Kent Academic Repository

Hlahleni, Emmanuel Tafara (2022) ***ZFY:A putative transcription factor with a poorly conserved N-terminus and a highly conserved C-terminus.*** **Master of Science by Research (MScRes) thesis, University of Kent,.**

ZFY: A putative transcription factor with a poorly conserved N-terminus and a highly conserved C-terminus

A thesis submitted to the University of Kent

M.Sc. Genetics in the Faculty of Science,

Technology and Medical Studies

2020

Emmanuel Tafara Hlahleni

School of Biosciences

Supervisor: Dr Peter Ellis

I Declaration

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

Emmanuel Tafara Hlahleni

December 2020

II Acknowledgements

First and foremost, I would like to express my sincere gratitude to my excellent supervisor Dr Peter Ellis, who made this project possible. I am appreciative for his guidance, advice, patience, enthusiasm, and benevolence throughout the entirety of this project. He made sure the experience was as pleasant as possible despite the hardships of the pandemic. I would like to also thank my co-supervisor Dr Tim Fenton for also ensuring the project was possible. An additional thanks to Ane Stranger, Alex Hendry, and Lucas Kazim for their assistance throughout my project. I would also like to thank Corey Kirkland and Dr Marta Farré Belmonte for assistance with the bioinformatic analysis, and my other lab friends within the Griffin lab, whom I shared bench spaces and my other friends from the Fenton lab.

My sincere appreciation also goes to Laura Freeman, an undergraduate student whom I supervised along with Dr Peter Ellis, for trusting in my limited supervisory skills. I would also like to thank the Ortega lab as they also assisted me with troubleshooting some areas of my project.

Last but not least, I would like to thank my family for the endless support they have offered, and friends who also provided motivation throughout the difficulties of the global pandemic.

III Contents

Table of Contents

IV List of Figures

V Abbreviations

9aa TAD	9 amino acid transactivation domain
9G8	Splicing factor 9G8
aa	Amino acid(s)
AD	Acidic domain
ATR	Ataxia telangiectasia and Rad3-related
AR	Androgen receptor
AZFb	Azoospermia factor
BIC	Bayesian Information criterion
BRCA1	Breast cancer type-1 susceptibility protein
BSA	Bovine serum albumin
C2H2	Cystiene2-Histidine2
cDNA	Complementary DNA
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeat/ CRISPR-associated protein-9
CRM1	Chromosome region maintenance 1
CYPT	Cysteine-rich perinuclear theca
DBD	DNA binding domain
DSB	Double Strand Breaks
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced green fluorescent protein
EJC	Exon junction complex
E. coli	*Escherichia coli*

FANTOM5	Functional Annotation of The Mammalian Genome
γH2AX	H2AX phosphorylated on serine 139
GTEx	Genotype-Tissue Expression
H2AFX	H2A histone family member X
HA-tag	Hemagglutinin-tag
HCC	Hepatocellular carcinoma
His-tagged	Histidine tagged
HMMER	Hidden Markov model
HNSCCs	Head and neck squamous cell carcinomas
HPA	Human Protein Atlas
HPV	Human Papilloma Virus
HRP-conjugated	Horseradish peroxidase-conjugated
hZFY	Human ZFY
hRBMY	Human RBMY
IPTG	isopropyl β-d-thiogalactopyranoside
kDa	Kilodalton
LB	Lysogeny Broth
MEGAX	Molecular Evolutionary Genetics Analysis
MI	First Meiotic Metaphase Checkpoint
MSCI	Meiotic Sex Chromosome Inactivation
MSY	Male specific region of the Y chromosome
NX	Normalized eXpression

Ni-NTA	Nickel Nitriloacetic acid
NCBI	National Center for Biotechnology Information
PAGE	Polyacrylamide gel electrophoresis
PVDF	Polyvinylidene difluoride
RBMY	RNA-binding motif gene on Y chromosome
RRM	RNA recognition motif
RT-PCR	Reverse transcription polymerase chain reaction
S. cerevisiae	*Saccharomyces cerevisiae*
SCP3	Synaptonemal Complex Protein 3
SDS	Sodium dodecyl sulphate
SRGY	Serine-arginine-glycine-tyrosine
SR/RS	Serine/arginine-rich
STAR	Signal transduction and activation of RNA
TAD	Transactivating domain
TAF9	TATA box binding protein (TBP)-associated factor
TBS	Tris buffered saline
TF	Transcription factor
TFIID	Transcription factor II D
Tra2-β	Transformer-2 protein homolog β
ZFY	Zinc finger Y-chromosomal protein

VI Abstract

ZFY is a male specific Y chromosome transcriptional factor with two splice variants: a long form that is ubiquitously expressed in most mammalian species, and a short form that is testis specific. Mouse models indicate that the physiological functions of full length ZFY include promoting meiotic sex chromosome silencing at the onset of pachytene, apoptotic elimination of aberrant cells during pachytene, and spermatid development following meiotic divisions. The testis specific short isoform has no known physiological role but is unable to activate transcription in *Saccharomyces cerevisiae*. Previous work in the Ellis-Fenton laboratory showed ectopic expression of the short form in HPV-negative oropharyngeal squamous cell carcinoma (OPSCC) cell lines, which could reveal the factor leading to higher head and neck squamous cell carcinoma incidences in males than females. However, the mechanisms leading to the difference in the transactivating activity between the two isoforms are still unknown. Thus, we sought to locate the regions responsible for the transactivation activity of full length ZFY by performing protein and bioinformatic structural studies. We successfully expressed the testis specific short ZFY, and bioinformatically identified a conserved nine amino acid transactivation domain (9aa TAD) motif SVVIQDVVEDVVIE within the alternatively spliced exon. Our results suggest that short form ZFY may competitively bind to the same genomic sites as the full length ZFY, but lacks the motif predicted to recruit the core transcription complex. Thus, it may have an oncogenic effect due to inhibition of the pro-apoptotic functions of full length ZFY.

1 Introduction

1.1 Head and neck cancers

1.1.1 Head and neck squamous cell carcinoma are more prevalent in men than women

Head and neck cancers (HNCs) are a heterogenous group of diseases that affect the upper aerodigestive tract, specifically the mucosal lining (Božinović *et al.*, 2019). These diseases affect the pharynx, oral cavity, sinonasal tract or the larynx, and affect predominantly the squamous cell epithelia as usually >90% of the cancers are head and neck squamous cell carcinomas (HNSCCs) (Vigneswaran and Williams, 2014; Božinović *et al.*, 2019). Of the worldwide 18 million new cancer cases reported in 2018, roughly >880,000 were HSNSCCs (lip, oral cavity, larynx, oropharynx, nasopharynx, hypopharynx, and salivary gland combined total) indicating that there is high incidence/burden (Bray *et al.*, 2018).

