the need to not only potentially standardise breed semen parameters, but also investigate each individual horse's sperm needs especially as stallion A (Warmblood) was found to have high fertility values, yet poor semen characteristics.

5.3.3.3 Semen analysis and stallion fertility

As seen with stallion A, the lipid peroxidation (LPO) for the samples from this stallion were found to be opposite to the other stallions which had both frozen and chilled samples analysed (stallions O and P). This stallion seemed to have better semen analysis results overall with its chilled samples than his frozen samples, further indicating a need for stallion-dependent chilling/freezing and extender type parameters. Ferrusola *et al.* (2009) and Ferreira *et al.* (2018) both stated similar results to these, with levels of LPO after freeze-thawing to be varied depending on the stallion being investigated. Although, it must be added that thawing after cryopreservation seems to cause peroxidative damage further to what the intrinsic levels of a specific stallion may be and this needs to be taken into account (Ferrusola *et al.*, 2009).

Overall, the stallions with "good" pregnancy results and the milk-based cryoprotectant samples had overall semen results better than those of the "bad" and egg-based cryoprotectant semen. Interestingly, even though elevated oxidation reduction potential (ORP) is associated with improved fertility in stallions (Gibb, Lambourne & Aitken, 2014; Gibb & Aitken, 2015; Mesa *et al.*, 2016), and thus should be directly associated with chromatin damage, and consequently increased DNA fragmentation index (DFI), DFI was lower in the "good" and milk-based cryoprotectant stallions when compared to the "bad" and egg-based cryoprotectant samples which is the same as described by Morrell *et al.* (2018) and Ferreira *et*

al. (2018). A similar relationship was seen on an extreme level for stallion A. This conflicting relationship between ORP and DFI in stallions is peculiar and seems to be highly time, stallion, cryoprotectant and temperature dependent. Vitality may be a more accurate way of analysing stallion semen due to the direct correlation with fertile ejaculates, whereby reduced vitality, especially in freeze-thawed samples, indicates more fertile samples, which was also found in this study (Gibb, Lambourne & Aitken, 2014; Gibb & Aitken, 2015).

5.3.4 Limitations and Future Perspectives

5.3.4.1 Humans

A number of factors may have played a role in this study that were not investigated. Male age in recent years has been shown to have a direct positive correlation with DNA fragmentation index (DFI) values, and thus this should have been taken into account as a potential fixed effect (Winkle *et al*., 2009). As with the stallion samples, only a single sample per patient was evaluated both at The Doctors Laboratory (TDL) and in-house and a small number of patients had information on all of the parameters investigated, making the reliability and significance of some of the correlations and values difficult to interpret. However, strong correlations were indeed identified between the main parameters being studied such as lipid peroxidation (LPO) and DFI, proving their relationship to one another (Gibb & Aitken, 2015; Agarwal *et al.*, 2019). An added piece of vital information would have been pregnancy outcomes from the patients tested. This would allow for a connection to be made between the samples tested and the fertility of the patient, however this information is not publicly available due to privacy policies at TDL.

As demonstrated in the initial optimisation of the LPO test, the time between testing samples is another problem which needs resolving due to the sensitivity of the lipid peroxidation sensor (LPS) and the added damage caused by reactive oxygen species (ROS). This could potentially be the factor which restricts this test to be used commercially as a large amount of labour and

time goes into preparing the sample for testing on the flow cytometer (FC). By the time the test is conducted, almost two hours have lapsed, while most semen analysis is conducted within an hour due to the damage the sperm undergoes over time (Evenson, 2016). Bias could be introduced in this manner. Ideally, LPO should be tested on fresh sperm as soon as it has been produced, while samples analysed in this study in-house have been cryopreserved, consequently needing thawing prior to LPS addition, washing and testing; causing further DNA lesions on genes and LPO damage (Gibb & Aitken, 2015). Further optimisation of this protocol with regard to reducing the time to testing could be a method forward.

One of the main positive influences on this study was the provision of human and stallion semen from a specific organization for each species, as this reduces the factors influencing the inconsistencies which can occur with handling semen in different facilities. Semen analysis is understood to be highly regulated through different accreditation bodies in the UK for humans, but this is still a major component lacking in the equine fertility industry. Even though there is uncertainty on the internal and external quality control (IQC and EQC, respectively), batch testing, equipment used, employee training received, and overall testing used for semen and sperm parameters in stallion semen centres, it is surprisingly unclear whether different human andrology centres are conducting the analysis to World Health Organization (WHO) guidelines or International Organization for Standardization (ISO15189) standards which is expected. Regulation of semen analysis in humans and stallions should be further explored to determine whether they are conducting semen analysis to some degree of certainty and accuracy.

5.3.4.2 Stallions

Although the provisional findings regarding general breed specificity and cryoprotectants are promising, the small number of not only horses in each breed and cryoprotectant category, but also the small number of straws evaluated for each horse (one or two straws), plus the lack

of separation between chilled and frozen samples for breed specifically, further adds bias to the results and would need more investigation to consolidate them (Sieme, Katila & Klug, 2004). Additionally, the stallions in this study had samples collected at random time points in different seasons and years as well as at varying ages and levels of competition, which all have an impact on not only the semen quality, but also the viability for artificial insemination (AI) and freezing (Dowsett & Knott, 1996; Gottschalk *et al.*, 2016; Wach-Gygax *et al.*, 2017; Wilson *et al.*, 2019; Castro, Morales & Parraguez, 2020). Naturally, horses would breed in the United Kingdom (UK) from March to September due to the increase in day length (Wach-Gygax *et al.*, 2017; Castro, Morales & Parraguez, 2020). A mixed model built with more stallion information and increased semen parameter data including fixed and variable effects might be a more appropriate method to interpret future data to further evaluate the significance of the information found, such as the one conducted by Vincente *et al.*, (2014).

Initial findings regarding the viability of the lipid peroxidation sensor (LPS) for lipid peroxidation (LPO) testing look promising yet further work is needed to identify the exact component of LPO the LPS is measuring. According to Gibb & Aitken (2015), there should be a definite correlation between lipid peroxidation and reactive oxygen species (ROS) generation due to the high level of mitochondrial oxidative phosphorylation (OXPHOS) occurring in the sperm. This will need further study in stallions due to the lack of LPO and oxidation reduction potential (ORP) information in the samples. The LPS is also highly time sensitive and needs to be made with fresh stock for each test, which is in itself time consuming and tedious and thus might not be appropriate for a commercial stud farm to use in their laboratory.

Growth of bacteria, debris, inflammation, and other factors beyond our control which may have negative effects on the samples, also may have had an influence on results before, during and after storage and cryopreservation (Gibb & Aitken, 2015; Boe-Hansen & Satake, 2019). Metabolic by-products could also have accumulated in the samples during incubation, as a lower pH was found at 6.2 while the normal range for raw semen is from 7.2-7.7 (Ball, 2008).

This would need further investigating but could be to do with the extenders and cryoprotectants used.

5.3.5 Conclusion

Due to the influence genetic abnormalities can have on fertility, and the pressure of breeding a viable foal, as well as on couples for a healthy child, combined with the long generation interval of a horse, and heartache experienced by couples unable to conceive, it is crucial for routine cytogenetic and seminal screening to be done. Based on this research, DNA fragmentation index (DFI), oxidation reduction potential (ORP) and vitality should be included as part of a standard semen analysis as a suite of tests and, in the previous chapter, I highlighted that it had been added to Stallion AI's portfolio. Even though basic semen analysis parameters give some information about the quality of a sample, for example many semen samples with high motility and concentration, had low DFI and ORP values, which in most mammals is associated with poor fertility, they do not explain the nature of the subfertility. By using DFI, ORP and LPO, the exact nature of the fertility issue can be deciphered. Horses (μ = 11.7%) have a much lower Sperm Chromatin Structure Assay® (SCSA®) DFI than humans (μ = 21.6%), while the correlation between LPO and DFI for the frozen-thawed human samples (R = -0.380) was much lower than for the frozen-thawed stallion sperm (R = -0.970), which backs up the hypothesis of a "live fast, die young" life that stallion sperm live. Provisional results demonstrate the novelty of creating a test to identify and quantify sperm lipid membrane damage and indicate possible insight into individualised male fertility. Additionally, Stallion AI has included sperm DNA damages screening as part of their directory of services based on the work conducted in this study (Figure 5-10), yet more work is needed on the LPO to ascertain its relevance and viability for commercial testing.

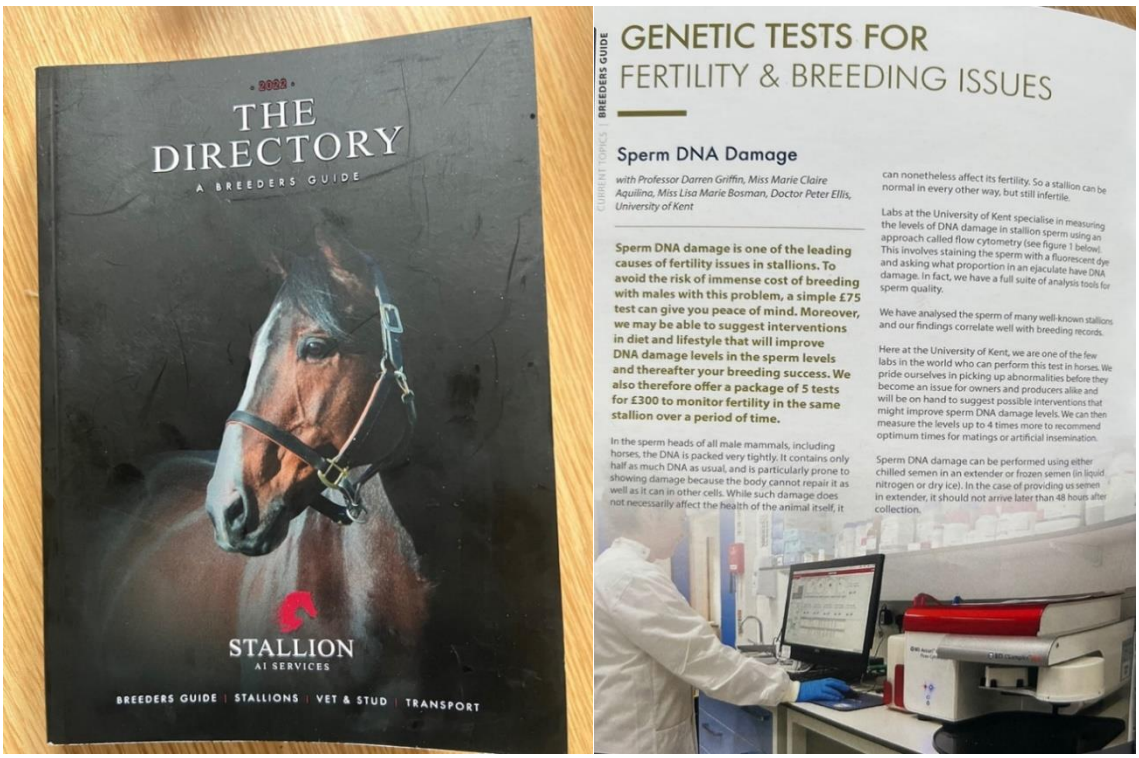


Figure 5-10 Two photos from the 2023 Stallion AI Ltd. *The Directory: A Breeders Guide*; including a piece regarding sperm DNA damage testing as an option for breeders (Image source: Own images).

6. Specific Aim 4 - Regulation of semen assessment in the fertility and pathology laboratory and clinic, based on accreditation: a survey

6.1 Background

6.1.1 Semen analysis

While the first step in introducing a new assay into a clinic is to ensure that it is reliable, practicable and relevant (see previous chapter), its widespread use depends on reproducibility in hands that are initially unfamiliar, and consistent use for the benefit of patients. It is here that agreed standards and their subsequent regulation is important. Currently one of the only methods of investigating male fertility and reproductive health is through standard semen analysis. Male infertility affects around 30 million men worldwide while unexplained male infertility is found in approximately 40% of male patients (Agarwal *et al*., 2015). The male ejaculate containing the seminal plasma and sperm, is studied in a semen analysis, while various parameters and components are measured individually and collectively, to provide a diagnosis and prognosis of a patient's fertility status (Baskaran *et al*., 2021; Tanga *et al*., 2021). For example, by measuring the sperm characteristics such as motility, morphology, and vitality separately to the seminal components of viscosity, pH, appearance and liquefaction, an overall picture of both genetic, sexual and physical health of the male can be seen (Baskaran *et al*., 2021; Tanga *et al*., 2021). Additionally, by examining these parameters independently, it allows the exact identification of the potential problem a patient may have. Other semen parameters such as leukocytes and antisperm antibodies can further give clinicians an indication of male accessory gland infection or inflammation, providing information on potential causes of infertility affected by environmental factors (Henkel *et al.,* 2021; Silva *et al*., 2021).

In the 1930s, semen analysis first became a test to determine male fertility status and is still used as the main method of detection today (Andrade-Rocha, 2017). Therefore, it needs to be reliable, repeatable, and accurate to ensure correct validation of results and for correct treatment to be provided. By missing factors or finding no specific cause of illness or subfertility, various pathologies and problems can be overlooked or misdiagnosed, therefore incorrect treatment could potentially be provided, causing further harm in the long run (NICE, 2014). This has the added stressor of emotional and financial burden on the patients (NICE, 2014) It is consequently crucial for correct determination of a male's fertility status to be made using a semen analysis which has been conducted in a manner corresponding to best practice guidelines and gold standard reference values, such as the World Health Organization (WHO) 2010 guide, while using quality-controlled equipment and procedures (NICE, 2014; Agarwal *et al*. 2022a). By following the International Organization for Standardization (ISO) 15189:2012 standard (ISO 15189:2012) and more recently, the ISO 23162:2021 standard, a semen analysis can provide unbiased, reliable, repeatable, and accurate results while maintaining professional quality assurance (Björndahl *et al*., 2022).

6.1.2 Regulation and accreditation

The United Kingdom Accreditation Service (UKAS) is a national accreditation body that assesses the competency of laboratories that provide diagnostic testing. UKAS oversees the implementation of the International Organization for Standardization (ISO) and will provide accreditation for those laboratories that meet those standards. Accredited laboratories must demonstrate their ability to meet minimum standards as well as adhere to internal as well as external quality control (QC) and must have a quality management system (QMS) in place (Pacey 2010; Agarwal *et al*., 2022a; Björndahl and Kirkman Brown, 2022). Andrology laboratories and clinics are assessed by UKAS using the ISO standards (found in the World Health Organization (WHO) guidelines) before, during and after a standard semen analysis,

thereby ensuring all components are conducted according to standard. However, many pathology facilities do not include the andrology department when applying for UKAS accreditation due to the cost and effort required for little return to the pathology laboratory service, consequently using the money for more widely used services such as cytology and microbiology (Holland & Maddox, 2019). Because of this, many United Kingdom (UK) laboratories carry out semen analysis without any accreditation, making regulation impossible.

In comparison to pathology services, many fertility clinics perform semen analysis in the UK. These centres are required by the Human Fertilisation and Embryology Authority (HFEA) and law to hold a license to practice as a fertility clinic. The HFEA "Code of Practice" is recognized and used internationally as a benchmark for various fertility practices, yet there is no requirement by law to have UKAS accreditation to conduct a semen analysis (HFEA Code of Practice 9th Edition 2021; Morgan, 2004). The main focus of the majority of HFEA licenced fertility clinics in the UK, is female fertility, which has been criticised by Deech (2009) and Tomlinson (2010) to unnecessarily create a barrier for male fertility to be correctly evaluated, often promoting the use of assisted conception treatments for men with substandard semen analysis parameters, instead of further determining the cause of their infertility (Dawson, 1997). These clinics often have no cause to regulate semen analysis, as they promote the use of assisted conception procedures and thus, they are not regulated in this regard. By ensuring semen analysis is standardised across different regulatory and accreditation bodies, male fertility management can be improved, and results can be compared across laboratories, thereby further aiding the understanding and causes of unexplained male infertility (Tomlinson, 2010).

6.1.3 Specific aims

With the above in mind, the aim of this study was to determine the quality of practice for performing semen analysis in various laboratories around the United Kingdom (UK), including both United Kingdom Accreditation Service (UKAS) and non-UKAS accredited laboratories, as well as Human Fertilisation and Embryology Authority (HFEA) licensed laboratories, and to determine whether routine diagnostic testing for male infertility is fit for purpose.

Specific aim 4 was thus to determine whether there are any differences in semen analyses between laboratories and clinics based on the regulatory body with which they are registered, as well as what type of entity they are. In particular:

- **Specific aim 4a.** To evaluate the distribution of laboratories conducting semen analysis in the United Kingdom
- **Specific aim 4b.** To assess the laboratory compliance for various andrology entities
- **Specific aim 4c**. To contrast sperm and seminal fluid parameter assessment between different laboratories and clinics
- **Specific aim 4d**. To define how semen analyses quality control and result reporting is conducted between the various accreditation bodies

6.2 Results

6.2.1 Specific aim 4a. To evaluate the distribution of laboratories conducting semen analysis in the United Kingdom

A total of 106 clinic and laboratory responses were included in this research. The responses were categorized based on their accreditation into Group 1: United Kingdom Accreditation Service (UKAS) only; Group 2: UKAS and Human Fertilisation and Embryology Authority (HFEA); Group 3: HFEA only; and Group 4: neither UKAS nor HFEA (Section 2.10.1). To ensure each

response was in the correct category, they were individually processed. Group numbers were stipulated in Section 2.10.1. Two organizations were at the time of the survey processing their UKAS applications and were consequently moved into Group 1 (UKAS only), while one clinic was CPA accredited, which follows World Health Organization (WHO) guidelines, and was HFEA licensed, hence it was moved to Group 2 (UKAS and HFEA accredited). Lastly, one laboratory had UKAS accreditation for microbiology and not andrology and was subsequently moved to Group 4 (no accreditation). Overall, 38 entities were UKAS accredited (Group 1), 17 had both HFEA and UKAS accreditation (Group 2), 42 had only HFEA licensing (Group 3), and nine were neither (Group 4).

A total of 30 pathology laboratories, 53 fertility clinics and 23 which were both pathology and fertility facilities responded to the survey. Of the pathology laboratory responses, almost all (n = 29) were National Health Service (NHS) government funded organizations while conversely the majority of the fertility clinics were privately funded (n = 31) and a portion were NHS-based (n = 22); 23 of the responses were organizations which had funding from both NHS and private subsidies (Table 6-1).

	Group 1 (n = 38)	**Group 2 (n = 17)**	**Group 3 (n = 42)**	**Group 4 (n = 9)**
NHS Pathology (n = 29)	63% (24)	18% (3)	0% (0)	22% (2)
NHS Fertility (n = 22)	16% (6)	18% (3)	24% (10)	33% (3)
Private Pathology (n = 1)	3% (1)	0% (0)	0% (0)	0% (0)
Private Fertility (n = 31)	5% (2)	35% (6)	52% (22)	11% (1)
Combination of above (n = 23)	13% (5)	29% (5)	24% (10)	33% (3)

Table 6-1 Distribution of laboratories performing semen analysis in the United Kingdom (UK). Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents are shown in each cell, with the number of respondents in brackets. The different answers the respondents could select from is shown in the left-hand column. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation

Other than the one UKAS accredited private pathology laboratory, there is a fairly even distribution within the different funding- and accreditation-type organizations which took part

in the survey. 55 of the laboratories had UKAS accreditation and 51 were non-UKAS while most of the respondents which had UKAS accreditation were part of the NHS and the majority of the HFEA licensed clinics formed part of the private fertility sector.

Table 6-2 describes the purpose of a semen analysis according to the respondent's laboratories and multiple answers could be selected. 100% of the Group 1 (UKAS only) entities and only 76% of the Group 3 (HFEA only) laboratories conduct semen analysis as a diagnostic test for male infertility ($p = 0.006$) while conversely, 18% of Group 1 and 86% of Group 3 laboratories use a semen analysis to select what assisted conception procedure to use for a patient/couple ($p < 0.00001$). Similarly, laboratories are more likely to refer patients for assisted conception procedures if they have HFEA accreditation (Group 2 = 65%; $p = 0.0007$), while laboratories in all the various groups provide post-vasectomy analysis and/or sperm cryopreservation, as well as offer sperm donor workups.