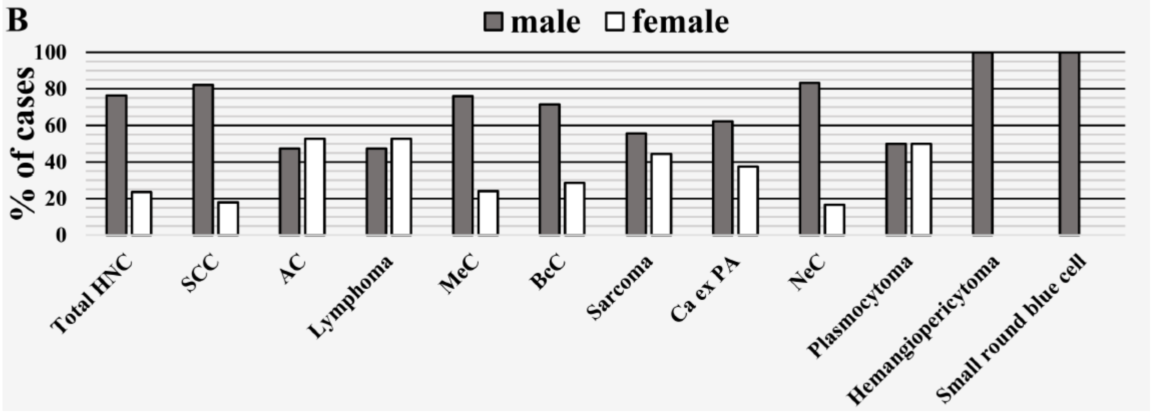


Figure 1. Head and neck cancer cases by percentages. Histopathological cancer type ratios shown next to the HNC cases. 619 subjects in total. SCC: Squamous cell carcinoma; AC: Adenocarcinoma; MeC: Mucoepidermoid carcinoma; BcC: Basocellular carcinoma; Ca ex PA: Carcinoma ex pleomorphic adenoma; NeC: Neuroendocrine carcinoma (Stoyanov *et al.*, 2017)

However, there are significant disparities between male and female incidence as males have higher HNSCCs incidences than females, as various studies have shown. One study in the *Cureus Journal of Medical Science* by (Stoyanov *et al.*,

2017) conducted in Bulgaria showed that the male to female ratio for HNC cases was 3.24:1 (**Figure 1**), and the male to female ratio for HNSCCs cases alone was roughly 4.56:1 which demonstrated that there were large disparities by sex. However, numbers vary as a 2018 study in the *CA: A Cancer Journal for Clinicians* showed a male to female HNSCCs ratio of 3:1, showing males consistently have a higher ratio (Bray *et al.*, 2018).

In terms of susceptibility of cancer incidence, the environment usually has a greater influence than genetics but genetic factors can modulate the effects of environmental factors (Lichtenstein *et al.*, 2000; Dorak and Karpuzoglu, 2012). Environmental factors that have been associated with HNSCCs incidence were excessive tobacco use and alcohol consumption (also both combined) as 70-80% of new HNSCC diagnoses were interpreted to be due to tobacco and alcohol use (Hashibe *et al.*, 2009; Praud *et al.*, 2016; Jethwa, Khariwala and Surgery, 2018). Cigarette smoke has been shown to be an HNSCC causative agent, and European and Americas region data between 1970-2000 from one study suggested that males consumed tobacco roughly five times more than females which likely increases the risk and contributes to the higher incidence in HNSCCs, but the smoking numbers likely vary as of date due to changes in culture as females smoke as much as their male counterparts (Guindon and Boisclair, 2003; Hashibe *et al.*, 2009; Domingo-vidal *et al.*, 2019).

Though the environmental factors contribute to cancer development, biological factors can also contribute to cancer development. Males generally have been shown to have less competent immune systems in comparison to females (Klein, 2012). Females usually have more competent cell mediated and humoral immune responses which are useful for efficient immune surveillance involved in clearing pathogens (usually viruses that induce tumours such as HPV), and elimination of nascent tumours by targeting tumour-specific antigens via better extrinsic tumour suppressor mechanisms (Swann and Smyth, 2007; Dorak and Karpuzoglu, 2012). Immune responses are interpreted to be modulated by hormones, namely oestrogens and androgens which both have different effects on the immune response (Dorak and Karpuzoglu, 2012). Oestrogens are thought to enhance the immune response by enhancing the innate immune response and increased cytokine/chemokine levels, whereas in males for instance, as testosterone is an important male androgen, high levels weaken the immune response by reducing antigen expression on antigen presenting cells (Dorak and Karpuzoglu, 2012). Thus, meaning that the immune surveillance in males is likely more incompetent than female immune surveillance, and HNSCCs develop more frequently.

Furthermore, females have two X chromosomes and one of the chromosome is usually inactivated. However, the inactivation is random so usually 50% of the cells contain one copy inactivate and the remaining 50% contain the other chromosome. Therefore, if there is a deleterious polymorphism or mutation in one X chromosome, 50% of the other cells contain the functional gene allowing dosage compensation for the mutated gene (Dorak and Karpuzoglu, 2012). However, males only have one X chromosome so if there were mutations within a tumour suppressor gene, the mutation leads to uncontrollable cell proliferation and likely increase the prevalence of HNSCC incidences in males (Dorak and Karpuzoglu, 2012). As males have a Y chromosome, this likely contains genes that can lead to oncogenesis. Ectopic expression of testis-specific Y-linked genes in one study by (Kido, Fai and Lau, 2015) was seen in somatic cells that were cancerous. This study suggested that Y chromosome genes potentially influence cancer development in the context of hepatocellular carcinomas (HCCs), but the same theory could possibly be applied to other cancers. As expression of male specific region of the Y-chromosome (MSY) genes is usually balanced in cells, lack of other MSY genes exacerbates oncogenesis as gene expression in the cancer cells is not properly regulated (Kido, Fai and Lau, 2015).

A recent qualitative review interpreted the enigmatic sex disparity in HNSCC for instance as being not fully explained by any environmental or genetic factors. Sex-specific biological factors were suggested to either act indirectly by modulating extrinsic oncogenic factors or directly as risk factors (Edgren, Liang and Chang, 2012). Therefore, the following thesis outlines two male-specific candidate oncogenes that could be involved in HNSCC oncogenesis when ectopically expressed. I will focus primarily on the zinc finger Y-chromosomal protein gene (ZFY), the focus of my lab work - however I will also briefly outline the biology of RNA-binding motif gene on Y chromosome (RBMY) as this is a proposed splicing regulator which was co-expressed with a testis-specific ZFY isoform in some HNSCCs cell lines in prior work from the Ellis-Fenton laboratory. It is thus possible that RBMY is responsible for the alternative splicing event generating the testis specific isoform.

1.2 ZFY

1.2.1 ZFY structure

Zinc finger Y-chromosomal protein gene (*ZFY* gene) is a gene that is principally expressed in placental mammals, is proposed to have an important role in male development, and is situated within the non-recombining region of the Y chromosome (Page *et al.*, 1987; Decarpentrie *et al.*, 2012; Jiang *et al.*, 2012). The ZFY protein is predicted to be located intracellularly in the nucleoplasm and nucleoli, and human ZFY is 801 amino acids long (90.5 kDa) with a predicted charge of -16, and an isoelectric point between 5.65-5.99. The protein is composed of a large N-terminal acidic activating domain, and a C-terminal DNA binding domain containing 13 zinc fingers encoded by a single exon, which are separated by what is proposed to be a short basic nuclear localization signal which implies it is a nuclear protein (Koopman, Ashworth and Lovell-Badge, 1991). Recently, an alternatively spliced version of the ZFY transcript was discovered by reverse transcription polymerase chain reaction (RT-PCR), which lacks the second coding exon that encodes half the acidic domain. Thus, two transcripts of ZFY are produced which are referred to as ZFY-long defined by the inclusion of the second coding exon and ZFY-short which is defined by the exclusion of the second coding exon (Decarpentrie *et al.*, 2012).

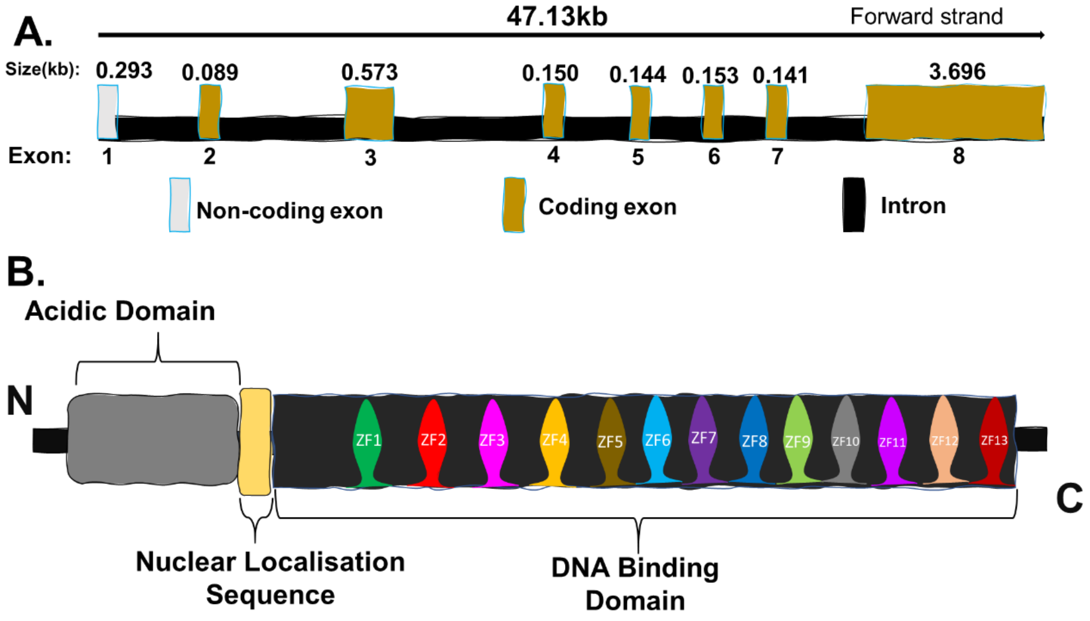


Figure 2. Human *ZFY* gene transcript and predicted protein product. (**A.**) The *ZFY* gene transcript shows 8 exons in total with 1 non-coding exon (grey) and 7 coding exons (coloured) not drawn to scale. The exons are all varying in sizes and the total gene (including introns) is 47.13kb, but the total exon length is 5.239kb. The resulting translated product is 801 amino acids long with a molecular weight of 90,504.87 Da. The diagram is an adaptation of the Archive Ensembl transcript ZFY-201 (transcript ID: ENST00000155093.7). (**B.**) An adapted depiction of the ZFY protein by (Koopman, Ashworth and Lovell-Badge, 1991) showing the N-terminal acidic domain, nuclear localisation sequence and the C-terminal DNA binding domain with 13 C2H2 zinc fingers (ZF).

ZFY proteins belong to the Krüppel-type family of C2H2-type zinc finger proteins (Koopman, Ashworth and Lovell-Badge, 1991; Poloumienko, 2004; Decarpentrie *et al.*, 2012). The acidic activating domain and the DNA binding domain (DBD) in combination suggested that the protein has hallmarks of conventional eukaryotic transcription factors (TF) (Mardon *et al.*, 1990). The zinc finger domains are usually 21 amino acids long, contain the $C\text{-}X_2\text{-}C\text{-}X_{12}\text{-}H\text{-}X_3\text{-}H$ pattern, and are usually referred to as poly-ZF as they contain $\geq$4 zinc finger repeats usually in tandem (Emerson and Thomas, 2009). Zinc fingers interact with nucleic acids, more specifically DNA, and they bind to DNA by binding to the trinucleotide using the 4 canonical positions of the zinc finger recognition helix (Persikov *et al.*, 2015).

The acidic domain was proposed to bind and recruit transcriptional regulatory machinery due to the negative charge (Decarpentrie *et al.*, 2012). As ZFY is a suspected TF, most mammalian transcription factors have been shown to interact with transcription factor II D (TFIID), a transcriptional complex, via the general transcriptional cofactor TATA box binding protein (TBP)-associated factor (TAF9) (Piskacek *et al.*, 2007). This cofactor recognises and interacts via protein-protein interactions with a transactivation domain (TAD) known as the nine amino acid transactivation domain (9aa TAD) and allows the orchestration of regulatory and transcription machinery. The 9aa TAD motif has been the best characterised binding element of various transcription factors. This type of TAD is found in the acidic domain of a group of yeast TFs known as Gal4 transcription factors and is important for Gal4 transactivation activity, as a fusion protein of the Gal4-TAD and a DBD of a repressor protein has been shown to have a similar strong transactivation activity to full length Gal4 in *Saccharomyces cerevisiae* (*S. cerevisiae*) reporter system (Piskacek *et al.*, 2007). ZFY-long acidic domain (AD) lacking the DBD has been demonstrated to have transactivation activity when fused to a Gal4-DBD in a *S. cerevisiae* reporter system (Piskacek *et al.*, 2007; Decarpentrie *et al.*, 2012). Thus, indicating the AD of ZFY-long possibly contains a 9aa TAD motif that recruits transcriptional machinery enabling transactivating properties. However, the mechanism of endogenous transactivation of full length ZFY is likely varied and does not operate in an identical fashion as shown by the ZFY-Gal4-DBD fusion protein. This transactivation domain is composed of two hydrophobic clusters and a hydrophilic region in between with nine amino acids, but the motif can be up to fourteen amino acids depending on adjacent amino acids as the function of some 9aa TAD is enhanced by adjacent amino acids (Piskacek *et al.*, 2017).

1.2.2 ZFY splicing in relation to transactivation ability

For most mammals such as humans, the general consensus is one *ZFY* gene copy on the Y chromosome. However, mice have two *Zfy* genes (*Zfy1* and *Zfy2*) compared to majority mammalian singular gene expression, and this is possibly a result of intrachromosomal duplication (Nagamine et al., 1989; Mardon et al., 1990).