	Group 1 (n = 38)	**Group 2 (n =17)**	**Group 3 (n = 42)**	**Group 4 (n = 9)**
Diagnostic test of male infertility	100% (38)	100% (17)	76% (32)* p = 0.006	100% (9)
To select which assisted conception procedure to use	18% (7)	65 % (11)* p = 0.0007	86% (36)* p < 0.00001	44% (4)
Sperm donor work-up	16% (6)	59% (10)* p = 0.001	50% (21)* p = 0.001	11% (1)
Other (e.g., post vasectomy analysis)	13% (5)	12% (2)	5% (2)	11% (1)

Table 6-2 The purpose of laboratory semen analysis. Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents are shown in each cell, with the number of respondents in brackets. The different answers the respondents could select from is shown in the left-hand column. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation. *Significantly different from Group 1.

6.2.2 Specific aim 4b. To assess the laboratory compliance for various andrology entities

When investigating how laboratories conduct semen analysis and laboratory compliance with best practice guidelines and World Health Organization (WHO) standards (Table 6-3), the majority of Group 1 (UKAS only) laboratories perform and report semen analysis according to WHO guidelines. Conversely, 88% of Group 2 (both UKAS and HFEA) laboratories and only 71% of Group 3 (HFEA only) clinics (p = 0.006) claim to follow these guidelines. However, all Group 1 and 2, as well as 98% of Group 3, laboratories use WHO 2010 reference values for non-diagnostic reasons, even if a large portion do not adhere to the WHO 2010 guidelines. One of the laboratories in Group 4 (no accreditation) uses in-house reference ranges for their semen analysis.

	Group 1 (n = 38)	**Group 2 (n = 17)**	**Group 3 (n = 42)**	**Group 4 (n = 9)**
Q5 Does your laboratory carry out semen analysis AND report ALL parameters strictly according to WHO 2010 guidelines?	95% (36)	88% (15)	71% (30)* p = 0.006	89% (8)
Q6 Do you use WHO 2010 reference values on your semen analysis report?	100% (38)	100% (17)	98% (41)	89% (8)
Q15 Does your laboratory adhere to best practice guidelines i.e., ISO15189 and WHO 2010 criteria	97% (37)	88% (15)	74% (31)* p = 0.003	78% (7)* p = 0.031

Table 6-3 Laboratory compliance questions - 5, 6 and 15. Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents answering "Yes" are shown in each cell, with the number of respondents in brackets. The different questions are shown in the left-hand column. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation. *Significantly different to Group 1.

When looking at International Organization for Standardization (ISO) 15189 best practice guidelines and if WHO 2010 criteria are followed, 97% of Group 1 (UKAS only) follow their recommendations, while only 74% and 78% of Group 3 (HFEA only) (p = 0.003) and Group 4 (no accreditation) (p = 0.031) laboratories, respectively, answered "yes" to this question.

Although the majority of Group 2 (UKAS and HFEA) laboratories were also in agreement with this question (88%), there was no significant difference between this category and Group 3 or 4.

6.2.3 Specific aim 4c. To contrast sperm and seminal fluid parameter assessment between different laboratories and clinics

6.2.3.1 Sperm parameters

The majority of laboratories use a manual method of motility testing, with 82% of Group 1 (UKAS only), 71% of Group 2 (UKAS and HFEA), 90% of Group 3 (HFEA only) and 100% of Group 4 (no accreditation), using this technique. The rest of the laboratories which did not choose manual motility, use computer assisted semen analysis (CASA) as their form of motility examination as well as some using both manual assessment and CASA depending on the sperm concentration in the sample (Group 1: 21%; Group 2: 29%; Group 3: 17% and Group 4: 0%). No significant differences were found between the different groups in this question.

The most reported types of motilities, include progressive (≥65%), non-progressive (≥64%) and immotile sperm (≥62%), but a small proportion of Group 2 (UKAS and HFEA) laboratories also report rapid (53%) and sluggish motility (59%) (Figure 6-1). Group 4 (no accreditation) laboratories also report rapid and sluggish motility at 44% and 33%, respectively, and n = 5 HFEA laboratories, Group 2 and 3, use their own in-house method of reporting progressive motility which consists of scoring out of 4. This method is not recommended by the World Health Organization (WHO) 2010 guidelines and is not used by any Group 1 (UKAS only) laboratories.

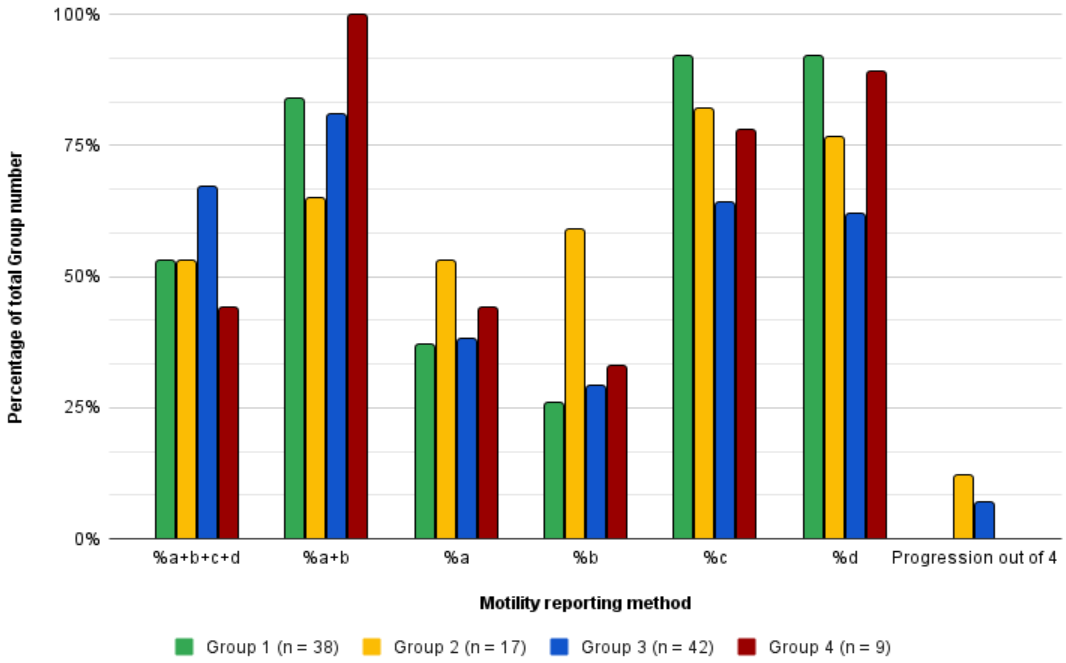


Figure 6-1 How the different Groups report sperm motility using differing techniques. Percentage of total Group number (y axis) is compared to the type of motility reporting method (x axis) for each Group. %a+b+c+d: Total motility; %a+b: Progressive motility; %a: Rapid-progressive motility; %b: Sluggish progressive motility; %c: Non-progressive motility; %d: Immotile. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation

When counting sperm, it is recommended by the WHO 2010 guidelines to be done using an improved Neubauer chamber under fixed conditions, yet there is a significant difference between laboratory methodology (Figure 6-2). Laboratories which are HFEA licenced (Groups 2 and 3) only had 59% and 57% of their total organizations perform this test with correct procedure under fixed conditions, respectively. While 84% of Group 1 (UKAS only) laboratories follow protocol, this result was significantly different to Groups 2 (UKAS and HFEA) ($p = 0.041$) and 3 (HFEA only) ($p = 0.008$), but not to Group 4 (no accreditation) (67%, $p = 0.229$). 95% of Group 1 laboratories use an improved Neubauer to count sperm which was significantly different to the 59% of Group 2 laboratories who use this technique ($p = 0.0009$). A significant difference between Group 1 and Group 3 was also found ($p = 0.0001$), as only 57% of Group 3 laboratories use the WHO 2010 recommended method of counting sperm. Additionally, the majority of Group 4 laboratories use an improved Neubauer to count sperm (78%). Other methods of counting sperm, specifically on motile sperm, used by Groups 2 and 3 include CASA

systems (29% and 17%, respectively), Makler chambers (23.5% and 36%, respectively) and 12% of both Groups use Leja slides or a Cell Vision counter.

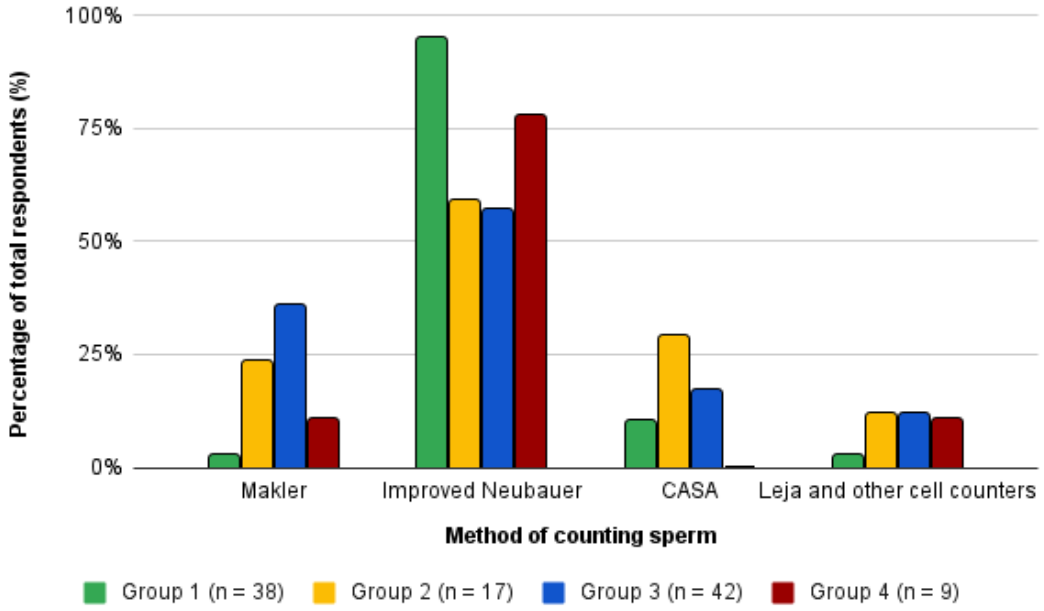


Figure 6-2 Comparison between the different Groups and how they count sperm. Percentage of total respondents (y axis) is compared to the type of counting method (x axis) for each Group. An improved Neubauer is the World Health Organisation (WHO) recommended method of counting sperm. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation

Morphology assessments are conducted under fixed and stained conditions according to the World Health Organization (WHO) 2010 guidelines (Figure 6-3). 97% and 100% of Group 1 (UKAS only), and 89% and 89% of Group 4 (no accreditation), laboratories do morphology testing using these techniques, respectively, whereas only 76.5% of Group 2 (UKAS and HFEA) laboratories fix their samples, while 82% of them stain for morphology testing. Even fewer Group 3 (HFEA only) laboratories follow gold-standard guidelines for morphology analysis (52% and 48%, respectively). Both Group 2 and 3 have significantly different results when compared to Group 1, for both fixing and staining samples (Group 2 fixed $p = 0.013$; Group 2 stained $p = 0.048$; Group 3 fixed and stained $p < 0.00001$). Upon looking further into the specific type of stain used by each laboratory, 89.5% of Group 1, 52.5% of Group 2, 16% of Group 3 and 55.5% of Group 4 laboratories use either Papanicolau, Diff-Quick or Shorr staining for morphology analysis which are the only recommended staining methods by the WHO (Figure 6-4). Group 1

was again significantly different to Groups 2 and 3 regarding these results (p=0.002 and p=0.0008, respectively).

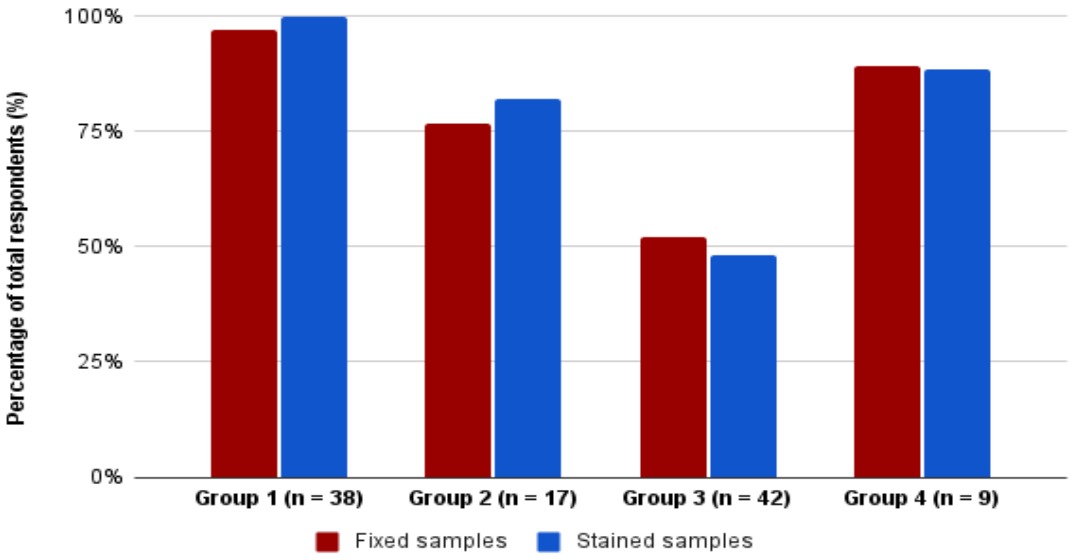


Figure 6-3 Percentage of total respondents in each group which fix and stain their samples for morphology analysis. Both fixing and staining samples are recommended by the World Health Organisation (WHO) guidelines. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation

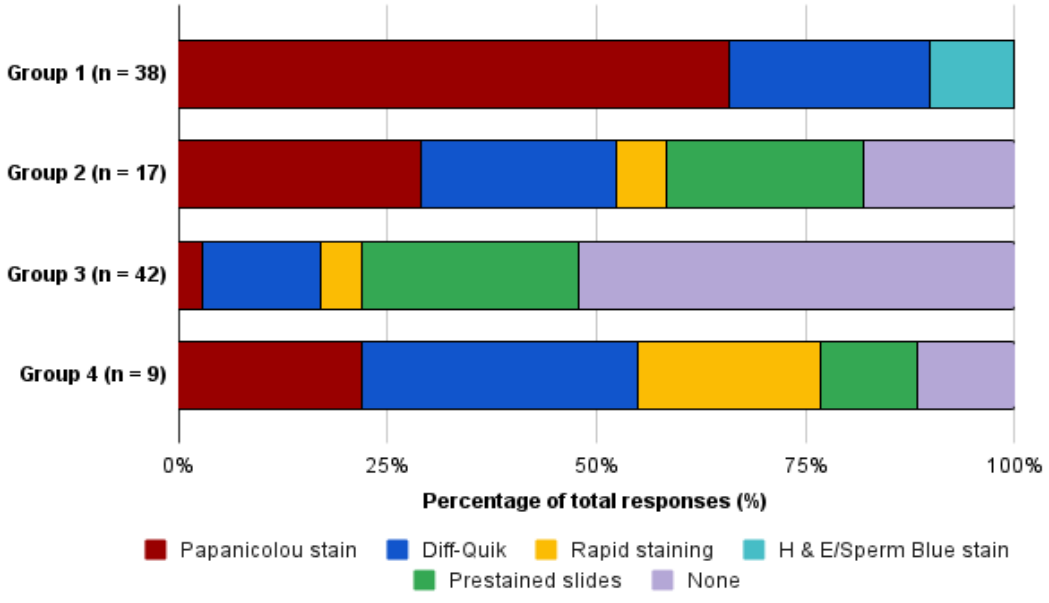


Figure 6-4 Different staining techniques for sperm for morphology assessment done by each group. Papanicolaou stain and Diff-Quick are the recommended World Health Organisation (WHO) standards for morphology assessment. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation

When investigating abnormal forms, individual sperm defect reporting is more common in Group 2 (UKAS and HFEA) and Group 3 (HFEA only) laboratories (76.5% and 71%, respectively) than Group 1 (UKAS only) (58%), while the majority of Group 4 (no accreditation) laboratories also report individual defects (89%). Interestingly, only Group 3 entities report specific abnormality types such as globozoospermia, pyriform and tapered heads, while Group 1 and Group 2 do not report the type of abnormality, only that there are abnormal forms. Some of the comments from Group 1 respondents included that it is not a World Health Organization (WHO) requirement to state exact morphology aberrations or that this type of information isn't useful to users. Two different Group 2 laboratory respondents stated they reporting exact abnormalities was unnecessary as treatment recommendations wouldn't be affected by the individual abnormality, while a different respondent stated all patients with abnormal morphology are always advised intracytoplasmic sperm injection (ICSI) treatment, regardless of the sperm defect type. A Group 4 respondent stated they are only a first line screening pathology service and thus specific abnormalities need to be investigated by a dedicated andrology laboratory and another respondent said something similar, but with referral to a general practitioner (GP) who doesn't need the details.

6.2.3.2 Seminal fluid parameters

In almost every laboratory and clinic, volume and viscosity were assessed, regardless of accreditation (Table 6-4). Appearance, pH and vitality were more important in Group 1 (UKAS only) laboratories (94%, 97% and 71%, respectively) when compared to the others, while liquefaction and appearance were most investigated in Group 2 (UKAS and HFEA) entities (88% and 82%, respectively). Group 3 (HFEA only) laboratories placed more emphasis on liquefaction (88%) and Group 4 (no accreditation) had very varied results. Groups 1, 2 and 3 all placed significance on round cells in the seminal fluid (89.5%, 100% and 90.5%, respectively) and Group 4 had a significantly different result to Group 1 (78%, $p = 0.043$). Interestingly,

Groups 2, 3 and 4 all placed significantly different emphasis on pH when compared to Group 1 (71%, p =0.003; 64%, p = 0.0002; 56%, p = 0.0003, respectively). Additionally, Groups 2 and 3 found peroxidase positive cells or leukocytes (29%, p = 0.036; 31%, p = 0.01, respectively) as well as antisperm antibodies (71%, p = 0.021; 69%, p = 0.004, respectively) to be significantly more important than Group 1 (8% and 31%, respectively). Based on these results, it seems even though Group 1 laboratories are accredited for semen analysis according to ISO15189 standards, many do not see the benefit in investigating seminal fluid parameters and focus more on the spermatozoa.

	Group 1 (n = 38)	**Group 2 (n = 17)**	**Group 3 (n = 42)**	**Group 4 (n = 9)**
Volume	100%	100%	100%	100%
pH	97%	71% *p = 0.003	64% *p = 0.0002	56% *p = 0.0003
Appearance	92%	82%	71% *p = 0.018	56% *p = 0.006
Liquefaction	74%	88%	88%	56% **p = 0.02
Viscosity	97%	100%	90.5%	100%
Vitality	71%	53%	40.5% *p = 0.006	67%
Round cells/other cells	89.5%	100%	95%	78% *p = 0.043
Peroxidase positive cells or leukocytes	8%	29% *p = 0.036	31% *p = 0.01	11%
Antisperm antibodies	37%	71% *p = 0.021	69% *p = 0.004	44%

Table 6-4 The proportion of laboratories reporting additional seminal fluid parameters. Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents answering "Yes" are shown in each cell, with the number of respondents in brackets. The different parameters which could be selected are shown in the left-hand column. Dark green (100-80%), light green (79-60%), yellow (59-40%), orange (39-20%), red (19-0%). Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation. *Significant when compared to Group 1. **Significant when compared to Group 3.