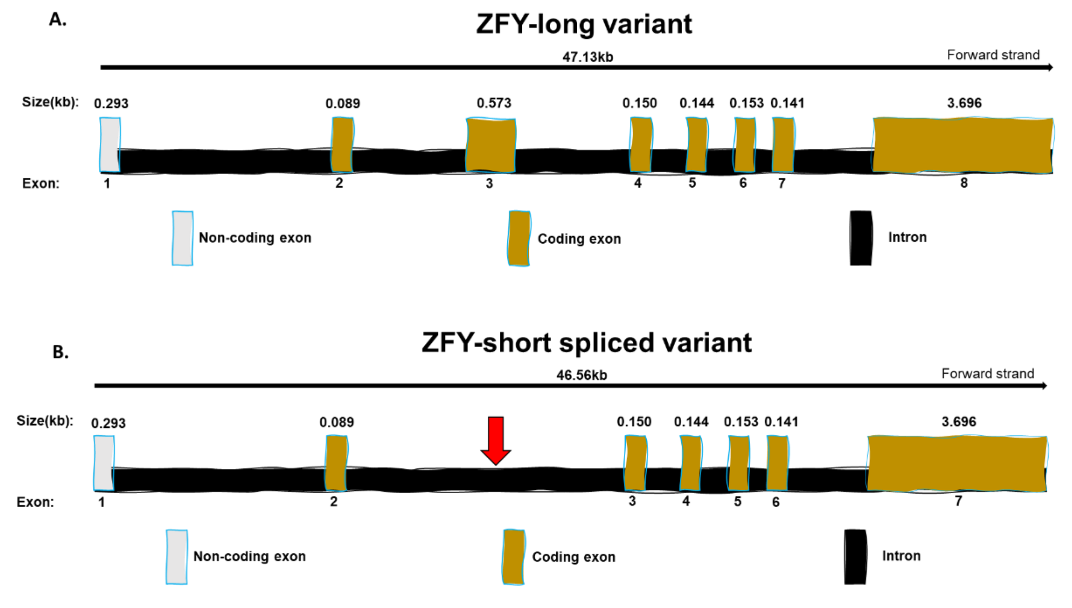


Figure 3. Human ZFY gene transcript variants. **(A.)** ZFY-long showing the full transcript with the exon-intron boundaries indicated. This transcript has one non-coding exon present, and eight coding exons as indicated, and the direction of the transcript is via the forward strand, and the total exon length is 5.239kb. **(B.)** ZFY-short lacks the second coding exon that is roughly 0.573kb which is indicated by the red arrow. The length of the transcript is shorter than the longer ZFY transcript with only 1 non-coding exon present and seven coding exons present, with the total coding exon length of 4.666kb. The diagram is an adaptation of the Archive Ensembl transcript ZFY-201 (transcript ID: ENST00000155093.7).

Mouse Zfy1 and Zfy2 are structurally homologous to human ZFY. Zfy1 is expressed as both short and long splice isoforms whereas Zfy2 is almost exclusively expressed as the long Zfy variant (Decarpentrie *et al.*, 2012; Nadège Vernet *et al.*, 2016). Therefore, mouse Zfy1 closely resembles human ZFY in its splicing pattern, while Zfy2 has lost the ability to produce the short isoform (Decarpentrie *et al.*, 2012). A second difference is that while most mammals including humans, the long isoform is ubiquitous and the short form is testis specific, in mouse Zfy1 and Zfy2 are expressed exclusively in the testis as demonstrated by northern blots and RT-PCR (Nagamine *et al.*, 1990). Mouse Zfy1 and Zfy2 have a Zfy promoter which is responsible for the roughly equal expression of the Zfy transcripts observed in spermatogonia (though Zfy1 is slightly higher) and early spermatocytes, but there was a difference in the expression of the Zfy transcripts in spermatids as mouse Zfy2 transcripts were expressed at a significantly higher level than Zfy1 (Decarpentrie *et al.*, 2012). Mouse Zfy2 contains a promoter that is spermatid-specific, derived from an X-linked gene known as CYPT, which promotes additional high-level expression of Zfy2 transcripts in spermatids (Decarpentrie *et al.*, 2012). This was shown by the fact that the Zfy promoter has weak reactivation when spermatocytes differentiate into spermatids, revealing that the Cypt-promoter leads to post meiotic spermatids with higher Zfy2 expression (Decarpentrie *et al.*, 2012).

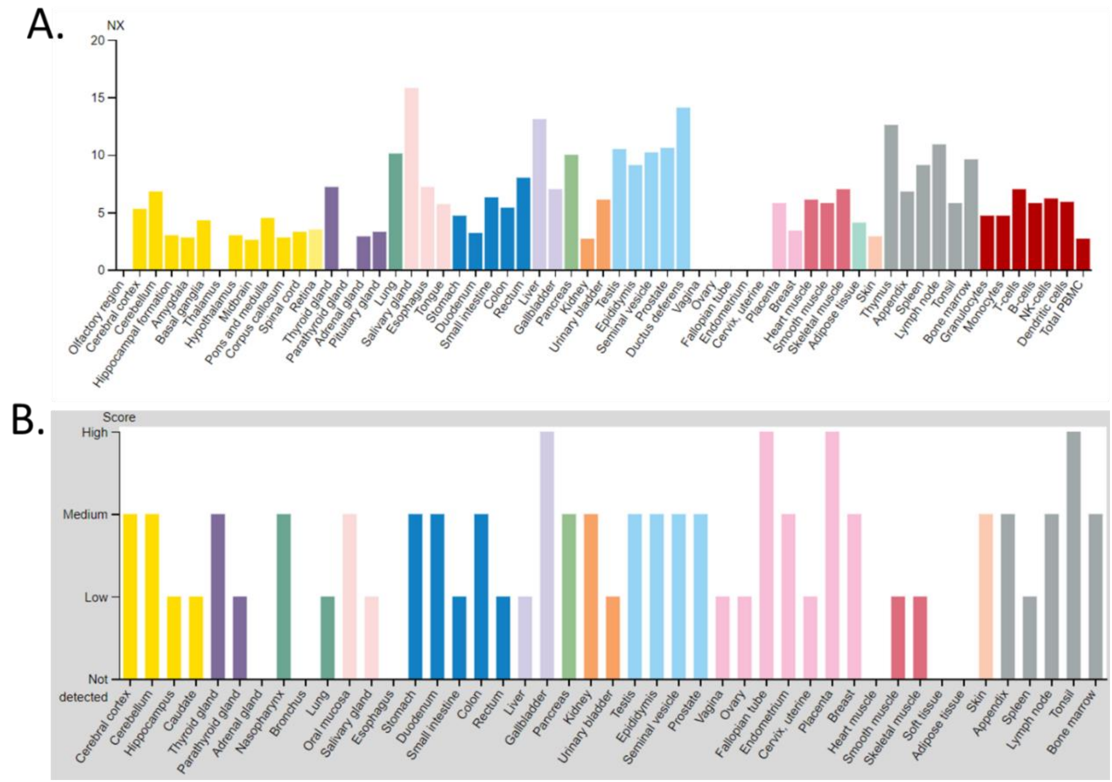


Figure 4. Human ZFY expression profiles in various organ tissues. The expression profiles show that there was ubiquitous expression of human ZFY/ZFX. However, the data is to be treated with caution as the expression was an aggregate of ZFY/ZFX as the homologous X-linked gene, ZFX, was likely mis-mapped. (**A.**) Consensus data showing the expression profiles of ZFY/ZFX by combining HPA, GTEx and FANTOM5 transcriptomics datasets. This was obtained by RNA-sequencing and it showed that the expression of ZFY RNA was ubiquitous as most of the tissues showed some degree of RNA expression, but the Normalized eXpression(NX) in each tissue was different as the salivary glands exhibited the highest RNA expression, and the cervix and other female reproductive tissue showed little RNA expression profiles likely due to the expression of the homologous gene on the X chromosome known as ZFX to no RNA expression profiles as ZFY is absent in female tissue. (**B.**) ZFY/ZFX protein expression in 44 tissues. The expression of protein was also shown to be ubiquitous in most tissue, but the degree of expression was shown to vary as some tissue showed little to no expression. The level of expression was measured as: no detection, low detection, medium detection, and high detection. The data presented was from the tissue atlas via the Human Protein Atlas (Ensembl: ENSG00000067646). There was expression seen in female-specific tissue because the antibody recognising ZFY is not specific to ZFY as it also recognises the X homolog (ZFX). Therefore, this data is to be treated with caution as the data represents aggregate expression of ZFY and ZFX.