6.2.4 Specific aim 4d. To define how semen analyses quality control and result reporting is conducted between the various accreditation bodies

	Group 1 (n = 38)	**Group 2 (n = 17)**	**Group 3 (n = 42)**	**Group 4 (n = 9)**
Q16 Does your laboratory take part in the UK NEQAS assessment?	100% (38)	94% (16)	98% (41)	89% (8) *p = 0.038
Q16a. Does your laboratory implement the exact same method of assessment for the patient samples, as they do for the UK NEQAS samples	95% (36)	87.5% (14)	88% (36)	87.5% (7)
Q17 Does your laboratory have internal quality controls in place?	100% (38)	100% (17)	90.5% (38) **p = 0.01	56% (5) *p < 0.001 **p = 0.003 ***p = 0.009

Table 6-5 Quality control in the laboratory. Questions 16, 16a and 17, are found in the left-hand column. Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents answering "Yes" to each question are shown in the cells, with the number of respondents for each in brackets. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation. *Significantly different from Group 1. **Significantly different from Group 2. ***Significantly different from Group 3.

An average of 97% of the survey respondents are National External Quality Assurance Scheme (NEQAS) participants regardless of accreditation (Table 6-5). Although 89% of Group 4 (no accreditation) participants take part in the survey, this was found to be significantly different from Group 1 (UKAS only), in which all respondents took part (p = 0.038). Another surprising finding was even though most of the laboratories take part in the survey, not all of them perform the same method of semen analysis as in the NEQAS survey for patient analysis, therefore making this external quality control program meaningless. Most of the laboratories which take part in NEQAS, but do not implement the same methodology in-house, use computer assisted semen analysis (CASA) systems for semen analysis. As you cannot use CASA for fixed samples, they only do fixed sperm counting for the external quality assurance program. It was also clear from the survey, that different groups implement quality control in different ways. Groups 1 and 2 (UKAS associated) all implement internal quality control schemes, yet significant fewer Group 3 (HFEA only) and 4 laboratories have these systems in

place (Group 3: 90.5%, compared to Group 1, p = 0.01; Group 4: 56%, compared to Group 1 p < 0.00001, and compared to Group 2, p = 0.003, respectively). Various Group 4 laboratories who do not conduct internal quality controls (IQC), mentioned they lack time, staff and/or resources to conduct IQC and use the NEQAS samples as their mode of quality control (QC).

For patients with semen parameters outside of reference ranges, the laboratories were asked what they suggest for those patients (Table 6-6). It was clear that Group 3 (HFEA only) laboratories and clinics comment the sample is only suitable for intracytoplasmic sperm injection (ICSI) treatment (64%) while not a single respondent from Group 1 (UKAS only) would advise this (p < 0.00001). A small proportion of Group 2 (UKAS and HFEA) (29%) and Group 4 (no accreditation) (22%) laboratories would also send their patients directly for ICSI treatment, which are also significantly different to Group 1 (p = 0.003 and p = 0.04, respectively). Many Group 2, and Group 3, laboratories would also refer their patients to a fertility clinic (23.5% and 36%, respectively) or urologist (35% and 31%, respectively), while the majority of Group 1 respondents would not give a comment on the report (39.5%).

	Group 1 (n = 38)	**Group 2 (n = 17)**	**Group 3 (n = 42)**	**Group 4 (n = 9)**
This sample suitable for ICSI treatment	0% (0)	29% (5) *p = 0.003	64% (27) *p < 0.00001 **p = 0.03	22% (2) *p = 0.04
Referral to a fertility clinic	13% (5)	23.5% (4)	36% (15) *p = 0.04	11% (1)
Referral to a urologist	8% (3)	35% (6) *p = 0.016	31% (13) *p = 0.022	11% (1)
Below/out of parameter range	34% (13)	12% (2)	29% (12)	33% (3)
Repeat analysis	10.5% (4)	23.5% (4)	24% (10)	22% (2)
None	39.5% (15)	18% (3)	7% (3) *p = 0.001	22% (2)

Table 6-6 Patient report comments included, if a value is found outside of the normal range. Groups and their respective respondent numbers are indicated in the top row. Percentages of respondents answering "Yes" are shown in each cell, with the number of respondents in brackets. The different answers respondents could select from are shown in the left-hand column. Group 1: UKAS only; Group 2: UKAS and HFEA; Group 3: HFEA only; Group 4: no accreditation. *Significant when compared to Group 1. **Significant when compared to Group 2.

6.3 Discussion

6.3.1 Concordance with World Health Organization (WHO) 2010 and International Organization for Standardization (ISO) 15189 guidelines

There is a clear difference in the approach of performing a semen analysis between the different accreditation bodies, United Kingdom Accreditation Service (UKAS) and Human Fertilisation and Embryology Authority (HFEA). Laboratories which are UKAS accredited seem to be almost always conduct semen analysis according to International Organization for Standardization (ISO 15189) standards, while HFEA and non-accredited entities are significantly less likely to adhere to these regulations. Although the discrepancy between accreditation bodies has been shown in this research, it is not a new concern. Many studies including Keel *et al*. (2002) in the USA, Alvarez *et al*. (2005) in Spain, Lu *et al*. (2010) in China, Filimberti *et al*. (2013) in Italy and others (Poland: Walczak-Jedrzejowska *et al.,* 2013; Belgium: Punjabi *et al.* 2016; Iran: Ahadi *et al.,* 2019; India: Kale *et al.,* 2022), have demonstrated a lack of coherent standardisation of semen analysis in their various countries with regards to quality control, reporting results and performance. This includes a lack of improvement and progression in the industry.

Almost all of the non-accredited and HFEA licensed organizations state they are conducting semen analysis within the World Health Organization (WHO) standards and using their guidelines, yet a large number of them are using motile samples for sperm counting and morphology testing, as well as using unstained samples or methodology which is not recommended. This could be an indication that these laboratories are either misinterpreting the WHO guidelines or do not understand them, for example, it is impossible according to the protocols to accurately count or address morphology if the sample is not fixed (WHO, 2010). Additionally, it is very difficult to assess morphology reliably without a staining method. In a study performed by Riddell *et al*. (2005), 35 laboratories in the United Kingdom (UK) stated

they were following the WHO guidelines yet only two of them were adhering to them, specifically with regards to the quality control facets. Similarly, in a more recent global study by Vasconcelos *et al*. (2022), almost three quarters of the laboratories worldwide (from 122 different papers) stated they were using WHO guidelines correctly, but in general the concordance with the methodology was poor.

6.3.2 Sperm and seminal fluid parameter reporting

The responses of the survey identified a number of discrepancies with sperm and seminal fluid parameter reporting regardless of licensing or accreditation body they are associated with. For example, many laboratories do not test or investigate individual sperm morphology, especially defects as it is not required of them or their laboratory, even though it is found in the World Health Organization (WHO) guidelines. Other organizations do not test for individual sperm defects as any patients with sperm morphological aberrations are automatically sent for intracytoplasmic sperm injection (ICSI). It is well known that genetic diseases such as globozoospermia, primary ciliary dyskinesia (PCD) and macrocephaly are causes of infertility and cannot be fixed, hence by identifying them immediately, the diagnosis has already been made without needing further investigation (Chemes and Rawe, 2003; Menkveld *et al*., 2007; Gatimel *et al*., 2017). If problems such as these are missed, patient management would be severely affected. Globozoospermia-based natural conception is almost impossible and due to this, both *in vitro* fertilisation (IVF) and ICSI would also have poor fertilisation rates. These patients would need to undergo ICSI as well as oocyte activation for a successful fertilisation (Kochhar & Ghosh 2018; Fesahat, *et al*., 2020). Another sperm characteristic which is not routinely investigated is sperm vitality. As vitality can be affected by both external environmental factors such as a poor lifestyle habits or structural problems which affect motility, it is important to investigate (Pizzol *et al.,* 2021; Agarwal *et al*., 2022b).

On the other hand, seminal fluid parameters are significant indicators of testicular and/or accessory gland inflammation, infection, obstruction, the presence of toxins and potentially an underlying system illness (Ludwig *et al*., 2003; La Vignera *et al*., 2013; Del Giudice, *et al*., 2020; Baskaran *et al*., 2021; Noweir *et al*., 2022). These different markers can help clinicians determine the health and reproductive well-being of a patient, hence why they are recommended by the WHO guidelines (WHO, 2010; WHO, 2021). From the responses, it is clear many laboratories do not investigate them, as once again, they are not required to report them. By testing these additional indicators, clinical infertility causes, and potential genitourinary tract pathological conditions can also be determined (Ludwig *et al*., 2003; WHO, 2010; La Vignera *et al*., 2013; Henkel & Solomon, 2017; Baskaran *et al.*, 2021). It has been hypothesized that as these parameters are often excluded, a diagnosis of "unexplained male infertility" is often reported as an incomplete result. Reporting a full and comprehensive list of the different semen parameters and their respective results, is crucial to diagnosis and prognosis.

6.3.3 Quality Control

Many laboratories are using World Health Organization (WHO) 2010 reference values, but do not follow the same protocols as the WHO guidelines intend, which means there will be a significant variation in the results due to differing methodology and consequently different reference values (Björndahl 2011; Boyd, 2010). Reference values are intended to be used for a specific methodology, and therefore if different methodology is used, new validated and verified reference values need to be developed (Keel, 2004; Lu *et al.,* 2010; Björndahl *et al.,* 2016). It is unclear whether these laboratories did develop their own reference values for internal quality control, yet based on the response answers, this doesn't seem likely.

Another vital quality control is standardisation and continued performance analysis of reagents, staff and equipment involved in semen analysis to ensure repeatable and reliable

results in all different components involved in testing (Agarwal *et al*., 2022a). Internal quality control (IQC) and external quality assurance (EQA) are being brought into question based on the survey answers, specifically between non-accredited and accredited laboratories. It seems the non-accredited laboratories are not performing these important quality assessments in their laboratories. For example, majority of the laboratories, accredited and non-accredited, take part in the National External Quality Assurance Scheme (NEQAS), yet only 56% of non-United Kingdom Accreditation Service (UKAS) accredited bodies actually implement the same methodology and protocols as the EQA. Additionally, not all the laboratories which take part in NEQAS use the same methodology in this scheme for semen analysis on their patients. Without quality control measures, there is increased intra- and inter-laboratory variability, subjectivity, and human error with not only measurements, but also results which could lead to misdiagnosis or poor patient treatment/management (Pacey 2010; Tomlinson, 2010; ; Long *et al*., 2018; Agarwal *et al*., 2022a; Björndahl & Kirkman Brown, 2022). Unfortunately, this also leads to reduced scientific progress with regards to understanding and improving undiagnosed male infertility issues in the industry (Björndahl *et al.,* 2022).

Within the United Kingdom (UK), the National Institute of Clinical Excellence (NICE) regulates medical practice. Specifically, within andrology, the organization recommends the use of the WHO guidelines and reference values based on the WHO's specific methodology for semen analysis (NICE, 2014). By following the protocols and using reference values which correspond to the techniques advised by the WHO guidelines, the results will be accurate, but only if quality control measures are used and implemented (NICE, 2014). Once again, but not following correct semen analysis, reliability is called into question and patient results are inaccurate (NICE, 2014). It is unclear why laboratories in the UK are not adhering to these protocols and methodologies (WHO 2010, WHO 2021) when they are clear, recommended by professional bodies such as NICE, globally used and are considered the gold standard (Björndahl *et al*., 2004; NICE, 2014; Minhas *et al*., 2021;). It could be due to the fact that many

of these laboratories offer different services which dictate why they use specific methodologies of semen analysis. For example, by using computer assisted semen analysis (CASA) in a fertility or pathology clinic, many parameters are given to the user in a much shorter time frame, with less manual labour and for a cheaper cost than if a comprehensive semen analysis was conducted (Björndahl *et al*., 2004) which are clear factors stated by respondents in this survey. Yet, using a technique such as CASA is not recommended by WHO 2010 guidelines due to the different machines available and a lack of standardised reference values across them (WHO, 2010). This lack of standardisation has been a problem for multiple years (Björndahl *et al*., 2004). Many staff also have different skills and training and the expertise and size of the clinic/laboratory, as well as the income made and management of the facility, all contribute to the lack of standardisation and compliance (Pacey, 2010; Tomlinson, 2010). If a facility is small, they may not prioritize external and internal quality controls, instead focusing their resources on solely determining if a sample can be used for *in vitro* fertilisation (IVF) or intracytoplasmic sperm injection (ICSI) rather than determining the root cause for infertility. This was another conclusion supported by the results of the survey. Some laboratories are also performing semen analysis reluctantly, for example pathology laboratories with a focus in cytology or microbiology might offer andrology, but as it is not their main focus, they are not motivated to follow strict protocols and procedures for testing (Holland and Maddox 2019). These laboratories also often don't value using their potentially limited resources on testing which is not their main source of income.

6.3.4 Diagnosis of male infertility

Semen analysis is the foundation of male infertility diagnosis and is currently the only test available for routine testing (Barratt, 2007; Esteves, 2014; Palermo *et al.,* 2014; Agarwal *et al.,* 2021; Barbăroșie *et al*., 2021), therefore it must be done correctly and to the highest standards to ensure accurate results. Developments in the assisted reproduction technology industry

have caused a reduced emphasis being placed on male fertility and finding the route cause, as many clinicians mistakenly think it doesn't need to be further investigated due to the ease of use of ICSI (Carrell & De Jonge, 2016). This undermines the fundamental basis of a semen analysis by identifying the root cause of male infertility (Carrell & De Jonge, 2016). Based on the survey results, this view is demonstrated by a large majority of Human Fertilisation and Embryology Authority (HFEA)-accredited laboratories and clinics, as they do not see a semen analysis as a diagnostic tool for identifying male infertility. If an underlying issue identified by a semen analysis can be treated, it will make the use of assisted conception procedures null and void, benefitting both the patient and the couple financially and emotionally. A correctly conducted semen analysis can also diagnose other markers of general healthy and thus is a moral obligation to perform correctly (Ventimiglia *et al*., 2015; Del Giudice, *et al*., 2020).

6.3.5 Limitations and future research

As with most surveys, they can be subjective and often misconstrued, consequently it is vital to outline this study's limitations and drawbacks for transparency. Firstly, it was unclear how many Group 1 (UKAS only) accredited clinics and laboratories were accredited for andrology or a different division such as histopathology, microbiology etc. Two answers were identified as not being accredited for andrology due to their own statements of this, yet it is unclear how many others there were. This increases the chance for bias and ambiguous results within Group 1 as well as influencing other groups if a laboratory/clinic was in fact in the wrong one. Secondly, as only a portion of the laboratories which were contacted, responded. In addition, laboratories performing semen analysis without a Human Fertilisation and Embryology Authority (HFEA) licence, and where the staff are not members of Association of Reproductive and Clinical Scientists (ARCS) or who do not partake in the National External Quality Assurance Scheme (NEQAS), would not have been contacted at all. Such laboratories are by default less likely to follow World Health Organization (WHO) guidelines which could have introduced bias

into the survey results, favouring those who perform analyses as per WHO guidelines. However, based on our results, this did not seem to be the case. Concurrently, 98% of the clinics and laboratories answering this survey were NEQAS participants, again, adding to the possible prejudice of this study.

With regards to quality control, the survey did not further investigate how internal quality controls (IQCs) are conducted, for example equipment or batch testing of supplies, all of which can impact the measurement uncertainty of a result (Björndahl *et al.,* 2016; Sanders *et al*., 2017; Long *et al*., 2018). The questions did not ask about the size of laboratory and/or clinic which could have an impact on why certain IQC, and external quality assurance (EQA) are used or not used in the first place. Future surveys or research could further delve into the IQC, and methods used in this regard, as well as how laboratories and clinics of various sizes operate on a day to day as well as patient to patient basis.

6.3.6 Conclusion

By performing semen analysis according to World Health Organization (WHO) 2010 guidelines and International Organization for Standardization (ISO) 15189 standards, with correct internal quality control (IQC) and external quality assurance (EQA) in place, accurate and reliable results should be received. This should be mandatory in any laboratory conducting a semen analysis to ensure reduced chance of bias and misdiagnosis of a patient. If this is done incorrectly, patients who have an underlying fertility issue which could be fixed with simple lifestyle changes, are directly referred to expensive and often emotionally draining fertility treatments. There are no other areas of medicine where lack of validated, controlled, and strict protocol adherence is acceptable. Standards of practice should be rigorously applied to semen analysis as with any other part of medicine, regardless of differing laboratory practices and aims. Detrimental implications to male infertility treatment could be caused by a lack of standardised procedures.

7. General Discussion

7.1 Achievement summary

In this study, the following aims were largely successful:

- 48 sequence and conservation score cattle BACs have been successfully used to assist in the production of chromosome-level genome assemblies. Using these BACs, a universally hybridising set of cattle BACs was developed to not only investigate within order Artiodactyl species such as bison, barasingha and sheep, but also out-of-order Laurasiatheria, dog and horse. The potential for further speciation identification and research, especially in poorly studied species such as bison and barasingha, was made possible with these BACs, as macro changes within chromosome X were traced, which could further aid bioinformaticians with sequence assembly of poorly studied mammals. Using the BAC sequences, fractional length relative to the p-terminus (FLPter) values and ideograms, highly conserved genomic regions were mapped specifically for chromosome 2 and X throughout the species studied; and similarly, six highly conserved genes with roles in neuronal development were identified within the BAC sequences. This provides the opportunity for FISH to be used not only for reproductive isolation analysis, but also for population chromosomal screening affecting fertility, based on the work done by O'Connor *et al.* (2017) and Jennings, Griffin & O'Connor (2021).
- A prototype multiprobe chromosomal screening device for horses was developed in this study off the back of the success of both the pig and cattle devices (O'Connor *et al*., 2017; Jennings, Griffin & O'Connor, 2021). Two novel chromosomal abnormalities, 64, X, t(20p) and 64, XX, t(15q; 28q) were identified with this equine device and the exact nature of another novel translocation 64, X, t(10p, Xp) and t(10q; Xq), identified during karyotyping, was established. The device permitted the correction of incorrectly

karyotyped horses and consequently, is a significant addition to the cytogenetic screening service currently available in the laboratory. Further development of this device is necessary to ensure a 100% success rate across all the probes for future possible commercial screening application. It could also be added to a suite of new tools to identify fertility issues in horses, specifically stallions.

- Fertility in both stallions and men is currently assessed using a standard semen analysis, yet this is often not sufficient to describe unexplained male infertility. A commercialised non-gametic flow cytometric assay for lipid peroxidation (LPO), was optimised for sperm, and then tested in both humans and horses to add another potential measure to identify unexplained male infertility. The use of this assay, and DNA damage screening, which is known to not only be caused by reactive oxygen species (ROS), but also create fertility issues, was efficiently implemented to associate the standard semen analysis parameters with the DNA fragmentation index (DFI) and LPO results. Additional research is needed to not only threshold values for the LPO assay, but also determine the underlying quantity of 4-hydroxynonenal (4HNE) in both stallion and male sperm. However, initial results are promising to add to a suite of potential fertility screening tools.
- Regulation of semen analysis and other tests for male fertility in the United Kingdom (UK) is another important aspect to consider, as quality control (internal quality control and external quality assurance) is used in every component of scientific research and commercial and clinical application. However, it is clear from our survey results, that many clinical laboratories are not performing semen analysis according to International Organization for Standardisation (ISO) 15189 standards or following World Health Organization (WHO) guidelines, which is mostly associated with the body with which the laboratory is accredited, United Kingdom Accreditation Service (UKAS) or Human Fertilisation and Embryology Authority (HFEA). This causes the misdiagnosis of patients as the underlying reasons for potential unexplained male infertility are not

diagnosed and treated. Improved standardisation of testing protocols is important to reduce bias as well as potential emotional stress placed on patients.

The overall goal of this research was to study genomic rearrangements and damage, and their connection to fertility as well as reproductive isolation. Through comparative genomics, highly conserved areas of the genome were identified between both closely and distantly related animals, from gross changes all the way down to specific gene conservation in orthologs. BAC clones are a vital tool which we used for not only comparing sequences that may have caused speciation across evolutionary diverse mammals, but also have aided in the detection of chromosomal changes which cause fertility issues in modern species. With these tools, the initial basis of chromosome-level assemblies was developed in poorly studied animals. By using species specific BAC clones, it has been clearly been proven with research in pigs by O'Connor *et al*. (2017), cattle by Jennings, Griffin & O'Connor (2021), and now in horses, that they are valuable tools not only in a research sense, but also as a commercial and modern tool for translocation screening, which has a major impact on fertility and their respective industries, as shown in papers by Lewis *et al*. (2021) and Lewis *et al*. (2022). BAC probes have a place for not only studying the evolutionary patterns of ancestors of modern mammals, but the mammals themselves, and the future offspring of these different species. By investigating chromosomal rearrangements in current animals, reproductive isolation can be mediated to ensure the conservation of a species for future generations.

Another aspect of fertility analysis which utilizes well-established tools such as flow cytometry for both novel research and commercial application, is semen analysis, as well as DNA damage screening and lipid peroxidation (LPO) ratio analysis, to identify unexplained male infertility. It is evident from this research that even though there is an opportunity to use sperm LPO as a potential screening tool in both men and stallions, more research needs to be done to understand and quantify the damage indicated by 4-hydroxynonenal (4HNE) and the lipid peroxidation sensor (LPS). However, DNA damage screening using Sperm Chromatin Structure

Assay® (SCSA®) and Terminal deoxynucleotidyl transferase-mediated fluorescein-dUTP nick end labelling (TUNEL) have both been highly researched and our data validates the use of these two tools to identify potential sources of unexplained male infertility. However, without the regulation of these types of screening assays and services to ensure standardisation and adherence to international guidelines, it is impossible to diagnose patients correctly and provide them with the optimal treatment. The research identified here, demonstrates how internal quality control (IQC) and external quality assurance (EQA) are not followed, and it is highly dependent on the body with which andrology services are accredited with in the United Kingdom (UK). This type of standardisation still needs to be investigated in horses.

7.2 Comparative genomics and speciation

The main aim of any work with genomics or sequence-based technology is to develop a full contiguous sequence from one end of the chromosome to another (Damas *et al*., 2016). However, sequencing and annotating whole mammalian genomes is currently not viable for many scientists in smaller studies such as this one. Based on previous work conducted in the avian project completed by Damas *et al*. (2016), lab research done by Jennings & Griffin (2019), and algorithm selected BACS based on the work done by Kim *et al*. (2013), we attempted to develop a universally hybridising set of BAC clones to map evolutionary conserved regions across phylogenetically different mammals which may have led to their reproductive isolation.

The work described in Specific aim 1 (section 3) describes the successful nature of our findings. As these BACs were selected based on their evolutionary breakpoint regions (EBRs), as well as for conservation score and sequence specification, which are known to be in gene dense regions, both small and gross chromosomal changes were identified (Larkin *et al*., 2009). The use of these BACs allowed for the identification of the chromosomal location where currently unplaced genomic scaffolds, in both pigs and horses, are situated with the use of ideograms,

fractional length relative to the p-terminus (FLPter) values, BLAST sequences and classic cytogenetic tools such as karyotyping. This demonstrates the necessity of physical mapping. Additionally, it is no surprise however, that highly conserved genes (PAX2, ZEB2, COPS9, TENM1, SH2D1A and PCDH9) involved in neuronal development (Dressler & Woolf, 1999; Yan *et al*., 2003; Bruining *et al*., 2015; Alkelai *et al*., 2016; Epifanova *et al*., 2019), were found in the various BAC sequence regions, as well as the identification of their orthologues in all the tested species, which had full genomes. However, in adult mammals, the role of these genes was often quite different. For example, both TENM1 and PCHD9 affect fertility (Strehl *et al*., 1998; Asahina *et al*., 2012; Bruining *et al*., 2015; Serranito *et al*., 2021; Zhou *et al*., 2022), which further expands on how these BACs can identify genes with potential influences on reproductive isolation and fertility potential. As these genes have maintained vital functions to embryonic development, while concurrently changing their main purpose, it reassures the molecular clock hypothesis whereby animals cannot create hybrids due to the change in gene function, resulting in speciation (Ho, 2020; Graves, 2016).

7.3 A suite of fertility tools in the stallion industry

Evidently, BACs not only identify genes situated within their sequences which could be highly conserved through evolution, but also can be used to evaluate gross chromosomal changes such as translocations and inversions. It is well studied, that these types of rearrangements can cause hypoprolificacy and subfertility/infertility in agricultural animals such as pigs (O'Connor *et al*., 2017), cattle (Jennings, Griffin & O'Connor, 2021) and horses (Bugno-Poniewierska & Raudsepp, 2021), which could originally have led to their speciation in the first place. However, these aberrations are costly to their respectively industries (Lewis *et al*., 2021; Lewis *et al*., 2022) and identifying them is crucial to not only breeders, but also for animal welfare and environmental purposes. Detection of the causes of unexplained infertility in the agricultural industry has also been investigated using a standardised semen analysis, even though DNA fragmentation testing (Sperm Chromatin Structure Assay® and Terminal deoxynucleotidyl

transferase-mediated fluorescein-dUTP nick end labelling) and identification of oxidation reduction potential (MiOXSYS) have been proven to have higher correlations with fertility outcomes. Based on these various aspects of fertility, the potential to develop a suite of fertility tools in the stallion industry seems promising and was attempted in this research.

In specific aim 2 (section 4), a prototype equine screening device was developed. Sequence assembly errors were identified in the horse genome with the new probes, whereby subtelomeric probes which were selected for specific chromosomes were found in the incorrect locations, which is what was found by O'Connor *et al*. (2017) in the pig genome, using the same methodology., By using the device, we identified of two novel translocations with chromosomes 5, 10, 15, 20, 28 and X, which were identified in the literature as mostly being unstable in equines (Power, 1987; Lear & Layton, 2002; Brito *et al.*, 2008; Bugno-Poniewierska & Raudsepp, 2021). For example, the Przewalksi's horse (a close relative of the domestic horse) chromosomes 23 and 24 either fused to become equine chromosome 5 or fission occurred in the reverse manner which would indicate instability of this chromosome (Huang *et al*., 2014). Additionally, one of the translocations we identified (chromosome 16) has been associated in a study by Lear *et al*. (2008) with causing repeated early embryonic loss in domestic Thoroughbred mares through the use of FISH BACs, demonstrating the potential importance for these tools in a diagnostic setting. However, until a full panel of BACs, with 100% hybridisation success rate, is developed, karyotyping is vital to identifying potential gross chromosomal changes in horses, even though it can't detect cryptic rearrangements (O'Connor *et al*., 2017). A commercial FISH translocation testing system for horses would be beneficial to the equine industry as there is no service like this currently available worldwide.

Similarly evaluated in Specific aim 3 (section 5), there are no DNA damage, reactive oxygen species (ROS) or lipid peroxidation (LPO) screening services available for horses in the equine trade, even though the individual parameters have been extensively studied (Wach-Gygax *et al*., 2017). The Sperm Chromatin Structure Assay® (SCSA®) is a cost-effective, accurate, robust,

and well researched test in horses, and as our results demonstrated the high correlation between Terminal deoxynucleotidyl transferase-mediated fluorescein-dUTP nick end labelling (TUNEL) and SCSA®, there is the potential to use the SCSA® as a primary form of commercial testing of DNA damage, especially due to its relationship with fertilising capacity in stallions (Wach-Gygax *et al*., 2017). As DNA fragmentation index (DFI) is a direct indicator of DNA integrity, it could be of more use than a standard semen analysis as an indicator of equine fertility and chromosomal stability. ROS (Oxidation reduction potential (ORP) measurement) causes DNA damage (Castro, Morales & Parraguez, 2020), yet stallion sperm have a "live fast, die young" type of existence due to their oxidative phosphorylation (OXPHOS) method of ATP production, whereby the most fertile sperm produce the most ROS (section 1.4.5) (Gibb, Lambourne & Aitken, 2014). This contradictory relationship needs further investigation, as ROS also cause LPO damage (Ferrusola *et al*., 2009; Muñoz *et al*., 2015; Wach-Gygax *et al*., 2017), which has been shown in our study to be correlated with DFI. The LPO test in this research needs further investigation as a commercial assay, yet it seems that the use of ORP (MiOXSYS) and DFI (SCSA®) values, may be the more logical and commercially appropriate testing methods for fertility in stallion, due to their simple, fast, accurate, fertility-correlated, and cost-effective means of assessment. Thresholds also need to be determined for each of the parameters depending on the variables involved such as breed, frozen/chilled or cryoprotectant type, due to their clear influences on stallion semen parameters based on this study. As equine karyotyping is already conducted in the lab, this, in combination with a multiprobe chromosomal screening device and SCSA® (DFI), MiOXSYS (ORP) and LPO screening, could potentially offer a multifaceted range of tools for causes of fertility issues in horses.

7.4 Andrology testing and accreditation

As shown in our research as well as the literature, reactive oxygen species (ROS), and DNA damage both have be evident as contributors to male fertility related issues in stallions, yet this is also the case in humans (Drevet & Aitken, 2020; Aitken, 2017). Although stallions with increased ROS are considered more fertile, humans with more ROS are less so, which once again comes down to the method used for ATP production, whereby humans use glycolysis (Griffin *et al.,* 2019; Peña *et al.,* 2019). DNA damage in men is also highly correlated with subfertility/infertility (Dutta, Majzoub & Agarwal, 2019; Nowicka-Bauer & Nixon, 2020), and consequently it is vital to test for this along with other semen parameters as the level of unexplained male infertility is ~15% globally (Hamada *et al*., 2012). However, lipid peroxidation (LPO) is not commonly tested for even though it has strong links with morphology and motility (Toor & Sikka, 2019), and thus, we aimed to develop a novel LPO test for investigating 4-hydroxynonenal (4HNE) production in patients in Specific aim 3 (section 5). Furthermore, these types of tests are normally conducted as an extension to a standard semen analysis, which is currently the main form of testing men for fertility issues (Agarwal *et al.,* 2019; Agarwal *et al.,* 2022a; Agarwal *et al*., 2022b). However, in the United Kingdom (UK), research-based and commercial/clinical semen analysis is regulated by different bodies such as the Human Fertilisation and Embryology Authority (HFEA) and United Kingdom Accreditation Service (UKAS), but to what extent they are following quality control (QC) regulations such as International Organization for Standardization (ISO) 15189 and the World Health Organization (WHO) guidelines, is unknown. This was investigated in Specific aim 4 (section 6).

The results of this study (Specific aim 3, section 5) demonstrate the potential of using tests such as DNA fragmentation testing, oxidation reduction potential (ORP) evaluation and LPO investigation for future screening in the andrology sector. For example, DNA fragmentation index (DFI) was significantly correlated with LPO, and as DFI is a known test for investigating fertility potential, often more so than a standard semen analysis (Agarwal *et al*., 2019;

Hamilton & Asssumpção, 2019; Homa *et al*., 2019; Tanga *et al*., 2021), our results are validated and further proof of principle. Although this is the case, it is unclear as to what exact parameters the LPS is measuring due to it being Abcam propriety information, and a non-gametic test, and more work is needed on this assay. However, the Sperm Chromatin Structure Assay® (SCSA®) and MiOXSYS were highly successful tests, having correlations with multiple standard semen analysis parameters. As ORP affects and causes LPO and DFI, it may be the more worthwhile indicator of unexplained male infertility which should be investigated, especially due to its links with poor embryonic development (Williams & Ford, 2005; Gibb & Aitken, 2015; Dutta *et al*., 2019; Homa *et al*., 2019; Yurchuk *et al*., 2021).

Nevertheless, if tests such as these are not conducted according to quality standards such as ISO 15189 or WHO guidelines, the results could be misleading, and patients could be misdiagnosed. In Specific aim 4 (section 6), our research demonstrated the clear difference in the approach of performing semen analysis, such as parameter reporting, between HFEA licensed and UKAS accredited clinical laboratories. Unfortunately, this is not the first time this kind of discrepancy has been shown as similar studies worldwide have revealed a lack of adherence to quality control, semen analysis performance and result reporting (USA: Keel *et al*., 2002; Spain: Alvarez *et al.,* 2005; China: Lu *et al.,* 2010; Italy: Filimberti *et al*., 2013; Poland: Walczak-Jedrzejowska *et al.,* 2013; Belgium: Punjabi *et al.* 2016; Iran: Ahadi *et al.,* 2019; India: Kale *et al.,* 2022). Quality control is also not done in a repeatable and reliable manner in many HFEA accredited laboratories, yet this is a vital part of continued performance analysis in the laboratory (Agarwal *et al*., 2022a) which can lead to misdiagnosis, subjectivity, human error, and poor patient management (Agarwal *et al*., 2022a; Björndahl & Kirkman Brown, 2022; Long *et al*., 2018; Pacey 2010; Tomlinson, 2010). As with any laboratory, research, or study, this can also cause a barrier to scientific progress (Björndahl *et al.,* 2022) and is vital to maintain.

7.5 Future work

Much of this work established and validated research which has already been published in the literature, yet there are many opportunities these studies found and provided. Further exciting and novel work which could be done include:

- With a new set of improved sequence- and conservation score-based bovine BACs, more comparative work can be done, specifically looking into the Perissodactyla order of mammals. As previously discussed, the majority of Perissodactyl animals are on the red list with regards to being vulnerable, near threatened, endangered, or critically endangered and only a small number of comparative studies have been done with them (Steiner & Ryder, 2011). Therefore, conservation research, specifically investigating their reproductive isolation, could be crucial for these species, potentially using the equine BACs from Section 4, even though they have not been selected for conservation score.
- To improve the equine multiprobe screening device, new horse BAC clones should be selected with more stringent criteria, then isolated, labelled, and tested to eventually produce a commercially viable test for fertility screening. Additionally, by either acquiring an equine hybloc or using a cattle hybloc, repeat signals could be suppressed, improving the signal strength of the probes. These will hopefully be the next and final steps of producing this potentially industry changing device for horses.
- In addition, equine blood culturing and harvesting should be improved on and optimized. This will provide more accurate, repeatable, and consistent results for translocation screening in horses, and based on the literature, and to our knowledge, this has not been studied, making it a novel option for perfecting the metaphase spreads of equines.
- Due to the multiple variables which affect semen quality of a stallion such as breed (Sieme, Katila & Klug, 2004), day length/season (Wach-Gygax *et al*., 2017; Castro,

Morales & Parraguez, 2020) as well as ages and levels of competition of the stallions (Dowsett & Knott, 1996; Gottschalk *et al*., 2016; Wach-Gygax *et al*., 2017; Wilson *et al*., 2019; Castro, Morales & Parraguez, 2020), a mixed model including semen parameter data, stallion information and fixed/variable effects, might be a different and more suitable method of interpreting the semen analysis, DNA damage, reactive oxygen species (ROS) and lipid peroxidation (LPO) results. This type of study could be more beneficial to the industry and breeders alike as to determine the fertility potential of different stallions.

7.6 Personal perspectives

As I write up the last couple of pages of my PhD, edit the finer details and summarize my findings, it is necessary to reflect on the last three years of my life. I have learnt so much more than I ever expected, from accreditation to cell culture to sperm analysis and beyond. This journey has allowed me to develop my presentation, writing, reading and discussion skills, from an analytic and scientific standpoint. I've had the opportunity to meet some incredibly talented people who have not only motivated me, but also believed in my potential as a researcher and academic and I can truly say are some of the best friends I will have for the rest of my life. There are no words to describe how much that means and meant to me.

Unfortunately, due to COVID-19, I did not have the opportunity to travel abroad to conferences and give talks or poster presentations in person, yet I had the chance to attend a couple of locally held and online conferences, for example, European Society for Human Reproduction and Embryology (ESHRE) 2021, Fertility 2022 and the Symposium in honour of John A Woolliams. I had poster abstracts accepted for ESHRE 2021 in the category of Clinical Science – Andrology; and Fertility 2022 in the category of Sperm and Testis. Two papers were released on which I am a co-author (Lewis *et al*., 2021; Lewis *et al*., 2022) and I finally wrote and submitted a paper of my own to Human Reproduction overseen by Professor Sherly

Homa. Additionally, I had the opportunity to take part in the IVF summer course, work on two different grants (EIRA and BBSRC) as well as work for TDL Andrology as an andrologist and visit Stallion AI (a dream of mine).

Although this PhD has tested me, it has given me multiple different opportunities and overall, I am incredibly proud of what I have achieved and what I have written in this thesis. It has truly been a labour of love, with many tears involved, but so unbelievably worth it.

8. References

Abcam (2019) *AB243377 assay kit (cell-based) lipid peroxidation - Abcam*, Abcam. Available at: https://www.abcam.com/ps/products/243/ab243377/documents/Lipid-Peroxidation-Assay-Kit-(Cell-based)-v2a-ab243377%20(website).pdf (Accessed: March 21, 2023).

Agarwal, A., *et al*. (2015) 'A unique view on male infertility around the Globe', *Reproductive Biology and Endocrinology,* 13 (37), pp. 1-9. https://doi.org/10.1186%2Fs12958-015-0032-1

Agarwal, A., *et al*. (2016a) 'Clinical utility of sperm DNA fragmentation testing: practice recommendations based on clinical scenarios', *Translational andrology and urology*, *5* (6), pp. 935–950. https://doi.org/10.21037/tau.2016.10.03

Agarwal, A., *et al*. (2016b) 'MiOXSYS: a novel method of measuring oxidation reduction potential in semen and seminal plasma', *Fertility and Sterility*, 106 (3), pp. 577-573.e10. https://doi.org/10.1016/j.fertnstert.2016.05.013

Agarwal, A., Samanta, L. and Henkel, R. (eds) (2018) *Oxidants, antioxidants, and impact of the oxidative status in male reproduction*. Academic Press. https://doi.org/10.1016/C2016-0-03860-3

Agarwal, A., *et al*. (2019) 'Male Oxidative Stress Infertility (MOSI): Proposed Terminology and Clinical Practice Guidelines for Management of Idiopathic Male Infertility', *The world journal of men's health*, *37* (3), pp. 296–312. https://doi.org/10.5534/wjmh.190055

Agarwal, A., *et al*. (2022a) 'Sperm Vitality and Necrozoospermia: Diagnosis, Management, and Results of a Global Survey of Clinical Practice', *The World Journal of Men's Health,* 40, pp. 228–242. https://doi.org/10.5534%2Fwjmh.210149

Agarwal, A., *et al*. (2022b) 'Standardized Laboratory Procedures, Quality Control and Quality Assurance Are Key Requirements for Accurate Semen Analysis in the Evaluation of Infertile Male', *The world journal of men's health*, 40(1), pp. 52–65. https://doi.org/10.5534/wjmh.210022

Ahadi, M., *et al*. (2019) 'Evaluation of the standardization in semen analysis performance according to the WHO protocols among laboratories in Tehran, Iran', *Iranian Journal of Pathology,* 14, pp. 142–147. https://doi.org/10.30699%2FIJP.14.2.142

Aitken, R. J., Clarkson, J. S., and Fishel, S. (1989) 'Generation of reactive oxygen species, lipid peroxidation, and human sperm function', *Biology of reproduction*, 41 (1), pp. 183–197. https://doi.org/10.1095/biolreprod41.1.183

Aitken, R. J., and Baker, M. A. (2006) 'Oxidative stress, sperm survival and fertility control', *Molecular and cellular endocrinology*, 250 (1-2), pp. 66–69. https://doi.org/10.1016/j.mce.2005.12.026

Aitken, R.J., *et al*. (2007) 'Analysis of lipid peroxidation in human spermatozoa using BODIPYC_{11}', *Molecular Human Reproduction*, 13 (4), pp. 203-211. https://doi.org/10.1093/molehr/gal119

Aitken R. J. (2017) 'Reactive oxygen species as mediators of sperm capacitation and pathological damage', *Molecular reproduction and development*, 84 (10), pp. 1039–1052. https://doi.org/10.1002/mrd.22871

Alharbi, Y. H., *et al*. (2022) 'Reciprocal Translocation T(Y;16) in a Male Patient With Non-Obstructive Azoospermia: A Case Report and Literature Review', *Cureus*, 14 (8), pp. e28365. https://doi.org/10.7759/cureus.28365

Alkelai, A., *et al*. (2016) 'A role for TENM1 mutations in congenital general anosmia', *Clinical genetics*, 90 (3), pp. 211–219. https://doi.org/10.1111/cge.12782

Alvarez C., *et al*. (2005) 'External quality control program for semen analysis: Spanish experience', *Journal of Assisted Reproduction and Genetics*, 22, pp. 379-387. https://doi.org/10.1007%2Fs10815-005-7461-2

Amann, R. P., and Waberski, D. (2014) 'Computer-assisted sperm analysis (CASA): capabilities and potential developments', *Theriogenology*, 81 (1), pp. 5–17.e173. https://doi.org/10.1016/j.theriogenology.2013.09.004

Andrade-Rocha, F.T. (2017) 'On the origins of the semen analysis: A close relationship with the history of the Reproductive Medicine', *Journal of Human Reproductive Sciences,* 10, pp. 242-255. https://doi.org/10.4103%2Fjhrs.JHRS_97_17

Applied Genetics Laboratories (2020) *Hybloc™ Competitor DNA, a cot-1 DNA® equivalent!*, *Hybloc™ Competitor DNA (i.e., Cot-1 DNA®) : Applied Genetics Laboratories, Inc.* Available at: https://www.appliedgenetics.com/products/hybloc-competitor-dna/ (Accessed: March 18, 2023).