Although human and mouse ZFY variants have been interpreted to have specific roles in spermatogenesis and spermiogenesis in the testis, the complete mechanisms of these genes are yet to be determined. Though, human ZFY-long (hZFY-long), mouse Zfy1-long and Zfy2-long have been shown to possess potent transactivating activity when the negatively charged acidic domains were fused to the Gal4-DBD in *S. cerevisiae* reporter systems, with Zfy2-long exhibiting the highest transactivation activity (Mardon *et al.*, 1990; Decarpentrie *et al.*, 2012; Vernet *et al.*, 2014; Nadège Vernet *et al.*, 2016). Human ZFY-short (hZFY-short), mouse Zfy1-short and mouse Zfy2-short however had no transactivation activity likely due to the truncated acidic domain. Thus, the domain acidity has strongly been correlated to potency of activation but however, there could possibly be other factors affecting activation efficiency (Mardon *et al.*, 1990). We hypothesised that hZFY-short, mouse Zfy1-short and Zfy2-short possibly have direct or competitive repressing abilities that antagonize hZFY-long and mouse Zfy2-long.

1.3 Physiological functions of ZFY

1.3.1 MSCI

Male meiosis entails an epigenetic silencing process known as meiotic sex chromosome inactivation (MSCI), which occurs at prophase of meiosis during early pachytene and is vital for sperm development (Turner, 2007; Burgoyne, Mahadevaiah and Turner, 2009; Royo *et al.*, 2010; Vernet *et al.*, 2011). In early spermatocytes (prior prophase I), chromosomes undergo an event at which homologous chromosomes synapse together, and the autosomes X and Y chromosomes are still transcriptionally active at this stage (Turner, 2007). As autosomal chromosomes share homology, they fully synapse together and remain transcriptionally active during in the pachytene spermatocytes during prophase I. This synapsis stage of spermatogenesis is an important aspect referred to as the synapsis checkpoint that facilitates MSCI. However, X and Y chromosomes do not share homology as the X chromosome is larger than the Y chromosome (rest of the regions remain unsynapsed), and the gene content varies significantly between each chromosome sex chromosome (Turner, 2007). The regions which remain unsynapsed then undergo MSCI and this leads to the eventual silencing of the X and Y chromosomes and a sex body is formed that transcriptionally represses the sex chromosome until the sperm develops. Thereafter, some of the genes of the sex chromosome are then reactivated.

Various experimental mouse models showed MSCI to be implicated in Y chromosome silencing by using XYY spermatocytes. As Y chromosomes are homologous and synapse together fully to form Y-Y bivalents, silencing in XYY spermatocytes was impaired and the spermatocytes did not facilitate MSCI which resulted in apoptosis mid-pachytene (stage IV of pachytene) as the Y chromosomes were still transcriptionally active. However, the X chromosome remained unaffected as it was silenced and there was no transcriptional activity like in wildtype XY mouse when they examined X-linked gene expression (Royo *et al.*, 2010; Nadège Vernet *et al.*, 2016).

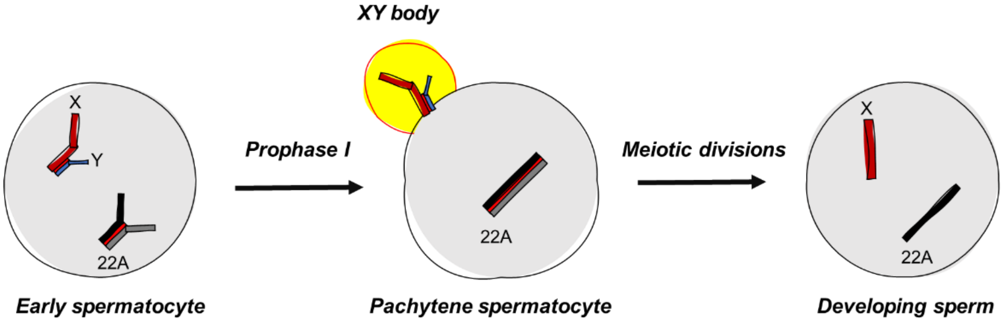


Figure 5. Human meiotic sex chromosome inactivation model. As demonstrated in the early spermatocytes, there is synapsis of homologous chromosomes and both the autosomal chromosomes (22A), and sex chromosomes(X and Y) are producing transcripts as they are active during leptotene and zygotene. However, as prophase I continues to pachytene spermatocyte, the autosomal chromosomes are fully synapsed whereas the XY chromosomes are only synapsed at the pseudoautosomal regions (PAR) with the rest of the chromosome portion being unsynapsed, resulting in the formation of a peripheral nuclear subdomain known as the 'XY body', and eventual silencing of X and Y chromosomes. As the spermatocytes develop and undergo further differentiation by desynapsis and cell division, the X chromosome (in X-bearing cells) is still transcriptionally repressed but then some of the genes eventually are reactivated the further the sperm develops. The image is an adaptation of Current Biology and (Turner, 2007). 22A represents 22 autosomal chromosomes and the red colour in between the homologous chromosomes represent synapse formation.

MSCI is regulated by a chromatin structure comprised of specialized proteins and modifications. For the meiotic silencing to occur in the male gametes, the histone protein H2A histone family member X (H2AFX) is phosphorylated at serine-139 into H2AX phosphorylated on serine 139 (γH2AX) as a response to DNA double strand breaks (DSBs) (Rogakou *et al.*, 1999). Subsequently, other histone proteins are modified by methylation and ubiquitination (Mckee and Handel, 1993; Burgoyne, Mahadevaiah and Turner, 2009; Nadège Vernet *et al.*, 2016; Jan *et al.*, 2018). Thereafter, H2AFX is phosphorylated for the second time by the phosphorylase Ataxia telangiectasia and Rad3-related (ATR), which is recruited to unsynapsed X and Y chromosome axis by Breast cancer type-1 susceptibility protein (BRCA1) to form a stable chromatin structure known as the XY body (Mahadevaiah *et al.*, 2001; Bellani *et al.*, 2005; Murr *et al.*, 2007; Turner, 2007).

1.3.2 Physiological functions of *ZFY*

ZFY function outside of the testis remains obscure making it difficult to clarify what other roles *ZFY* genes are involved in. The majority of ZFY studies conducted in the testis have predominantly involved mutant mice rather than human testis, as the mouse model is relatively comparable to a human model. However, demonstrations from mouse models are to be used cautiously when paralleled with unknown mechanisms in humans, but the principles should remain consistent. However, since mouse Zfy1 and Zfy2 are testis-specific, mouse studies cannot inform us about effects of ZFY outside the testis in other species with ubiquitous expression of ZFY-long. Within the testis context, multiple studies point to ZFY preventing aneuploidies or aberrations ensuring viable germ cells survive.

Zfy genes have been implicated in promoting MSCI because when male mouse germ cells were deficient in *Zfy1* and *Zfy2* (X^EO,Sry, XO males transgenic for Sry [conferring maleness] and Eif2sy [spermatogonia proliferation]), there was MSCI leakage during pachytene (Nadège Vernet *et al.*, 2016). When there was transgenic restoration of *Zfy1* or *Zfy2* in germ cells with univalent chromosomes, the germ cells with leaky MSCI were reduced as the *Zfy* genes corrected the MSCI leakage. This implied that premeiotic expression of *Zfy* genes in juvenile germ cells was essential for the germ cells to progress into MSCI before pachytene (Nadège Vernet *et al.*, 2016). Though essential for germ cells to express *Zfy* pre-meiotically, it is essential that Y chromosome silencing occurs before pachytene for the progression of pachytene (Royo *et al.*, 2010).