Arishima, T., *et al*. (2017) 'Maternal variant in the upstream of FOXP3 gene on the X chromosome is associated with recurrent infertility in Japanese Black cattle', *BMC genetics*, 18 (103), pp. 1-12. https://doi.org/10.1186/s12863-017-0573-8

Asahina, H., *et al*. (2012) 'Distribution of protocadherin 9 protein in the developing mouse nervous system', *Neuroscience*, 225, pp. 88–104. https://doi.org/10.1016/j.neuroscience.2012.09.006

Aurich, J.E. (2012) 'Artificial insemination in horses—more than a century of practice and Research', *Journal of Equine Veterinary Science*, 32 (8), pp. 458–463. Available at: https://doi.org/10.1016/j.jevs.2012.06.011.

Ayala, F.J. and Coluzzi, M. (2005) 'Chromosome speciation: Humans, drosophila, and mosquitoes', *Proceedings of the National Academy of Sciences*, 102 (suppl_1), pp. 6535–6542. Available at: https://doi.org/10.1073/pnas.0501847102.

Ball, B.A. (2008) 'Diagnostic methods for evaluation of stallion subfertility: a review', *Journal of Equine Veterinary Science,* 28 (11), pp. 650-665. https://doi.org/10.1016/j.jevs.2008.10.003

Barbăroșie, C., Agarwal, A., and Henkel, R. (2021) 'Diagnostic value of advanced semen analysis in evaluation of male infertility', *Andrologia,* 53 (2), pp. 1-18. https://doi.org/10.1111/and.13625

Barratt, C.L.R. (2007) 'Semen analysis is the cornerstone of investigation for male infertility', *Practitioner,* 251 (1690), pp. 8-17. link.gale.com/apps/doc/A158608514/AONE?u=anon~5c272b13&sid=googleScholar&xid=dcd650b5

Barrier Battut, I., *et al*. (2016) 'Development of a new fertility prediction model for stallion semen, including flow cytometry', *Theriogenology*, 86(4), pp. 1111–1131. https://doi.org/10.1016/j.theriogenology.2016.04.001

Baskaran, S., *et al*. (2021) 'Diagnostic value of routine semen analysis in clinical andrology', *Andrologia,* 53, pp. 1-12 https://doi.org/10.1111/and.13614

Baudhuin, L. M., *et al*. (2015) 'Confirming Variants in Next-Generation Sequencing Panel Testing by Sanger Sequencing', *The Journal of molecular diagnostics : JMD*, 17(4), pp. 456–461. https://doi.org/10.1016/j.jmoldx.2015.03.004

Bejjani, B.A., *et al*. (2005) 'Use of targeted array-based CGH for the clinical diagnosis of chromosomal imbalance: Is less more?', *Am. J. Med. Genet.*, 134A, pp. 259-267. https://doi.org/10.1002/ajmg.a.30621

Bhatt, R. K., and Agarwal, M. (2020) 'Study of Spectrum of Chromosomal Rearrangements in Recurrent Pregnancy Loss', *Journal of obstetrics and gynaecology of India*, 70(3), pp. 189–194. https://doi.org/10.1007/s13224-020-01305-9

Bishop, R. (2010) 'Applications of fluorescence *in situ* hybridization (FISH) in detecting genetic aberrations of medical significance', *Bioscience Horizons*, 3(1), pp. 85–95. Available at: https://doi.org/10.1093/biohorizons/hzq009.

Björndahl, L., Tomlinson, M. and Barratt, C.L. (2004) 'Raising standards in semen analysis: professional and personal responsibility', *Journal of Andrology,* 25, pp. 862-863. https://doi.org/10.1002/j.1939-4640.2004.tb03152.x

Björndahl, L. (2011) 'What is normal semen quality? On the use and abuse of reference limits for the interpretation of semen analysis results', *Human Fertility,* 14, pp. 179-186. https://doi.org/10.3109/14647273.2011.580823

Björndahl, L., *et al*. (2016) 'How to count sperm properly: checklist for acceptability of studies based on human semen analysis', *Human Reproduction*, 31, pp. 227-232. https://doi.org/10.1093/humrep/dev305

Björndahl, L. (2022) 'A paradigmatic shift in the care of male factor infertility: How can the recommendations for basic semen examination in the sixth edition of the WHO manual and the ISO 23162:2021 standard help?', *Reproductive BioMedicine Online,* 45, pp. 731–736. https://doi.org/10.1016/j.rbmo.2022.03.011

Björndahl, L., *et al*. (2022) 'Standards in semen examination: publishing reproducible and reliable data based on high-quality methodology', *Human Reproduction*, 37, pp. 2497-2502. https://doi.org/10.1093/humrep/deac189

Björndahl, L. and Kirkman Brown, J. (2022) 'The sixth edition of the WHO Laboratory Manual for the examination and processing of human semen: Ensuring Quality and standardization in BASIC examination of human ejaculates', *Fertility and Sterility*, 117, pp. 246–251. https://doi.org/10.1016/j.fertnstert.2021.12.012

Blue, M. G., Bruère, A. N., and Dewes, H. F. (1978) 'The significance of the XO syndrome in infertility of the mare', *New Zealand veterinary journal*, 26(6), pp. 137–141. https://doi.org/10.1080/00480169.1978.34521

Boe-Hansen, G. B., *et al*. (2005) 'Increasing storage time of extended boar semen reduces sperm DNA integrity', *Theriogenology*, 63 (7), pp. 2006–2019. doi: 10.1016/j.theriogenology.2004.09.006

Boe-Hansen, G. B., Fortes, M.S. and Satake, N. (2018) 'Morphological defects, sperm DNA integrity, and protamination of bovine spermatozoa', *Andrology,* 6, pp. 627–33. doi: 10.1111/andr.12486

Boe-Hansen, G. B., and Satake, N. (2019) 'An update on boar semen assessments by flow cytometry and CASA', *Theriogenology*, 137, pp. 93–103. https://doi.org/10.1016/j.theriogenology.2019.05.043

Bohlander, S.K., Kakadia, P.M. (2015) *DNA Repair and Chromosomal Translocations*. In: Ghadimi, B., Ried, T. (eds) *Chromosomal Instability in Cancer Cells. Recent Results in Cancer Research*, vol 200. Springer, Cham. https://doi.org/10.1007/978-3-319-20291-4_1

Bowling, A. T., *et al*. (1997) 'International system for cytogenetic nomenclature of the domestic horse. Report of the Third International Committee for the Standardization of the domestic horse karyotype', Davis, CA, USA, *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 5 (7), pp. 433–443. https://doi.org/10.1023/a:1018408811881

Boyd, J.C. (2010) 'Defining laboratory reference values and decision limits: populations, intervals, and interpretations', *Asian Journal of Andrology,* 12, pp. 83–90. https://doi.org/10.1038/aja.2009.9

Breen, M., *et al*. (1997) 'Detection of equine X chromosome abnormalities in equids using a horse X whole chromosome paint probe (WCPP)', *Veterinary journal (London, England : 1997)*, 153 (3), pp. 235–238. https://doi.org/10.1016/s1090-0233(97)80057-3

Breen, M., Bullerdiek, J., and Langford, C. F. (1999) 'The DAPI banded karyotype of the domestic dog (*Canis familiaris*) generated using chromosome-specific paint probes', *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 7 (5), pp. 401–406. https://doi.org/10.1023/a:1009224232134

Brinsko, S. P., et al. (2000) 'Motility, morphology and triple stain analysis of fresh, cooled and frozen-thawed stallion spermatozoa', *Journal of reproduction and fertility. Supplement*, 56, pp. 111–120.

British Equestrian Trade Association (2022) *British Equestrian Annual Report 2020-2021*. rep. British Equestrian Trade Association. Available at: https://www.britishequestrian.org.uk/assets/About%20the%20BEF/Annual%20Report/Annual%20Report/Annual%20Review%2020_21_FINAL_sm.pdf (Accessed: 2022).

Brito, L.F.C., *et al*. (2008) 'Autosomic 27 Trisomy in a Standardbred Colt', *Journal of Equine Veterinary Science*, 28 (7), pp. 431-436. https://doi.org/10.1016/j.jevs.2008.06.003

Brosnahan, M. M., Brooks, S. A., and Antczak, D. F. (2010) 'Equine clinical genomics: A clinician's primer', *Equine veterinary journal*, 42 (7), pp. 658–670. https://doi.org/10.1111/j.2042-3306.2010.00166.x

Bruining, H., *et al*. (2015) 'Genetic Mapping in Mice Reveals the Involvement of Pcdh9 in Long-Term Social and Object Recognition and Sensorimotor Development', *Biological psychiatry*, 78 (7), pp. 485–495. https://doi.org/10.1016/j.biopsych.2015.01.017

Bugno, M. *et al*. (2009) 'Identification of chromosome abnormalities in the horse using a panel of chromosome-specific painting probes generated by microdissection', *Acta Veterinaria Hungarica,* 57 (3), pp. 369-381, Available From: AKJournals https://doi.org/10.1556/avet.57.2009.3.3

Bugno, M., Jabłońska, Z., and Słota, E. (2009) 'Optimization of fluorescence *in situ* hybridization conditions in mare oocytes and mouse embryos', *Folia biologica*, 57 (1-2), pp. 49–55.

Bugno-Poniewierska, M., and Raudsepp, T. (2021) 'Horse Clinical Cytogenetics: Recurrent Themes and Novel Findings', *Animals : an open access journal from MDPI*, 11 (3), pp. 831-857. https://doi.org/10.3390/ani11030831

Burssed, B. *et al.* (2022) 'Mechanisms of structural chromosomal rearrangement formation', *Mol Cytogenet,* 15 (23), pp. 1-15. https://doi.org/10.1186/s13039-022-00600-6

Burton, S. (2022) *Turning a profit a major challenge for racehorse owners*, *BloodHorse.com*. BloodHorse. Available at: https://www.bloodhorse.com/horse-racing/articles/265285/turning-a-profit-a-major-challenge-for-racehorse-owners (Accessed: March 18, 2023).

Bustani, G.S. and Baiee, F.H. (2021) 'Semen extenders: An evaluative overview of preservative mechanisms of semen and semen extenders', *Vet World,* 14 (5), pp. 1220-1233. doi: 10.14202/vetworld.1220-1233. PMID: 34220124; PMCID: PMC8243668.

Capilla, L., *et al*. (2016) 'Mammalian Comparative Genomics Reveals Genetic and Epigenetic Features Associated with Genome Reshuffling in Rodentia', *Genome Biology and Evolution*, 8 (12), pp. 3703–3717. https://doi.org/10.1093/gbe/evw276

Carbone, L. *et al*. (2006) 'Evolutionary movement of centromeres in horse, donkey, and zebra', *Genomics*, 87 (6), pp. 777-782. https://doi.org/10.1016/j.ygeno.2005.11.012

Carrell, D.T. and De Jonge, C.J. (2016) 'The troubling state of the semen analysis', *Andrology,* 4, pp. 761-762. https://doi.org/10.1111/andr.12257

Carvalho, C.V. *et al.* (2019) 'Influence of X-chromosome markers on reproductive traits of beef cattle', *Livestock Science*, 220, pp. 152–157. Available at: https://doi.org/10.1016/j.livsci.2018.12.021.

Caspersson, T., Zech, L. and Johansson, C. (1970) 'Differential binding of alkylating fluorochromes in human chromosomes', *Experimental Cell Research,* 60 (3), pp. 315-319. https://doi.org/10.1016/0014-4827(70)90523-9.

Castro, R., Morales, P. and Parraguez, V.H. (2020) 'Post-thawing Sperm Quality in Chilean Purebred Stallions: Effect of Age and Seasonality', *Journal of Equine Veterinary Science,* 92, pp. 103170. https://doi.org/10.1016/j.jevs.2020.103170

Chakraborty, A. *et al*. (2021) 'A Novel Balanced Chromosomal Translocation in an Azoospermic Male: A Case Report', *Journal of reproduction and infertility*, 22 (2), pp. 133–137. https://doi.org/10.18502/jri.v22i2.5802

Champroux, A., *et al*. (2016) 'Mammalian sperm nuclear organization: resiliencies and vulnerabilities', *Basic and clinical andrology*, 26 (17), pp. 1-22. https://doi.org/10.1186/s12610-016-0044-5

Chemes, H.E. and Rawe, V.Y. (2003) 'Sperm pathology: a step beyond descriptive morphology. Origin, characterization and fertility potential of abnormal sperm phenotypes in infertile men', *Human Reproduction Update*, 9, pp. 405-428. https://doi.org/10.1093/humupd/dmg034

Chen, A. K., *et al*. (2008) 'Assessing the sensitivity of commercially available fluorophores to the intracellular environment', *Analytical chemistry*, 80 (19), pp. 7437–7444. https://doi.org/10.1021/ac8011347

Chiarelli, B. (1962) 'Comparative morphometric analysis of primate chromosomes I. The chromosomes of anthropoid apes and of man', *Caryologia*, 15 (1), pp. 99-121. DOI: 10.1080/00087114.1962.10796052.

Chon, K.H., Dash, S. and Ju, K. (2009) 'Estimation of respiratory rate from photoplethysmogram data using time-frequency spectral estimation', *IEEE Trans Biomed Eng.*, 56 (8), pp. 2054-2063. doi: 10.1109/TBME.2009.2019766.

Crossley, B. M. *et al*. (2020) 'Guidelines for Sanger sequencing and molecular assay monitoring', *Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc*, 32 (6), pp. 767–775. https://doi.org/10.1177/1040638720905833

Damas, J. *et al.* (2016) 'Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set', *Genome Research*, 27 (5), pp. 875–884. Available at: https://doi.org/10.1101/gr.213660.116.

Damas, J., Corbo, M. and Lewin, H.A. (2021) 'Vertebrate chromosome evolution', *Annual Review of Animal Biosciences*, 9 (1), pp. 1–27. Available at: https://doi.org/10.1146/annurev-animal-020518-114924.

Deakin, J.E., Waters, P.D. and Marshall Graves, J.A. (eds) (2010) *Marsupial genetics and genomics*. Dordrecht Heidelberg London New York: Springer. https://doi.org/10.1016/s0168-9525(02)02772-5

Deakin, J. E. *et al*. (2019) 'Chromosomics: Bridging the Gap between Genomes and Chromosomes', *Genes*, 10 (8), pp. 1-17. https://doi.org/10.3390/genes10080627

Del Giudice, F. *et al*. (2020) 'Clinical correlation among male infertility and overall male health: a systematic review of the literature', *Investigative and Clinical Urology*, 61 (4), pp. 355-371. https://doi.org/10.4111%2Ficu.2020.61.4.355

Demyda-Peyrás, S. *et al.* (2014) 'The use of a novel combination of diagnostic molecular and cytogenetic approaches in horses with sexual karyotype abnormalities: A rare case with an abnormal cellular chimerism', *Theriogenology*, 81 (8), pp. 1116–1122. Available at: https://doi.org/10.1016/j.theriogenology.2014.01.040.

Di Santo, M., *et al*. (2011) 'Human Sperm Cryopreservation: Update on Techniques, Effect on DNA Integrity, and Implications for ART'. *Adv Urol.,* 2012 (854837), pp. 1-30. doi: 10.1155/2012/854837.

Dixit, R. *et al*. (2014) 'PHYSICAL MAPPING OF GENOME AND GENES', *Journal of Biological Engineering Research and Review*, 1, pp. 6-11.

Dodman, N., *et al*. (2016) 'Genomic risk for severe canine compulsive disorder, a dog model of human OCD', , 14 (1), pp. 1-18. http://www.jarvm.com/articles/Vol14Iss1/Vol14%20Iss1Dodman.pdf

Dominguez, K., Arca, C.D.R. and Ward, W.S. (2011) *The Relationship Between Chromatin Structure and DNA Damage in Mammalian Spermatozoa*. In: Zini, A., Agarwal, A. (eds) *Sperm Chromatin*. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-6857-9_4

Dowsett, K.F. and Knott, L.M. (1996) 'The influence of age and breed on stallion semen', *Theriogenology*, 46 (3), pp. 397-412. https://doi.org/10.1016/0093-691x(96)00162-8

Dressler, G. R., and Woolf, A. S. (1999) 'Pax2 in development and renal disease', *The International journal of developmental biology*, 43 (5), pp. 463–468.

Drevet, J. R., and Aitken, R. J. (2020) 'Oxidation of Sperm Nucleus in Mammals: A Physiological Necessity to Some Extent with Adverse Impacts on Oocyte and Offspring', *Antioxidants (Basel, Switzerland)*, 9 (2), pp. 1-14. https://doi.org/10.3390/antiox9020095

Dutta, S., Majzoub, A., and Agarwal, A. (2019) 'Oxidative stress and sperm function: A systematic review on evaluation and management', *Arab journal of urology*, 17 (2), pp. 87–97. https://doi.org/10.1080/2090598X.2019.1599624

Dürrbaum, M., Storchová, Z. (2015) *Consequences of Aneuploidy in Cancer: Transcriptome and Beyond*. In: Ghadimi, B., Ried, T. (eds) *Chromosomal Instability in Cancer Cells. Recent Results in Cancer Research*, vol 200. Springer, Cham. https://doi.org/10.1007/978-3-319-20291-4_9

Easthope, E. (2020) *Flow cytometry sample prep mistakes to avoid - biocompare, Flow Cytometry Sample Prep Mistakes to Avoid*. Future Lab - Biocompare. Available at: https://www.biocompare.com/Editorial-Articles/561165-Flow-Cytometry-Sample-Prep-Mistakes-to-Avoid/ (Accessed: March 19, 2023).

Eisfeldt, J. *et al*. (2019) 'Comprehensive structural variation genome map of individuals carrying complex chromosomal rearrangements', *PLoS genetics*, 15 (2), pp. 1-29. https://doi.org/10.1371/journal.pgen.1007858

El Hachem, H., *et al*. (2017) 'Recurrent pregnancy loss: current perspectives', *International journal of women's health*, 9, pp. 331–345. https://doi.org/10.2147/IJWH.S100817

Eldridge, M.D. (2010) 'Marsupial population and conservation genetics', *Marsupial Genetics and Genomics*, pp. 461–497. https://doi.org/10.1007/978-90-481-9023-2_22.

Epifanova, E., *et al*. (2019) 'Role of Zeb2/Sip1 in neuronal development', *Brain research*, 1705, pp. 24–31. https://doi.org/10.1016/j.brainres.2018.09.034

Esteves, S.C. (2014) 'Clinical relevance of routine semen analysis and controversies surrounding the 2010 World Health Organization criteria for Semen Examination', *International Braz J Urol*, 40, pp. 433–453. https://doi.org/10.1590/s1677-5538.ibju.2014.04.02

Eurodressage (2022) *Oldenburg's top "secrets" in 2022 Autumn elite auction in Vechta*, *Eurodressage*. Available at: https://www.eurodressage.com/2022/09/28/oldenburgs-top-secrets-2022-autumn-elite-auction-vechta (Accessed: March 18, 2023).

Evenson DP, Darzynkiewicz Z and Melamed MR. (1980) 'Relation of mammalian sperm chromatin heterogeneity to fertility', *Science*, 210, pp. 1131–1133. doi: 10.1126/science.7444440.

Evenson, D. P., *et al*. (1985) 'Flow cytometric analysis of mouse spermatogenic function following exposure to ethylnitrosourea', *Cytometry,* 6, pp. 238-53. 10.1002/cyto.990060311.