However, when silencing did not occur in mouse germ cells, *Zfy1* and *Zfy2* were inappropriately expressed as a result of MSCI leakage. This implied that *Zfy* genes are executioner genes with possibly a negative feedback loop as they likely upregulate themselves leading up to pachytene, downregulate themselves during pachytene and then reactivated in spermatids predominantly as *Zfy2* due to the Cypt-promoter (Royo *et al.*, 2010; Nadège Vernet *et al.*, 2016). A consequence of MSCI leakage when there was impaired Y-silencing in XYY males for instance was mid-pachytene (stage IV) arrest that eventually led to germ cell apoptosis (Royo *et al.*, 2010; Decarpentrie *et al.*, 2012; Nadège Vernet *et al.*, 2016).

Furthermore, *Zfy* mediates apoptotic elimination of germ cells at the first meiotic metaphase spindle assembly checkpoint (MI). When mouse germ cells had a univalent chromosome with the Y-chromosome derived sex reversal factor with the testis-determining factor (XS*xr*^{a}O), there was observed elimination of spermatocytes at the MI by apoptosis (Kot and Handel, 1990; Vernet *et al.*, 2011). However, germ cells with univalent X chromosome with Y-short arm gene deletion (XS*xr*^{b}O) and X^{E}O,Sry males were not eliminated as they completed MI and became interphasic secondary spermatocytes. The secondary spermatocytes however were arrested between the first and second meiotic division and eventually eliminated by the delayed apoptosis (Vernet *et al.*, 2011, 2014; Nadège Vernet *et al.*, 2016). In spite of this, with transgenic restoration of *Zfy2* into X^{E}O,Sry males, the secondary spermatocytes rapidly underwent apoptosis at the first meiotic metaphase due to *Zfy2* reinstating the efficient apoptotic response (Vernet *et al.*, 2011; Nadège Vernet *et al.*, 2016).

Although *Zfy2* enables the efficient apoptosis of germ cells, it has also been shown to promote the second meiotic division. Male mice germ cells deficient in *Zfy* genes with univalent X chromosomes completed first meiotic phase and were arrested subsequently preventing the second meiotic division (Vernet *et al.*, 2011, 2014). A small number of the germ cells however completed meiosis I and consequently, a large proportion of those germ cells became diploid secondary spermatocytes that entered interphase and became diploid round spermatids

(Mahadevaiah *et al.*, 2012; Vernet *et al.*, 2014). Though, reinstating Zfy2 into the males led to the majority of spermatocytes to complete meiosis I and meiosis II as the frequency of haploid round spermatids increased significantly (Vernet *et al.*, 2014).

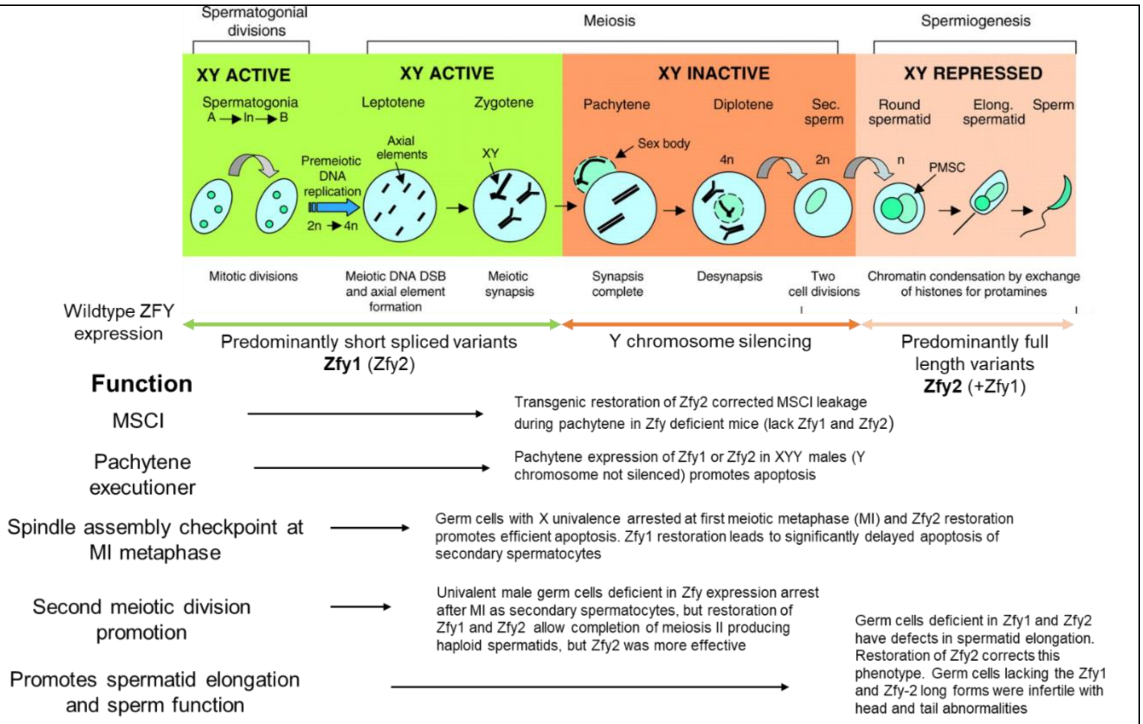


Figure 6. Overview of ZFY function. Mouse Zfy transgenes were used experimentally to analyse the physiological functions at various stages of sperm development from spermatogenesis to spermiogenesis. (Turner, 2007).

Zfy has been shown to promote elongation of spermatids and aid in sperm function (Yamauchi *et al.*, 2015; Nadège. Vernet *et al.*, 2016). Germ cells with univalent X chromosomes had abnormal and delayed spermatid development as they did not elongate, and the chromatin was not remodelled which led to round spermatid arrest and elimination by apoptosis (Mahadevaiah *et al.*, 2012; Nadège. Vernet *et al.*, 2016). However, reinstatement of *Zfy2* transgenes promoted spermiogenesis progression as it enabled round spermatids that were randomly orientated to transition into spermatids undertaking sperm morphogenesis (spermatid elongation, nuclei condensation and tail formation) (Nadège. Vernet *et al.*, 2016). In addition, CRISPR/Cas9 double knockout of *Zfy1* and *Zfy2* showed that the sperm produced had head and tail defects, abnormal mitochondria, and were infertile with chromosome aberrations that led to failure in early embryonic development (Yamauchi *et al.*, 2015; Nakasuji *et al.*, 2017). Combined, research concerning *Zfy* shows that these genes are crucial factors in spermatogenesis and spermiogenesis for the development of sperm, more significantly *Zfy2*.

Though the function of *Zfy*-long primarily concerns sperm development, the function for *Zfy*-short is yet to be discovered. However, both splice variants must share the same genomic binding sites since they share a common DNA binding domain on the terminal zinc finger exon. Since hZFY-short has been shown to lack transactivation properties, it is likely that it will competitively inhibit the function of hZFY-long, and may potentially even serve to directly repress the same genes activated by hZFY-long (Decarpentrie *et al.*, 2012). As previous papers heavily suggest that *Zfy*-long genes are involved in promoting apoptosis and cell remodelling, the testis specific *hZFY*-short is predicted to have anti-apoptotic properties. Dysregulation in proliferation of cells is a universal trait of cancer, thus it is possible that cancer cells ectopically expressing *hZFY*-short survive and become malignant, relative to the head and neck region.

1.3.3 *ZFY* as an oncogene

ZFY has been proposed to possess indirect oncogenic activity as (Tricoli and Bracken, 1993) insinuated that excessive activation of ZFY transcription in human prostate cancer cells possibly leads to deregulation of growth regulatory genes and consequently prostate malignancy. This was due to the fact that *ZFY* was found to be more frequently expressed in malignant prostate tissue, and absent in benign hyperplastic tissue, which postulated that the gene became transcriptionally active during prostate malignancy (Tricoli and Bracken, 1993). This research was conducted before the discovery of a second hZFY-short transcript variant so there needs to be further research establishing which variant was present in the adenocarcinomas. Interestingly, a novel 4.3kb transcript was observed by northern analysis which could represent the hZFY-short variant, as the coding exons of the transcript shown by **Figure 3B** has the same number of bases.

ZFY is interpreted as a favourable prognostic marker in head and neck squamous cell carcinomas by *The Human Protein Atlas*. However, this data should be interpreted very cautiously as the data does not properly distinguish ZFY from its X chromosome homolog ZFX. Additionally, previous data from a former student showed both *ZFY* spliced forms with h*ZFY*-long observed in all the male cell lines as expected and h*ZFY*-short in one of the HPV-negative oropharyngeal squamous carcinoma cell lines, which could indicate that the short form of h*ZFY* was oncogenic as it is usually expressed only in the testis.

1.3.4 Regulation of *ZFY* splicing

It is not properly known how *ZFY* alternative splicing occurs which means the consequences of mis-splicing these genes are also not fully known. Therefore, questions that can be raised are: how is splicing of *ZFY* initiated or triggered and if *ZFY* is implicated in cancer does mis-splicing lead to cancer? Another Y-linked gene known as *RBMY* is expressed in testis upstream of *ZFY* and is a known regulator of alternative splicing. *RBMY* was co-expressed with h*ZFY*-short in a OPSCC cell line, thus suggesting that these genes act to collectively promote malignancy in males.

While RBMY is not the focus of this project, I review its known structure and functions below for completeness.

1.4 RBMY

1.4.1 RBMY structure

RNA-binding motif gene on Y chromosome (*RBMY*) are 30 gene and pseudogenes copies located on Y-chromosome arms, with the functional copies clustered within the azoospermia factor interval known as the AZFb locus. *RBMY* gene copies encode a male germ cell-specific RNA binding protein (RBMY) composed of 496 amino acids (Tsuei *et al.*, 2004). RBMY is expressed specifically in the testis of all mammals and located in the nucleus of adult male germ cells between spermatogonia and round spermatid phase, with a single RNA recognition motif (RRM) at the N-terminus, a linker region containing RS (arginine/serine) dipeptides, and internally repeating tetrapeptide motif known as the serine-arginine-glycine-tyrosine (SRGY) at the C-terminal domain (Tsuei *et al.*, 2004, 2011; Dreumont *et al.*, 2009; Chua *et al.*, 2015). The protein was demonstrated to be nuclear localised transfected cells showed the protein was concentrated in the nucleus, emphasizing that the protein must have a predominant role in the nucleus (Dreumont *et al.*, 2009; Liu *et al.*, 2009). The RRM is composed of 3 exons and the C-terminal auxiliary domain is made up of 4 tandem exons, with each individual exon coding for repeats of the 37-amino acid SRGY box (Tsuei *et al.*, 2004).

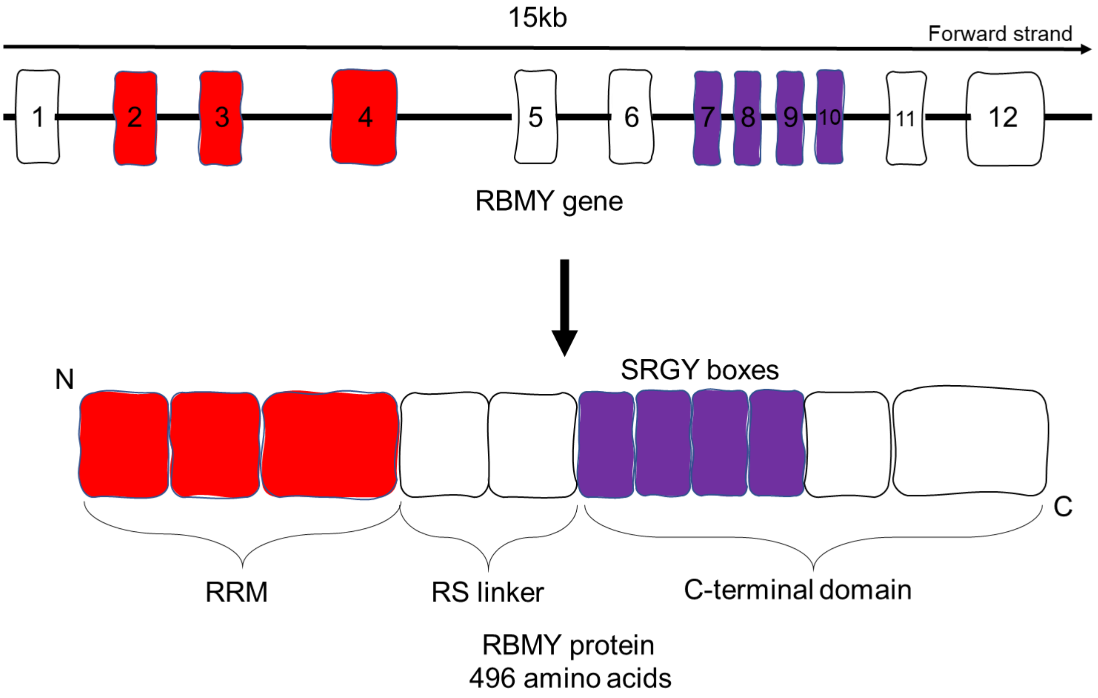


Figure 7. Human RBMY gene and* RBMY *protein structure. The human *RBMY* gene transcript consists of 12 exons and is 15kb in total length. The gene is then transcribed into a transcript of 1878 bases and then spliced into a protein with consisting of an N-terminal RNA recognition motif, RS linker region and a C-terminal domain that contains 4 tandem SRGY boxes. This diagram is an adaptation of (Tsuei et al., 2004).

Human RBMY (hRBMY) has been shown to interact with high affinity with RNA stem-loop structures referred to as pentaloops which contain a conserved CA/UCAA consensus sequence (Skrisovska *et al.*, 2007; Dreumont *et al.*, 2009). This demonstrated the sequence-specific interaction of hRBMY as it recognized specifically the trinucleotide CAA of RNA which was stabilized by the protein main chain and side chain interactions. Additionally, the RRM was shown to stabilize the 5' end of the pentaloops to maximise the affinity for the binding target. However, the RRM was also proposed to bind in a shape-specific manner as the hRBMY has β2-β3 hairpin loop that is formed from 7 amino acids that inserts into the major groove of the RNA helix and result in stable intermolecular interactions (Skrisovska *et al.*, 2007).

1.4.2 RBMY Physiological Function

As the AZFb locus has been implicated in spermatogenesis of germ cells, *RBMY* has been suggested to be a component involved in spermatogenesis (Elliott *et al.*, 1997). The deletion of AZFb locus (where the functional *RBMY* gene was shown to be located) was shown to lead to arrest of germ cells at the meiotic stage of spermatogenesis as germ cells with absent *RBMY* develop up to meiosis, but failed to complete meiosis signifying its necessity for adult mammalian male fertility (Elliott *et al.*, 1997; Mahadevaiah *et al.*, 1998; Tsuei *et al.*, 2011). This arrest is phenotypically similar to XYY males, and also to transgenic mice overexpressing *Zfy*-long during pachytene (Royo *et al.*, 2010). This potentially implies that RBMY acts to promote alternative formation of the *ZFY/Zfy*-short antagonistic isoform. In this case, lack of RBMY would result in excess lead to upregulation of *ZFY*-long during pachytene, which in turn leads to pachytene arrest and apoptosis.