Evenson, D. P. (2016) 'The Sperm Chromatin Structure Assay (SCSA®) and other sperm DNA fragmentation tests for evaluation of sperm nuclear DNA integrity as related to fertility', *Animal reproduction science*, 169, pp. 56–75. https://doi.org/10.1016/j.anireprosci.2016.01.017

Evenson, D.P., *et al*. (2020) 'Relationships between the age of 25,445 men attending infertility clinics and sperm chromatin structure assay (SCSA®) defined sperm DNA and chromatin integrity', *Fertility and Sterility*, 114 (2), pp. 311-320. https://doi.org/10.1016/j.fertnstert.2020.03.028

Falchi, L., *et al*. (2018) 'Liquid storage of ram semen for 96 h: Effects on kinematic parameters, membranes and DNA integrity, and ROS production', *Livest Sci.*, 207, pp. 1–6. doi: 10.1016/j.livsci.2017.11.001

Fan, Y. *et al*. (2002) 'Genomic structure and evolution of the ancestral chromosome fusion site in 2q13-2q14.1 and paralogous regions on other human chromosomes', *Genome research*, 12 (11), pp. 1651–1662. https://doi.org/10.1101/gr.337602

Farhadi, A. *et al*. (2013) 'Comparative FISH mapping of BMPR1B, BMP15 and GDF9 fecundity genes on cattle, river buffalo, sheep and goat chromosomes', *Journal of genetics*, 92 (3), pp. 595–597. https://doi.org/10.1007/s12041-013-0301-8

Ferguson-Smith, M. A., and Trifonov, V. (2007) 'Mammalian karyotype evolution', *Nature reviews. Genetics*, 8 (12), pp. 950–962. https://doi.org/10.1038/nrg2199

Ferreira, H. N., *et al*, (2018) 'Variable Inter-assay Estimation of Sperm DNA Fragmentation In Stallions Classified as Good and Bad Semen Freezers', *Cryo letters*, 39 (1), pp. 67–71. https://pubmed.ncbi.nlm.nih.gov/29734417/

Fesahat, F., Henkel, R. and Agarwal, A. (2020) 'Globozoospermia syndrome: an update', *Andrologia*, 52 (13459), pp. 1-13. https://doi.org/10.1111/and.13459

Filimberti, E., *et al*. (2013) 'High variability in results of semen analysis in andrology laboratories in Tuscany (Italy): the experience of an external quality control (EQC) programme', *Andrology*, 1, pp. 401-407. https://doi.org/10.1111/j.2047-2927.2012.00042.x

Fraser L. (2004) 'Structural damage to nuclear DNA in mammalian spermatozoa: its evaluation techniques and relationship with male infertility', *Polish journal of veterinary sciences*, 7 (4), pp. 311–321. https://pubmed.ncbi.nlm.nih.gov/15633792/

Frazer, K. A. *et al*. (2003) 'Cross-species sequence comparisons: a review of methods and available resources', *Genome research*, 13 (1), pp. 1–12. https://doi.org/10.1101/gr.222003

Frohlich, J. *et al*. (2017) 'Karyotype relationships among selected deer species and cattle revealed by bovine FISH probes', *PloS one*, 12 (11), pp. 1-17. https://doi.org/10.1371/journal.pone.0187559

Gallagher, D. S. *et al*. (1999) 'Applications of chromosomal fish in the Bovidae with emphases on physical mapping in domestic cattle and comparative cytogenetic analyses of the tribe Bovini', *Animal biotechnology*, 10 (3), pp. 105–108. https://doi.org/10.1080/10495399909525930

García, R.C. and Gutiérrez, S.A. (2021) *Does male infertility play a role in recurrent miscarriage?,* inviTRA. Available at: https://www.invitra.com/en/does-male-infertility-play-a-role-in-recurrent-miscarriage/ (Accessed: March 19, 2023).

Gatimel, N., *et al*. (2017) 'Sperm morphology: assessment, pathophysiology, clinical relevance, and state of the art in 2017', *Andrology*, 5, pp. 845-862. https://doi.org/10.1111/andr.12389

Gautam, R., *et al*. (2019) 'Oxidative stress-mediated alterations on sperm parameters in male Wistar rats exposed to 3G mobile phone radiation', *Andrologia*, 51 (e13201), pp. 1-9. https://doi.org/10.1111/and.13201

Gersen, S.L. and Keagle, M.B. (eds) (2013) "*Principles of Clinical Cytogenetics"*. 3rd edn. New York, New York: Scholars Portal. https://doi.org/10.1007/978-1-4419-1688-4

Ghosh, S. *et al*. (2021) 'Two Novel Cases of Autosomal Translocations in the Horse: Warmblood Family Segregating t(4;30) and a Cloned Arabian with a *de novo* t(12;25)', *Karger,* 160 (11-12), pp. 1-10. DOI: 10.1159/000512206

Ghosh, S. *et al*. (2022) 'The Second Case of Non-Mosaic Trisomy of Chromosome 26 with Homologous Fusion 26q;26q in the Horse', *Animals : an open access journal from MDPI*, 12 (7), pp. 803-814. https://doi.org/10.3390/ani12070803

Gibb, Z., Lambourne, S. R., and Aitken, R. J. (2014) 'The paradoxical relationship between stallion fertility and oxidative stress', *Biology of reproduction*, 91 (3), pp. 1-10. https://doi.org/10.1095/biolreprod.114.118539

Gibb, Z., and Aitken, R. J. (2016) 'The Impact of Sperm Metabolism during *In Vitro* Storage: The Stallion as a Model', *BioMed research international*, 2016 (9380609), pp. 1-9. https://doi.org/10.1155/2016/9380609

Gosálvez, J., *et al*. (2011) 'Relationships between the dynamics of iatrogenic DNA damage and genomic design in mammalian spermatozoa from eleven species', *Mol Reprod Dev*, 78, pp. 951– 961. https://doi.org/10.1002/mrd.21394

Gosálvez, J., Holt, W.V., and Johnston, S.D. (2014) *Sperm DNA Fragmentation and Its Role in Wildlife Conservation*. In: Holt, W., Brown, J., Comizzoli, P. (eds) *Reproductive Sciences in*

Animal Conservation. Advances in Experimental Medicine and Biology, 753. Springer, New York, NY. https://doi.org/10.1007/978-1-4939-0820-2_15

Gottschalk, M., *et al*. (2016) 'Analysis of breed effects on semen traits in light horse, warmblood, and draught horse breeds', *Theriogenology*, 85 (8), pp. 1375–1381. https://doi.org/10.1016/j.theriogenology.2015.11.030

Graham J. K. (2001) 'Assessment of sperm quality: a flow cytometric approach', *Animal reproduction science*, 68 (3-4), pp. 239–247. https://doi.org/10.1016/s0378-4320(01)00160-9

Graphodatsky, A.S., Trifonov, V.A. and Stanyon, R. (2011) 'The genome diversity and karyotype evolution of mammals', *Mol Cytogenet,* 4 (22), pp. 1-16. https://doi.org/10.1186/1755-8166-4-22

Graves J.A. (1991) 'Mammalian genome evolution: new clues from comparisons of eutharians, marsupials and monotremes', *Comp Biochem Physiol A Comp Physiol.*, 99 (1-2), pp. 5-11. https://doi.org/10.1016/0300-9629(91)90225-2

Graves J.A. (2016) 'Did sex chromosome turnover promote divergence of the major mammal groups? *De novo* sex chromosomes and drastic rearrangements may have posed reproductive barriers between monotremes, marsupials and placental mammals', *Bioessays,* 38, pp. 734–743. 10.1002/bies.201600019

Green, A.D., Kavanagh-Wright, L. and Lee, G.R. (2020 ' Investigation of the long-term yield of auditing for conformity with the ISO 15189:2012 Quality Standard in a hospital pathology laboratory', *Practical Laboratory Medicine*, 20 (e00159). https://doi.org/10.1016/j.plabm.2020.e00159

Griffin, R. A., *et al*. (2019) 'What makes a fertile sperm? Unique molecular attributes of stallion fertility', *Reproduction (Cambridge, England)*, 158 (4), pp. R125–R137. https://doi.org/10.1530/REP-19-0060

Griffin, D.K. and Bruford, M.W. (2020) *Conservation efforts and genomics: Not all genomes are created equal*. Open Access Government. Available at: https://www.openaccessgovernment.org/conservation-efforts-and-genomics-not-all-genomes-are-created-equal/87572/ (Accessed: March 25, 2023).

Gustavsson I. (1988) 'Standard karyotype of the domestic pig. Committee for the Standardized Karyotype of the Domestic Pig', *Hereditas*, 109 (2), pp. 151–157. https://doi.org/10.1111/j.1601-5223.1988.tb00351.x

Halbert, N. D. and Derr, J. N. (2007) 'A Comprehensive Evaluation of Cattle Introgression into US Federal Bison Herds', *Journal of Heredity*, 98 (1), pp. 1-12, https://doi.org/10.1093/jhered/esl051

Hamada, A., *et al*. (2012) 'Unexplained male infertility: diagnosis and management', *International braz j urol : official journal of the Brazilian Society of Urology*, 38 (5), pp. 576–594. https://doi.org/10.1590/s1677-55382012000500002

Hamilton, T.R.D.S., and Assumpção, M.E.O.D. (2020) 'Sperm DNA fragmentation: causes and identification', *Zygote (Cambridge, England)*, 28 (1), pp. 1–8. https://doi.org/10.1017/S0967199419000595

Han, H. *et al.* (2018) 'Eight Y chromosome genes show copy number variations in horses', *Archives Animal Breeding*, 61 (3), pp. 263–270. Available at: https://doi.org/10.5194/aab-61-263-2018.

Harris, I.D. (2021) *Effect of advanced paternal age on fertility and pregnancy*. Edited by L.L. Simpson and K. Eckler. Available at: https://www.medilib.ir/uptodate/show/415 (Accessed: March 18, 2023).

Hassanin, A., *et al*. (2021) 'Evolutionary history of Carnivora (Mammalia, Laurasiatheria) inferred from mitochondrial genomes', *PLoS ONE*, 16 (3), pp. 1-28. https://doi.org/10.1371/journal.pone.0240770

Hassold, T., Hall, H., and Hunt, P. (2007) 'The origin of human aneuploidy: where we have been, where we are going', *Human molecular genetics*, 16 (2), pp. R203–R208. https://doi.org/10.1093/hmg/ddm243

Henkel, R., *et al*. (2010) 'TUNEL assay and SCSA® determine different aspects of sperm DNA damage', *Andrologia*, 42, pp. 305-313. https://doi.org/10.1111/j.1439-0272.2009.01002.x

Henkel, R. and Solomon, M. (2017) 'Semen culture and the assessment of Genitourinary Tract infections', *Indian Journal of Urology,* 33, pp. 188-193. https://doi.org/10.4103/iju.iju_407_16

Henkel, R., Offor, U., and Fisher, D. (2021) 'The role of infections and leukocytes in male infertility', *Andrologia*, 53 (13743), pp. 1-19. https://doi.org/10.1111/and.13743

Hill-Baskin, A., Lander, E. and Nadeau, J. (2006) 'Chromosome Substitution Strains: A New Way to Study Genetically Complex Traits', *Methods in molecular medicine*, 128, pp. 153-72. 10.1385/1-59745-159-2:153.

Ho, S.Y.W. (2008) 'The Molecular clock and estimating species divergence', *Nature Education*, 1 (1), pp.168.

Ho, S.Y.W. (2020) *The Molecular Clock and Evolutionary Rates Across the Tree of Life*. In: Ho, S.Y.W. (eds) *The Molecular Evolutionary Clock*. Springer, Cham. https://doi.org/10.1007/978-3-030-60181-2_1

Holland, L. and Maddox, A. (2019) 'Andrology and accreditation – an opportunity for cytologists', *Cytopathology*, 30, pp. 650–652. https://doi.org/10.1111/cyt.12766

Homa, S. T., *et al*. (2019) 'A Comparison Between Two Assays for Measuring Seminal Oxidative Stress and their Relationship with Sperm DNA Fragmentation and Semen Parameters', *Genes*, 10 (3), pp. 1-14. https://doi.org/10.3390/genes10030236

Hossain, M. S., *et al*. (2011) 'Flow cytometry for the assessment of animal sperm integrity and functionality: state of the art', *Asian journal of andrology*, 13 (3), pp. 406–419. https://doi.org/10.1038/aja.2011.15

Hotaling, S., Kelley, J. L., and Frandsen, P. B. (2021) 'Toward a genome sequence for every animal: Where are we now?', *Proceedings of the National Academy of Sciences of the United States of America*, 118 (52), pp. 1-8. https://doi.org/10.1073/pnas.2109019118

Howe, B., Umrigar, A., and Tsien, F. (2014) 'Chromosome preparation from cultured cells', *Journal of visualized experiments : JoVE*, (83). https://doi.org/10.3791/50203

Hu, Q., Maurais, E. G., and Ly, P. (2020) 'Cellular and genomic approaches for exploring structural chromosomal rearrangements', *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 28 (1), pp. 19–30. https://doi.org/10.1007/s10577-020-09626-1

Huang, J., *et al.* (2014) 'Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype', *Sci Rep,* 4 (4958), pp. 1-8. https://doi.org/10.1038/srep04958

Hughes, C.M., *et al*. (1996) 'A comparison of baseline and induced DNA damage in human spermatozoa from fertile and infertile men, using a modified comet assay', *Molecular Human Reproduction*, 2 (8), pp. 613-619. https://doi.org/10.1093/molehr/2.8.613

Human Fertilisation and Embryology Authority (2021) 'Code of Practice 9th Edition', *Human Fertilisation and Embryology Authority*; 9. https://portal.hfea.gov.uk/media/ihkjnfqq/2022-07-01-code-of-practice-2021.pdf

Iannuzzi, L. (1996) 'G- and R-banded prometaphase karyotypes in cattle (*Bos taurus L.*)', *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology,* 4, pp. 448-456. 10.1007/BF02265052.

Iannuzzi, L. *et al*. (2014) 'An improved characterization of horse (*Equus caballus*, 2n=64) chromosomes by using replicating G and R banding patterns', *Caryologia: International Journal of Cytology, Cytosystematics and Cytogenetics*, 56, pp. 205-211. 10.1080/00087114.2003.10589326.

International Organization for Standardization (ISO). (2012). *EN-ISO15189 - medical laboratories - requirements for quality and competence.* International Organization for Standardization (ISO); 3. https://www.iso.org/standard/76677.html

International Organization for Standardization (ISO). (2021). *ISO 23162: 2021 Basic semen examination—Specification and test methods.* International Organization for Standardization (ISO); 1. https://www.iso.org/standard/74800.html

Ioannou *et al*. (2016) 'Impact of sperm DNA chromatin in the clinic', *J Assist Reprod Genet.* 33 (2), pp. 157-166. https://doi.org/10.1007%2Fs10815-015-0624-x

Jagannathan, V., *et al*. (2021) 'Dog10K_Boxer_Tasha_1.0: A Long-Read Assembly of the Dog Reference Genome', *Genes*, 12 (6), pp. 1-17. https://doi.org/10.3390/genes12060847

Jalali, S., *et al*. (2021) 'Progression from Monoclonal gammopathy of undetermined significance of the immunoglobulin M class (IgM-MGUS) to Waldenstrom Macroglobulinemia is associated with an alteration in lipid metabolism', *Redox biology*, 41 (101927), pp. 1-12. https://doi.org/10.1016/j.redox.2021.101927

Jennings, R. and Griffin, D. (2019) *Development and application of mammalian molecular cytogenetic tools for genome reconstruction, evolution and reproductive screening*. thesis. University of Kent. https://kar.kent.ac.uk/id/eprint/82794

Jennings, R. L., Griffin, D. K., and O'Connor, R. E. (2020) 'A new Approach for Accurate Detection of Chromosome Rearrangements That Affect Fertility in Cattle', *Animals : an open access journal from MDPI*, 10 (1), pp. 1-9. https://doi.org/10.3390/ani10010114

Jiang, Y. *et al*. (2009) 'The application of computer-assisted semen analysis parameters in semen quality analysis', *2009 International Conference on Future BioMedical Information Engineering (FBIE)*, Sanya, China, pp. 256-258, doi: 10.1109/FBIE.2009.5405901.

Jiménez, J. M. *et al*. (2022) 'Effect of the rob(1;29) translocation on the fertility of beef cattle reared under extensive conditions: A 30-year retrospective study', *Reproduction in domestic animals = Zuchthygiene*, 57 (4), pp. 349–356. https://doi.org/10.1111/rda.14073

Kaiser-Rogers, K. and Rao, K. (2005) 'Structural Chromosome Rearrangements', *The Principles of Clinical Cytogenetics – Clinical Cytogenetics,* 165 (206). *10*.1385/1-59259-833-1:165.

Kalbfleisch, T.S., *et al*. (2018) 'EquCab3, an Updated Reference Genome for the Domestic Horse', *bioRxiv*, 1 (197), pp. 1-8. https://doi.org/10.1038/s42003-018-0199-z

Kale, S., *et al*. (2022) 'Survey of semen analysis practices in India and need for standardization and improvement', *Indian J Pathology and Microbiology,* 65, pp. 117-122. https://doi.org/10.4103/ijpm.ijpm_475_21

Keel, B.A., *et al*. (2002) 'Lack of standardization in performance of the semen analysis among laboratories in the United States', *Fertility and sterility*, 78, pp. 603-608. https://doi.org/10.1016/s0015-0282(02)03296-x

Keel, B.A. (2004) 'How reliable are results from the semen analysis?', *Fertility and Sterility,* 82, pp. 41–44. https://doi.org/10.1016/j.fertnstert.2004.04.007

Kherzri, A., *et al*. (2019) 'DNA methylation patterns vary in boar sperm cells with different levels of DNA fragmentation', *BMC Genomics,* 20 (897), pp. 1-15. doi: 10.1186/s12864-019-6307-8

Kiani, J., & Rassoulzadegan, M. (2013) 'A load of small RNAs in the sperm – how many bits of hereditary information?', *Cell Res*, 23, pp. 18-19. https://doi.org/10.1038/cr.2012.181

Kiazim, L. G., *et al*. (2021) 'Comparative Mapping of the Macrochromosomes of Eight Avian Species Provides Further Insight into Their Phylogenetic Relationships and Avian Karyotype Evolution', *Cells*, 10 (2), pp. 1-16. https://doi.org/10.3390/cells10020362

Kim, G.Y. (2018) 'What should be done for men with sperm DNA fragmentation?', *Clin Exp Reprod Med*, 45 (3), pp. 101-109. https://doi.org/10.5653%2Fcerm.2018.45.3.101

Kim, J., *et al*. (2013) 'Reference-assisted chromosome assembly', *Proceedings of the National Academy of Sciences of the United States of America*, 110 (5), pp. 1785–1790. https://doi.org/10.1073/pnas.1220349110

Kochhar, P.K. and Ghosh, P. (2018) 'Intracytoplasmic sperm injection with assisted oocyte activation resulting in successful pregnancies and live birth in couples with globozoospermia: A report of two cases', *Journal of Human Reproductive Sciences*, 11, pp. 72-74. https://doi.org/10.4103/jhrs.jhrs_47_17

Koina, E., Fong, J., and Graves, J. A. M. (2006) 'Marsupial and monotreme genomes', *Genome dynamics*, 2, pp. 111–122. https://doi.org/10.1159/000095099

Krawetz, S.A., *et al.* (2011) ' A survey of small RNAs in human sperm', *Human Reproduction'*, 26 (12), pp. 3401-3412. https://doi.org/10.1093/humrep/der329

Kubiak, B. B. *et al*. (2020) 'Hybridization between subterranean tuco-tucos (Rodentia, Ctenomyidae) with contrasting phylogenetic positions', *Scientific reports*, 10 (1502), pp. 1-13. https://doi.org/10.1038/s41598-020-58433-5

Kumar, S., *et al*. (2017) 'TimeTree: A Resource for Timelines, Timetrees, and Divergence Times', *Molecular biology and evolution*, 34(7), pp. 1812–1819. https://doi.org/10.1093/molbev/msx116

Kumar, S., *et al*. (2022) 'TimeTree 5: An Expanded Resource for Species Divergence Times', *Molecular Biology and Evolution*, 39 (8), pp. 1-6. msac174, https://doi.org/10.1093/molbev/msac174

Kumaresan, A., *et al*. (2020) 'Sperm DNA integrity and male fertility in farm animals: A review', *Front. Vet. Sci.*, 7 (321), pp. 1-15. https://doi.org/10.3389/fvets.2020.00321

Kuroda, S., *et al*. (2020) 'Genetic disorders and male infertility', *Reprod Med Biol*., 19, pp. 314– 322. https://doi.org/10.1002/rmb2.12336

La Vignera, S., *et al*. (2013) 'Markers of semen inflammation: supplementary semen analysis?', *Journal of Reproductive Immunology*, 100, pp. 2-10. https://doi.org/10.1016/j.jri.2013.05.001

Lander, E.S., *et al.* (2001) 'Initial sequencing and analysis of the human genome', *Nature*, 409 (6822), pp. 860-921. http://dx.doi.org/10.1038/35057062

Lanzafame, F. M., *et al*. (2009) 'Oxidative stress and medical antioxidant treatment in male infertility', *Reproductive biomedicine online*, 19 (5), pp. 638–659. https://doi.org/10.1016/j.rbmo.2009.09.014

Larkin, D. M. *et al*. (2006) 'Comparative mapping of mink chromosome 8p: *in situ* hybridization of seven cattle BAC clones', *Animal genetics*, 37 (4), pp. 429–430. https://doi.org/10.1111/j.1365-2052.2006.01491.x

Larkin, D. M., *et al*. (2009) 'Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories', *Genome research*, 19 (5), pp. 770–777. https://doi.org/10.1101/gr.086546.108

Lear, T. L., and Layton, G. (2002) 'Use of zoo-FISH to characterise a reciprocal translocation in a thoroughbred mare: t(1;1 6)(q16;q21.3)', *Equine veterinary journal*, 34 (2), pp. 207–209. https://doi.org/10.2746/042516402776767295

Lee, E. J., *et al*. (2004) 'A 1.4-Mb interval RH map of horse chromosome 17 provides detailed comparison with human and mouse homologues', *Genomics*, 83 (2), pp. 203–215. https://doi.org/10.1016/j.ygeno.2003.07.002

Lewis, N.M. *et al.* (2021) 'Incidence, Reproductive Outcome, and Economic Impact of Reciprocal Translocations in the Domestic Pig', *DNA*, *MDPI AG*, 1 (2), pp. 68–76. https://doi.org/10.3390/dna1020007

Lewis, N. M., *et al*. (2022) 'The Economic Burden of Chromosome Translocations and the Benefits of Enhanced Screening for Cattle Breeding', *Animals : an open access journal from MDPI*, 12 (1982), pp. 1-11. https://doi.org/10.3390/ani12151982

Li, X. *et al*. (2015) 'Testicular Characteristics and The Block to Spermatogenesis in Mature Hinny', *Asian-Australasian Journal of Animal Sciences*, 29 (6), pp. 793-800. 10.5713/ajas.15.0670.

Li, X., *et al.* (2020) 'Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits', *Nat Commun,* 11 (2815), pp. 1-16. https://doi.org/10.1038/s41467-020-16485-1

Liu, Y. *et al.* (2009) '*Bos taurus* genome assembly', *BMC Genomics,* 10 (180), pp. 1-11. https://doi.org/10.1186/1471-2164-10-180

Long, S.E. (1996) 'Tandem 1;30 translocation: a new structural abnormality in the horse (*Equus caballus*)', *Karger*, 72, pp. 162-163.

Long, S., Woodward, B. and Tomlinson, M. (2018) 'Sperm toxicity testing: UK best practice guideline from the Association of Biomedical andrologists', *British Journal of Biomedical Science*, 75, pp. 53–60.

Loomis, P.R. (2001) 'The equine frozen semen industry', *Anim Reprod Sci.*, 68, pp. 191-200. https://doi.org/10.1016/s0378-4320(01)00156-7

Lu, J.C. *et al*. (2009) 'A survey on the status of Semen Analysis in 118 laboratories in China', *Asian Journal of Andrology,* 12, pp. 104–110. https://doi.org/10.1038%2Faja.2008.41

Lu, X., *et al*. (2021) 'Genome-Wide Association Study on Reproduction-Related Body-Shape Traits of Chinese Holstein Cows', *Animals : an open access journal from MDPI*, 11 (7), pp. 1-17. https://doi.org/10.3390/ani11071927

Ludwig, M., *et al*. (2003) 'Chronic prostatitis/chronic pelvic pain pyndrome): seminal markers of inflammation', *World Journal of Urology*, 21, pp. 82-85. https://doi.org/10.1007/s00345-003-0330-7

Ly, P. *et al.* (2019) 'Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements', *Nat Genet.,* 51, pp. 705–715. https://doi.org/10.1038/s41588-019-0360-8

Madlung, A. (2013) 'Polyploidy and its effect on evolutionary success: old questions revisited with new tools', *Heredity,* 110, pp. 99–104. https://doi.org/10.1038/hdy.2012.79

Mantere, T., Kersten, S., and Hoischen, A. (2019) 'Long-Read Sequencing Emerging in Medical Genetics', *Frontiers in genetics*, 10 (426), pp. 1-14. https://doi.org/10.3389/fgene.2019.00426

Martin Muñoz, P., *et al*. (2015) 'Depletion of Intracellular Thiols and Increased Production of 4-Hydroxynonenal that Occur During Cryopreservation of Stallion Spermatozoa Lead to Caspase Activation, Loss of Motility, and Cell Death', *Biology of reproduction*, 93 (6), pp. 1-11. https://doi.org/10.1095/biolreprod.115.132878

Mastrangelo, S., *et al*. (2019) 'Novel and known signals of selection for fat deposition in domestic sheep breeds from Africa and Eurasia', *PLoS ONE*, 14 (6), pp. 1-15. https://doi.org/10.1371/journal.pone.0209632

McCue, P.M. (1998) 'Review of Ovarian Abnormalities in the Mare', *American Association of Equine Practitioners (AAEP) Conference Proceedings*, 44, pp. 125-133. https://www.ivis.org/sites/default/files/library/aaep/1998/Mccue.pdf

Menkveld, R., Holleboom, C.A. and Rhemrev, J.P. (2011) 'Measurement and significance of sperm morphology', *Asian journal of andrology*, 13, pp. 59-68. https://doi.org/10.1038%2Faja.2010.67

Mesa, A.M. *et al*. (2017) 'Stallion Semen Incubated with Hydrogen Peroxide Decreased DNA Fragmentation as Measured by the TUNEL Assay', *Journal of Equine Veterinary Science,* 47, pp. 81-86. https://doi.org/10.1016/j.jevs.2016.09.007

Middelkamp, S., *et al*. (2020) 'Sperm DNA damage causes genomic instability in early embryonic development', *Science Advances*, 6 (16), pp. 1-12. https://doi.org/10.1126/sciadv.aaz7602

Mikko, S. *et al*. (1997) 'A comparative analysis of Mhc DRB3 polymorphism in the American bison (*Bison bison*)', *The Journal of heredity*, 88 (6), pp. 499–503. https://doi.org/10.1093/oxfordjournals.jhered.a023144

Miller, D. *et al*. (2019) 'Physiological, hyaluronan-selected intracytoplasmic sperm injection for infertility treatment (HABSelect): a parallel, two-group, randomised trial', *Lancet*, 393 (10170), pp. 416-422. https://doi.org/10.1016%2FS0140-6736(18)32989-1

Miller, D. (2019) 'Hyaluronan-selected sperm should not be considered an add-on', *Lancet*, 394 (10206), pp. 1320. https://doi.org/10.1016/S0140-6736(19)31248-6

Millon, L.V. and Penedo, C.T. (2009) *Chromosomal abnormalities and DNA genotyping in horses*. rep. Davis, California: University of California, pp. 1–5. Fedequinas Journal, 57:34-40.
https://www.morganhorse.com/upload/photos/3231828Chromosomal_abnormalities_modified_for_AMHA.pdf

Minhas, S., *et al*. (2021) 'European association of urology guidelines on male sexual and reproductive health: 2021 update on male infertility', *European urology*, 80, pp. 603-620. 10.1016/j.eururo.2021.08.014

Morgan, D. (2004) 'Ethics, economics and the exotic: The early career of the HFEA', *Health Care Analysis*, 12, pp. 7–26. https://doi.org/10.1023/b:hcan.0000026650.11708.82

Morrell, J. M., *et al*. (2008) 'Sperm morphology and chromatin integrity in Swedish warmblood stallions and their relationship to pregnancy rates', *Acta veterinaria Scandinavica*, 50 (2), pp. 1-7. https://doi.org/10.1186/1751-0147-50-2

Nagarajan, N. and Pop, M. (2013) 'Sequence assembly demystified', *Nat Rev Genet,* 14, pp. 157–167. https://doi.org/10.1038/nrg3367

National Institute for Health and Care Excellence. (2014) *NICE guidelines (2014) Fertility problems. Quality standard [QS73] Quality statement 4: Semen analysis*. National Institute for Health and Care Excellence; 1 : 16-17. https://www.nice.org.uk/guidance/qs73/resources/fertility-problems-pdf-2098846428613

Ni, K., *et al*. (2016) 'The impact of sperm protamine deficiency and sperm DNA damage on human male fertility: a systematic review and meta-analysis', *Andrology*, 4, pp. 789-799. https://doi.org/10.1111/andr.12216

Nie, G.J., Momont, H.W., Buoen, L. (1993) 'A Survey of Sex Chromosome Abnormalities In 204 Mares Selected for Breeding', *Journal of Equine Veterinary Science*, 13 (8), pp. 456-459, https://doi.org/10.1016/S0737-0806(07)80284-X.

Nie, W., *et al.* (2012) 'Chromosomal rearrangements and karyotype evolution in carnivores revealed by chromosome painting', *Heredity,* 108, pp. 17–27. https://doi.org/10.1038/hdy.2011.107

Nili, H.A., Mozdarani, H. and Pellestor, F. (2011) 'Impact of DNA damage on the frequency of sperm chromosomal aneuploidy in normal and subfertile men', *Iran Biomed J*., 15 (4), pp. 122-129. https://doi.org/10.6091%2FIBJ.990.2012

Noweir, H.A., Modgil, V. and Pearce, I. (2022) 'Male accessory gland inflammation (MAGI): an evolving entity', *Journal of Clinical Urology,* 15, pp. 414-423. https://doi.org/10.1177/2051415820987687

Nowicka-Bauer, K., and Nixon, B. (2020) 'Molecular Changes Induced by Oxidative Stress that Impair Human Sperm Motility', *Antioxidants (Basel, Switzerland)*, 9 (2), pp. 1-22. https://doi.org/10.3390/antiox9020134

Nurk, S., *et al*. (2022) 'The complete sequence of a human genome', *Science (New York, N.Y.)*, 376 (6588), pp. 44–53. https://doi.org/10.1126/science.abj6987

O'Connor, C. (2008) 'Karyotyping for chromosomal abnormalities', *Nature Education,* 1 (1), pp. 27. https://www.nature.com/scitable/topicpage/karyotyping-for-chromosomal-abnormalities-298/

O'Connor, R. E. *et al*. (2017) 'Isolation of subtelomeric sequences of porcine chromosomes for translocation screening reveals errors in the pig genome assembly', *Animal genetics*, 48 (4), pp. 395–403. https://doi.org/10.1111/age.12548

O'Connor, R.E., *et al.* (2018) 'Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes', *Genome Biol,* 19, pp. 1-15. https://doi.org/10.1186/s13059-018-1550-x

Onofre, J., *et al*. (2021) 'Simplified sperm testing devices: A possible tool to overcome lack of accessibility and inconsistency in male factor infertility diagnosis. an opportunity for low- and middle- income countries', *Facts, Views and Vision in ObGyn,* 13, pp. 95–98. https://doi.org/10.52054/fvvo.13.1.011

Ortega Ferrusola, C., *et al*. (2009) 'Lipid peroxidation, assessed with BODIPY-C11, increases after cryopreservation of stallion spermatozoa, is stallion-dependent and is related to apoptotic-like changes', *Reproduction (Cambridge, England)*, 138 (1), pp. 55–63. https://doi.org/10.1530/REP-08-0484

Osoegawa, K. *et al*. (1998) 'An improved approach for construction of bacterial artificial chromosome libraries', *Genomics*, 52 (1), pp. 1–8. https://doi.org/10.1006/geno.1998.5423

Pacey, A.A. (2010 'Quality Assurance and Quality Control in the laboratory andrology', *Asian Journal of Andrology,* 12, pp. 21–25. https://doi.org/10.1038%2Faja.2009.16

Pareek, C. S., Smoczynski, R., and Tretyn, A. (2011) 'Sequencing technologies and genome sequencing', *Journal of applied genetics*, 52 (4), pp. 413–435. https://doi.org/10.1007/s13353-011-0057-x

Patel, A. S., Leong, J. Y., and Ramasamy, R. (2017) 'Prediction of male infertility by the World Health Organization laboratory manual for assessment of semen analysis: A systematic review', *Arab journal of urology*, 16 (1), pp. 96–102. https://doi.org/10.1016/j.aju.2017.10.005

Pauciullo, A. *et al*. (2014) 'Sequential cross-species chromosome painting among river buffalo, cattle, sheep and goat: a useful tool for chromosome abnormalities diagnosis within the family Bovidae', *PloS one*, 9 (10), pp. 1-7. https://doi.org/10.1371/journal.pone.0110297

Peña, F. J., Ferrusola, C.O., and Muñoz, P.M. (2016) 'New flow cytometry approaches in equine andrology', *Theriogenology*, 86 (1), pp. 366–372. https://doi.org/10.1016/j.theriogenology.2016.04.050

Peña, F. J. *et al*. (2019) 'Redox Regulation and Oxidative Stress: The Particular Case of the Stallion Spermatozoa', *Antioxidants (Basel, Switzerland)*, 8 (11), pp. 1-23. https://doi.org/10.3390/antiox8110567

Perucatti, A. *et al*. (2012) 'Advanced comparative cytogenetic analysis of X chromosomes in river buffalo, cattle, sheep, and human', *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 20 (4), pp. 413–425. https://doi.org/10.1007/s10577-012-9285-0

Petersen, J.L., *et al*. (2013) 'Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data', *PLoS ONE*, 8 (1), pp. 1-15. https://doi.org/10.1371/journal.pone.0054997

Petrunkina, A. M., and Harrison, R. A. (2011) 'Cytometric solutions in veterinary andrology: Developments, advantages, and limitations', *Cytometry. Part A : the journal of the International Society for Analytical Cytology*, 79 (5), pp. 338–348. https://doi.org/10.1002/cyto.a.21044

Phillippy A. M. (2017) 'New advances in sequence assembly', *Genome research*, 27 (5), pp. xi–xiii. https://doi.org/10.1101/gr.223057.117

Piasecka, M., *et al*. (2007) 'Flow cytometry application in the assessment of sperm DNA integrity of men with asthenozoospermia', *Folia histochemica et cytobiologica*, 45 (1), pp. S127–S136. https://pubmed.ncbi.nlm.nih.gov/18292820/

Pieńkowska-Schelling, A., Kaul, A., and Schelling, C. (2020) 'X chromosome aneuploidy and micronuclei in fertile mares', *Theriogenology*, 147, pp. 34–38. https://doi.org/10.1016/j.theriogenology.2020.02.020

Pizzol, D., *et al*. (2020) 'Pollutants and sperm quality: A systematic review and meta-analysis', *Environmental Science and Pollution Research*, 28, pp. 4095–4103. https://doi.org/10.1007/s11356-020-11589-z

Powell, C.M. (2013) *Sex Chromosomes, Sex Chromosome Disorders, and Disorders of Sex Development*. In: Gersen, S., Keagle, M. (eds) *The Principles of Clinical Cytogenetics*. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-1688-4_10

Power M. M. (1987) 'Equine half sibs with an unbalanced X;15 translocation or trisomy 28', *Cytogenetics and cell genetics*, 45 (3-4), pp. 163–168. https://doi.org/10.1159/000132448

Pruitt, K., *et al*. (2020) *RefSeq Frequently Asked Questions (FAQ)* 2010 Nov 15 [Updated 2020 Jan 31]. In: RefSeq Help [Internet]. Bethesda (MD): National Centre for Biotechnology Information (US); 2011-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK50679/

Punjabi, U., *et al*. (2016) 'Fifteen Years of Belgian experience with external quality assessment of semen analysis', *Andrology*, 4, pp. 1084–1093. https://doi.org/10.1111/andr.12230

Rambags, B. P. *et al*. (2005) 'Numerical chromosomal abnormalities in equine embryos produced *in vivo* and *in vitro*', *Molecular reproduction and development*, 72 (1), pp. 77–87. https://doi.org/10.1002/mrd.20302

Rao, B., *et al*. (1989) 'Lipid peroxidation in human spermatozoa as related to midpiece abnormalities and motility', *Gamete research*, 24 (2), pp. 127–134. https://doi.org/10.1002/mrd.1120240202

Raudsepp, T., *et al*. (2019) 'Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era', *Anim Genet*, 50, pp. 569-597. https://doi.org/10.1111/age.12857

Rejón, M.R. (2019) *The origin of the human species: A chromosome fusion?*, OpenMind. Available at: https://www.bbvaopenmind.com/en/science/bioscience/the-origin-of-the-human-species-a-chromosome-fusion/ (Accessed: March 25, 2023).

Rens, W., *et al*. (1999) 'Karyotype relationships between four distantly related marsupials revealed by reciprocal chromosome painting', *Chromosome research: an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 7 (6), pp. 461–474. https://doi.org/10.1023/a:1009249813617

Rhie, A., *et al.* (2023) 'The complete sequence of the human Y chromosome', *Nature*, 621 (7978), pp. 344-354. https://doi.org/10.1038/s41586-023-06457-y

Richer, C. L. (1990) 'Standard karyotype of the domestic horse (*Equus caballus*), Committee for standardized karyotype of *Equus caballus',* The Second International Conference for Standardization of Domestic Animal Karyotypes, INRA, Jouy-en Josas, France, 22nd-26th May 1989. *Hereditas*, 112 (3), pp. 289–293. https://doi.org/10.1111/j.1601-5223.1990.tb00069.x

Riddell, D., Pacey, A. and Whittington, K. (2005) 'Lack of compliance by UK Andrology Laboratories with World Health Organization recommendations for sperm morphology

assessment', *Human Reproduction*, 20, pp. 3441–3445. https://doi.org/10.1093/humrep/dei230

Ried, T. *et al*. (1997) 'Tumour cytogenetics revisited: Comparative genomic hybridization and spectral karyotyping', *J Mol Med,* 75, pp. 801-814. 10.1007/s001090050169.

Ried, T. *et al*. (1998) 'Chromosome Painting: A Useful Art', *Human Molecular Genetics*, 7 (10), pp. 1619–1626, https://doi.org/10.1093/hmg/7.10.1619

Rizzi, R. *et al.* (2019) 'Overlap graphs and *de Bruijn* graphs: data structures for *de novo* genome assembly in the big data era', *Quant Biol,* 7, pp. 278–292. https://doi.org/10.1007/s40484-019-0181-x

Robinson, L., *et al*. (2012) 'The effect of sperm DNA fragmentation on miscarriage rates: a systematic review and meta-analysis', *Human Reproduction*, 27 (10), pp. 2908-2917. doi:10.1093/humrep/des261

Ruiz, A. J., *et al*. (2019) 'Azoospermia and Y Chromosome-Autosome Translocation in a Friesian Stallion', *Journal of equine veterinary science*, 82 (102781). https://doi.org/10.1016/j.jevs.2019.07.002

Sakamoto, M., *et al*. (1995) 'Semiautomated DNA probe mapping using digital imaging microscopy: II. System performance', *Cytometry*, 19, pp. 60-69. https://doi.org/10.1002/cyto.990190108

Samper J. C. (2001) 'Management and fertility of mares bred with frozen semen', *Animal reproduction science*, 68 (3-4), pp. 219–228. https://doi.org/10.1016/s0378-4320(01)00158-0

Sanders, D., Fensome-Rimmer, S., and Woodward, B. (2017) 'Uncertainty of measurement in andrology: UK best practice guideline from the Association of Biomedical Andrologists', *British Journal of Biomedical Science,* 74, pp. 157-162. https://doi.org/10.1080/09674845.2017.1333682

Schaeffer, A. J., *et al*. (2004) 'Comparative genomic hybridization-array analysis enhances the detection of aneuploidies and submicroscopic imbalances in spontaneous miscarriages', *American journal of human genetics*, 74 (6), pp. 1168–1174. https://doi.org/10.1086/421250

Schmid, T.E., *et al*. (2007) 'The effects of male age on sperm DNA damage in healthy non-smokers', *Human Reproduction*, 22 (1), pp. 180-187. https://doi.org/10.1093/humrep/del338

Schoen D. J. (2000) 'Comparative genomics, marker density and statistical analysis of chromosome rearrangements', *Genetics*, 154 (2), pp. 943–952. https://doi.org/10.1093/genetics/154.2.943

Scriven, P. N., *et al*. (2001) 'Robertsonian translocations--reproductive risks and indications for preimplantation genetic diagnosis', *Human reproduction (Oxford, England)*, 16 (11), pp. 2267–2273. https://doi.org/10.1093/humrep/16.11.2267

Seabright, M. (1971) 'A rapid banding technique for human chromosomes', *Lancet,* 30 (2), pp. 971-972. doi: 10.1016/s0140-6736(71)90287-x

Serranito, B., *et al.* (2021) 'Local adaptations of Mediterranean sheep and goats through an integrative approach', *Sci Rep,* 11 (21363), pp. 1-17. https://doi.org/10.1038/s41598-021-00682-z

Sharma, M. and Leslie, S.W. (2023) Azoospermia. In: *StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing*; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK578191/

Sharma, R., Agarwal, A. (2018) *Defective Spermatogenesis and Sperm DNA Damage.* In: Zini, A., Agarwal, A. (eds) *A Clinician's Guide to Sperm DNA and Chromatin Damage*. Springer, Cham. https://doi.org/10.1007/978-3-319-71815-6_14

Sieme, H., Katila, T., and Klug, E. (2004) 'Effect of semen collection practices on sperm characteristics before and after storage and on fertility of stallions', *Theriogenology*, 61 (4), pp. 769–784. https://doi.org/10.1016/s0093-691x(03)00251-6

Silva, A.F., Ramalho-Santos, J. and Amaral, S. (2021) 'The impact of antisperm antibodies on human male reproductive function: an update', *Reproduction*, 162, pp. R55-R71. https://doi.org/10.1530/rep-21-0123

Soria-Meneses, P.J., *et al*. (2022) 'Determination of ram (*Ovis aries*) sperm DNA damage due to oxidative stress: 8-OHdG immunodetection assay vs. SCSA®', *Animals*, 12 (23), 1-13. https://doi.org/10.3390/ani12233286

Sparkes, R.S. and Crandall, B.F. (1972) 'Genetic disorders affecting growth and development', *Fetal-Placental Disorders*, pp. 207–267. https://doi.org/10.1016/b978-0-12-065502-1.50011-5.

Squires E. L. (2005) 'Integration of future biotechnologies into the equine industry', *Animal reproduction science*, 89 (1-4), pp. 187–198. https://doi.org/10.1016/j.anireprosci.2005.06.022

Stallion AI (2023) *Stallion directory*, *Stallion AI Services*. Available at: https://www.stallionai.co.uk/animals/stallion (Accessed: March 18, 2023).

Stanyon, R., *et al*. (2008) 'Primate chromosome evolution: ancestral karyotypes, marker order and neocentromeres', *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 16 (1), pp. 17–39. https://doi.org/10.1007/s10577-007-1209-z

Steiner, C.C., and Ryder, O.A. (2011) 'Molecular phylogeny and evolution of the Perissodactyla', *Zoological Journal of the Linnean Society, 163*, pp. 1289-1303. https://doi.org/10.1111/j.1096-3642.2011.00752.x

Strehl, S., *et al*. (1998) 'Characterization of two novel protocadherins (PCDH8 and PCDH9) localized on human chromosome 13 and mouse chromosome 14', *Genomics*, 53 (1), pp. 81–89. https://doi.org/10.1006/geno.1998.5467

Stroupe, S., *et al.* (2023) 'Chromosome-level reference genome for North America bison (Bison bison) and variant database aids in identifying albino mutation', *G3 (Bethesda)*, 30 (13), pp. 1-9. https://doi.org/10.1093/g3journal/jkad156

Stock, K.F., *et al*. (2016) 'Genomic applications in horse breeding', *Animal Frontiers*, 6 (1), pp. 45–52. https://doi.org/10.2527/af.2016-0007

Tanga, B.M., *et al*. (2021) 'Semen evaluation: Methodological advancements in sperm quality-specific fertility assessment', *Animal Bioscience*, 34, pp. 1253-1270. https://doi.org/10.5713/ab.21.0072

The Bovine Genome Sequencing and Analysis Consortium, *et al*. (2009) 'The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution', *Science,* 324 *(*5926), pp. 522-528. https://doi.org/10.1126/science.1169588

Tomlinson, M. (2010) 'Is your andrology service up to scratch?' *Human Fertility*, 13, pp. 194–200. https://doi.org/10.3109/14647273.2010.535089

Turleau, C., and de Grouchy, J. (1973) 'New observations on the human and chimpanzee karyotypes', *Humangenetik*, 20, pp. 151–157. https://doi.org/10.1007/BF00284851

Varner, D. D., Gibb, Z., and Aitken, R. J. (2015) 'Stallion fertility: a focus on the spermatozoon', *Equine veterinary journal*, 47 (1), pp. 16–24. https://doi.org/10.1111/evj.12308

Vasconcelos, A.L., *et al*. (2022) 'Do studies published in two leading reproduction journals between 2011 and 2020 demonstrate that they followed WHO5 recommendations for basic semen analysis?', *Human Reproduction,* 37, pp. 2255–2263. https://doi.org/10.1093/humrep/deac173

Venkatesh, S., *et al*. (2009) 'Correlation of sperm morphology and oxidative stress in infertile men', *Iranian J Reprod Med*., 7 (1), pp. 29–34. https://ijrm.ir/article-1-133-en.pdf

Ventimiglia, E., *et al*. (2015) 'Infertility as a proxy of general male health: results of a cross-sectional survey', *Fertility and sterility,* 104, pp. 48-55. https://doi.org/10.1016/j.fertnstert.2015.04.020

Vilar, J.M., *et al*. (2018) 'Histological, cytogenetic and endocrine evaluation in twenty-five unilateral cryptorchid horses', *Journal of Applied Animal Research*, 46 (1), pp. 441-444, DOI: 10.1080/09712119.2017.1333435

Waberski, D., Suarez, S. S., and Henning, H. (2022) 'Assessment of sperm motility in livestock: Perspectives based on sperm swimming conditions *in vivo'*, *Animal reproduction science*, 246 (106849). https://doi.org/10.1016/j.anireprosci.2021.106849

Wach-Gygax, L., *et al*. (2017) 'Seasonal changes of DNA fragmentation and quality of raw and cold-stored stallion spermatozoa', *Theriogenology*, 99, pp. 98–104. https://doi.org/10.1016/j.theriogenology.2017.05.025

Wade, C. M., *et al*. (2009) 'Genome sequence, comparative analysis, and population genetics of the domestic horse', *Science (New York, N.Y.)*, 326 (5954), pp. 865–867. https://doi.org/10.1126/science.1178158

Wagner, H., Cheng, J. W., and Ko, E. Y. (2017) 'Role of reactive oxygen species in male infertility: An updated review of literature', *Arab journal of urology*, 16 (1), pp. 35–43. https://doi.org/10.1016/j.aju.2017.11.001

Walczak-Jedrzejowska, R., *et al*. (2013) 'Semen analysis standardization: is there any problem in Polish laboratories?' *Asian journal of andrology,* 15, pp. 616-621. https://doi.org/10.1038%2Faja.2013.48

Walters, J. L. H., *et al*. (2018) 'Oxidative Stress in the Male Germline: A Review of Novel Strategies to Reduce 4-Hydroxynonenal Production', *Antioxidants (Basel, Switzerland)*, 7 (10), pp. 1-15. https://doi.org/10.3390/antiox7100132

Wang, G. D., *et al*. (2019a) 'Dog10K: the International Consortium of Canine Genome Sequencing', *National science review*, 6 (4), pp. 611–613. https://doi.org/10.1093/nsr/nwz068

Wang, R., et al. (2019b) 'Artificial intelligence in reproductive medicine', Reproduction, 158 (4), pp. R139-154. https://doi.org/10.1530/REP-18-0523

Wang, C., *et al.* (2021) 'A novel canine reference genome resolves genomic architecture and uncovers transcript complexity', *Commun Biol,* 4 (185), pp. 1-11. https://doi.org/10.1038/s42003-021-01698-x

Warnefors, M., and Kaessmann, H. (2013) 'Evolution of the correlation between expression divergence and protein divergence in mammals', *Genome biology and evolution*, 5 (7), pp. 1324–1335. https://doi.org/10.1093/gbe/evt093

Webber, C., and Ponting, C.P. (2005) 'Hotspots of mutation and breakage in dog and human chromosomes', *Genome Research,* 15 (12), pp. 1787-1797. https://doi.org/10.1101%2Fgr.3896805

Wessex Reg. Genetics Centre (no date) *Normal male 46,XY human karyotype, Wellcome Collection*. Available at: https://wellcomecollection.org/works/zd7rdetc (Accessed: March 21, 2023).

Westerman, M., Meredith, R.W., and Springer, M.S. (2010) 'Cytogenetics Meets Phylogenetics: A Review of Karyotype Evolution in Diprotodontian Marsupials', *Journal of Heredity*, 101 (6), pp. 690–702, https://doi.org/10.1093/jhered/esq076

Wilcox, R.R. (2021). *Chapter 3 - Estimating measures of location and scale, Introduction to Robust Estimation and Hypothesis Testing (Fifth edition).* Academic Press. https://doi.org/10.1016/C2019-0-01225-3

Williams, A.C. and Ford, W.C.L. (2005) 'Relationship between reactive oxygen species production and lipid peroxidation in human sperm suspensions and their association with sperm function', *Fertility and Sterility*, 83 (4), pp. 929-936. https://doi.org/10.1016/j.fertnstert.2004.11.031

Wilson, M., *et al*. (2019) 'Variance in stallion semen quality among equestrian sporting disciplines and competitions levels', *Animals (Basel)*, 9 (8), pp. 1-16. http://dx.doi.org/10.3390/ani9080485

Winkle, T., *et al.* (2009) 'The correlation between male age, sperm quality and sperm DNA fragmentation in 320 men attending a fertility center', *J Assist Reprod Genet*, 26, pp. 41-46. DOI 10.1007/s10815-008-9277-3

Wnuk, M., *et al*. (2010) 'Redox status of equine seminal plasma reflects the pattern and magnitude of DNA damage in sperm cells', *Theriogenology*, 74 (9), pp. 1677–1684. https://doi.org/10.1016/j.theriogenology.2010.07.007

Wolff, D.J., *et al*. (2007) 'Guidance for Fluorescence *in Situ* Hybridization Testing in Hematologic Disorders', *The Journal of molecular diagnostics : JMD*, 9, pp. 134-143. 10.2353/jmoldx.2007.060128.

Wolff, D.J. (2013) *Fluorescence In Situ Hybridization (FISH)*. In: Gersen, S., Keagle, M. (eds) *The Principles of Clinical Cytogenetics*. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-1688-4_17

World Health Organization (2010) 'WHO laboratory manual for the examination and processing of human semen', *World Health Organization*; 5. ISBN 978 92 4 154778 9

World Health Organization. (2021) 'WHO laboratory manual for the examination and processing of human semen', *World Health Organization*; 6. ISBN 978-92-4-003078-7

Yahaya, M. S. *et al*. (2019) 'Application of Veterinary Cytogenetics in Domestic Animals: A Review', *Annual Research* and *Review in Biology*, 33 (1), pp. 1-16. doi: 10.9734/arrb/2019/v33i130112.

Yan, J., *et al*. (2003) 'COP9 signalosome subunit 3 is essential for maintenance of cell proliferation in the mouse embryonic epiblast', *Molecular and cellular biology*, 23 (19), pp. 6798–6808. https://doi.org/10.1128/MCB.23.19.6798-6808.2003

Yang, F. *et al*. (2004) 'Refined genome-wide comparative map of the domestic horse, donkey and human based on cross-species chromosome painting: insight into the occasional fertility of mules', *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 12 (1), pp. 65–76. https://doi.org/10.1023/b:chro.0000009298.02689.8a

Yurchuk, T., *et al*. (2021) 'Lipid peroxidation and DNA fragmentation in fresh and cryopreserved spermatozoa of men at different spermatogenesis state', *The Ukranian Biochemical Journal*, 93 (5), pp. 24-29. DOI: 10.15407/ubj93.03.024

Zhao, C. *et al*. (2015) 'Detection of foetal subchromosomal abnormalities by sequencing circulating cell-free DNA from maternal plasma', *Clinical chemistry*, 61 (4), pp. 608–616. https://doi.org/10.1373/clinchem.2014.233312

Zhou, S., *et al*. (2015) 'A clone-free, single molecule map of the domestic cow (*Bos taurus*) genome', *BMC genomics*, 16 (644), pp. 1-19. https://doi.org/10.1186/s12864-015-1823-7

Zhou, K., *et al*. (2022) 'COP1 Acts as a Ubiquitin Ligase for PCDH9 Ubiquitination and Degradation in Human Glioma', *Molecular neurobiology*, 59 (4), pp. 2378–2388. https://doi.org/10.1007/s12035-021-02634-0

Zimin, A.V., *et al.* (2009) 'A whole-genome assembly of the domestic cow, *Bos taurus', Genome Biol,* 10 (4), pp. R42.1-10. https://doi.org/10.1186/gb-2009-10-4-r42

9. Appendix

I Semen Assessment in Fertility and Pathology Laboratories and Clinics

---

Page 1: Welcome

This survey follows on from the survey we sent out last year to gauge opinion regarding semen analysis practice in laboratories around the UK.

Last year's survey provided some very interesting results, but it was apparent that in order to obtain a more complete picture, we really need your help in answering some further questions.

This survey should only be answered by a technician who is routinely performing semen analysis in your laboratory. It should only take a maximum of 10 minutes of your time.

Thank you so much to those of you who contributed last year and thank you so much for your contribution today.

Page 2: About the Laboratory

1. **Describe your laboratory (tick which applies)**

- ○ NHS Pathology
- ○ NHS Fertility
- ◉ Private Pathology
- ○ Private Fertility
- ○ Other/Combination of above

a. If other/combination of the above (please specify)

……

2. **What is your position/job in your clinic/laboratory?**

- ○ Biomedical Scientist
- ○ Clinical Embryologist
- ○ Clinical Andrologist
- ○ MLA
- ○ Other

a. If you selected Other, please specify:

……..

3. **Is your laboratory accredited?**

- ☐ UKAS
- ☐ CPA
- ☐ HFEA
- ☐ None of the above
- ☐ Other

a. If you selected Other, please specify:

…….

4. **What is the purpose of your laboratory semen analyses (tick all that apply)?**

- ☐ Diagnostic test for male infertility/reproductive function
- ☐ To select which assisted conception procedures to use in the clinic
- ☐ Sperm donor work-up
- ☐ Other

a. If you selected Other, please specify:

……..

Page 3: Performance of Semen Analysis and Reference Values

5. **Does your laboratory carry out semen analysis AND report ALL parameters strictly according to WHO 2010 guidelines?**

- ○ Yes
- ○ No

6. **Which reference values do you use on your semen analysis report?**

- ☐ Validated in-house reference ranges
- ☐ WHO 2010 reference values
- ☐ Other

a. If you selected Other, please specify:

……..

7. **Does your lab use one of the following to assess motility?**

- ○ CASA system or equivalent
- ○ Manual motility assessment
- ○ Other

a. If you selected Other, please specify:

……..

8. **How does your lab report motility (please tick all that apply)?**

- ☐ Total motility (%)
- ☐ Progressive motility (% a + b)
- ☐ Rapid progressive motility (% a)
- ☐ Sluggish progressive motility (% b)
- ☐ Non-progressive motility (% c)
- ☐ Immotile (% d)
- ☐ Other, for example, progression out of 4, velocity etc.

a. If you selected Other, please specify:

……..

9. **Does your lab perform sperm counts on (tick relevant box):**

- ○ Motile sperm
- ○ Immobilised/Fixed sperm

10. **What chamber does your lab use for counting sperm?**

- ☐ Horwell
- ☐ Makler

- [ ] Improved Neubauer
- [ ] CASA
- [ ] Other

a. If you selected Other, please specify:

……..

11. **In your laboratory, is morphology performed on:**

- ○ Fixed samples
- ○ Motile samples

12. **Are samples stained for morphology in the lab using:**

- [ ] Papanicolau stain
- [ ] Diff-Quik
- [ ] Schorr
- [ ] Rapid staining
- [ ] None
- [ ] Other

a. If you selected Other, please specify:

……..

13. **Does your laboratory report specific types of sperm defects (e.g., globozoospermia, pyriform heads etc.)?**

- ○ Yes
- ○ No

a. If you selected No, please explain why:

……

14. **Does your lab report the following parameters (tick yes or no):**

	Yes	No
Volume	○	○
pH	○	○

Appearance	○	○
Liquefaction	○	○
Viscosity	○	○
Vitality	○	○
Round cells/other cells	○	○
Peroxidase positive cells or leukocytes	○	○
Antisperm antibodies	○	○

Page 4: Quality Control in the Laboratory

15. **Does your laboratory/clinic adhere to best practice guidelines (i.e., international standards (ISO15189) and WHO 2010 criteria)?**

○ Yes

○ No

16. **Does your laboratory take part in the UK NEQAS assessment?**

○ Yes

○ No

a. **If yes, does the lab implement the exact same method of assessment for the patient samples, as they do for the UK NEQAS samples? ~ For example, you may use a Neubauer chamber for NEQAS samples, but Horwell for your patient samples**

○ Yes

○ No

i. If you selected No, please explain why:

…….

17. **Does your lab have internal quality controls in place?**

○ Yes

○ No

a. If you selected No, please explain why:

……

18. **What comment do you include on the patient report if the values are outside of the normal range?**

- ☐ This sample is only suitable for ICSI treatment
- ☐ Referral to a fertility clinic
- ☐ Referral to an urologist
- ☐ None
- ☐ Other

a. If you selected Other, please specify:

……..

Page 5: Final page

Thank you for completing this survey.

10. Incorporation of Published Work

Original research manuscripts

Lewis, N.M., … , **Bosman, L.M**., *et al*. (2021) 'Incidence, reproductive outcome, and economic impact of reciprocal translocations in the domestic pig', *DNA*, 1, pp. 68-76. https://doi.org/10.3390/dna1020007

Lewis, N.M., … , **Bosman, L.M.**, *et al.* (2022) 'The economic burden of chromosome translocations and the benefits of enhanced screening for cattle breeding', *Animals*, 12 (15), pp. 1-11. https://doi.org/10.3390/ani12151982

Original research manuscripts in preparation

Bosman, L.M., Grosu, I.A., Griffin, D.K., Ellis, P.J., and Homa, S.T. 'Implementation of World Health Organization guidelines for semen analysis: a survey of laboratories in the UK'

Poster presentations

Bosman, L.M., Ellis, P., Homa, S., and Griffin, D. (2021) 'P–111 Development of a flow cytometric assay for membrane lipid oxidation in human sperm', *Human Reproduction*, 36 (1), deab130.110, https://doi.org/10.1093/humrep/deab130.110

Bosman, L.M., Grosu, I., Griffin, D., Ellis, P., and Homa, S. Fertility, January 2022. *004: Regulation of semen assessment in the fertility and pathology laboratory and clinic, based on accreditation: as survey,* https://my.ltb.io/#/showcase/fertilityconference/2022