In spermatogenesis, gene expression changes within spermatozoa during spermatogenesis occur due to tightly regulated processes, and a good regulation process resulting in protein diversity from multiple mRNAs is alternative splicing. RBMY was proposed to be a RNA splicing regulator that modulates activity of splicing factors that are constitutively expressed (Liu *et al.*, 2009; Tsuei *et al.*, 2011). This was because the interacting partners of RBMY included SR-related proteins (SRp20, 9G8 and Tra2β), STAR proteins (Sam68 and T-STAR proteins), and EJC components (Magoh) which are all RNA binding proteins involved in splicing regulation, signal transduction and activation of RNA (Venables *et al.*, 1999; Elliott *et al.*, 2000; Venables, 2000; Tsuei *et al.*, 2004).

These proteins interact with RBMY by forming protein interactions via the SRGY box and also the linker region as these regions contain RS dipeptides, but the

RRM does not interact due to the lack thereof of RS dipeptides (Venables, 2000; Dreumont *et al.*, 2009). Furthermore, hRBMY interactions with 9G8 and Tra2-β resulted in repression of 9G8 and Tra2-β splicing activity of pre-mRNA substrates, and the ectopic expression of hRBMY led to activation of splicing of a gene that 9G8 and Tra2-β were involved in repressing its splicing (Dreumont *et al.*, 2009). This insinuated that hRBMY functioned in an antagonistic manner in relation to 9G8 and Tra2-β. This therefore suggests that hRBMY possibly has a function in germ cells as a co-regulator of certain alternative splicing events (Dreumont *et al.*, 2009).

1.4.3 Proposed RBMY Oncogenic Effect

As RBMY is a splicing factor within the testis, we hypothesized that it indirectly alters proteins that are involved in splicing of substrates, and one we are suspicious of is ZFY as h*ZFY*-short was inappropriately expressed in OPSCC cell lines as demonstrated by a former Ellis-Fenton laboratory student. In addition, a great number of papers have demonstrated that aberrant expression of RBMY encourages the oncogenic capabilities in hepatocellular carcinomas (HCC), as one third of the male HCC tissue expressed RBMY explaining the prevalence in male HCC (Tsuei *et al.*, 2004). The mechanisms that have been interpreted are that the knockdown of RBMY correlated with increased expression of an androgen receptor (AR) inhibitor which reduced AR transactivation activity in an HCC cell line and allowed regulated AR gene expression. Therefore, expression of RBMY was proposed to be antagonistic as it directly reduced the AR inhibitory protein and this indirectly increased the risk of human male hepatocarcinogenesis by upregulation of AR target genes that have anti-apoptotic abilities (Tsuei *et al.*, 2011).

Although RBMY is a nuclear protein in testes, RBMY has been shown however to be expressed in the cytoplasm of HCC cell lines as RBMY contains a sequence that matches the consensus sequence of a nuclear export signal recognised by nuclear export chromosome region maintenance 1 (CRM1) (Chua *et al.*, 2015). This study showed that cytoplasmic RBMY has kinase activity as it is involved in the inactivation a repressor of the Wnt pathway by serine phosphorylation, but it is unclear whether the kinase activity is direct or indirect. This could lead to malignancy as cells hereby proliferate in an uncontrolled manner due to impeding of β-catenin destruction (Chua *et al.*, 2015). Further research needs to be conducted as these mechanisms could apply to other malignancies, more specifically head and neck cancers.

1.5 Project outline

Aims

As the roles of hZFY variants remain obscure, a few hypotheses have been proposed by (Nadège Vernet *et al.*, 2016) and Ellis-Fenton laboratory.

Based on the background information, we hypothesise that:

a) The difference in transactivation ability between human short and long ZFY isoforms is due to the presence of 9aa TAD motifs within the alternatively spliced second coding exon

b) The short ZFY isoform antagonises the transactivation activity of the long isoform either passively as a competitive inhibitor (by competing for binding at downstream promoters and preventing access by ZFY-long isoform) or by actively recruiting co-repressor complexes

c) hRBMY acts as a splicing factor to promote the production of the hZFY-short isoform
d) hZFY-short isoform will function as an oncogene when transfected into mammalian cell lines

The original project plan was addressing hypotheses (a) and (b) with another student focusing on RBMY function.

Using commercially synthesized constructs, the goals were:

- To express and purify the human ZFY-long and ZFY-short acidic domains (without DNA binding domains) in bacterial culture and do preliminary structural characterization e.g., circular dichroism measurement to assess protein folding
- To use purified human ZFY-long and ZFY-short acidic in a capture experiment to identify potential binding partners
- To use reporter constructs to validate the transactivation data for human ZFY-long and ZFY-short isoforms in mammalian cells, as thus far this work has been done in yeast reporter system
- To search for 9aa TAD domains within the human ZFY-long and ZFY-short sequences via bioinformatic analysis in preparation for future mutagenesis work

Owing to COVID-19 restrictions, the bulk of the laboratory work could not be completed, and thus the project goals were changed to encompass investigation of ZFY structure and sequence conservation across a wide range of species (including 9aa TAD and zinc finger prediction) to highlight key conserved regions likely to be functionally important in its activity. The goals therefore were:

- To use genetic sequence databases to identify ZFY DNA and protein sequences across various species
- To use in silico analysis for sequence alignment, sequence conservation and phylogeny across various species
- To characterise the structure of ZFY using similarities between the various species
- To screen and locate a 9aa TAD and locate binding interfaces of the DNA binding domain.

2 Materials & Method

2.1 ZFY protein and cDNA (nucleotide) search and alignment

The primary database used for ZFY protein and cDNA sequences are described in **Table 1** and **Table 2**. The search terms used were ‘ZFY’ and ‘zinc finger Y-chromosomal protein’, and the search was narrowed down by taxonomic groups to vertebrates using the ‘Results by taxon’ feature on NCBI taking the number of protein entries to 2826 and 466 species, and the number of gene entries to 577 and the 259 species. As most entries were only partial sequences, the number of species with full ZFY sequences was lower resulting in access to only 28 land vertebrates and 16 fish species as shown by **Table 1** and **Table 2**.

For the protein and cDNA sequences absent on NCBI, previous published papers, Uniprot and EMBL-EBI were the alternative databases used to obtain the sequences. In addition, to locate the exon length, we utilised the NCBI gene database and the gene table format to locate the gene.

Table 1. Land vertebrate **ZFY/ ZFX** ***protein and nucleotide sequences.*** The table illustrates each taxa given their common name and biological classification of each taxa, so that the arrangement is sequential. The ***nucleotide/cDNA*** sequence database is in *Italic*. Subsequently, the sequences were used for alignment. As *Neophocaena asiaeorientalis* is a cetacean, it was grouped as Cetartiodactyla along with Artiodactyla so for the entirety of the experiment they were examined with the rest of the land vertebrates.

Binomial Nomenclature	**Common Species Nomenclature**	**Taxonomic Classification**	**Protein Name**	**Database for Protein & *Nucleotide/ cDNA* Sequence**	**Database Protein Accession**	**Database Nucleotide Accession**
Homo sapiens	Human	Chordata/ Mammalia/ Primates/ Hominidae/ *Homo*	ZFY	NCBI GenPept - *NCBI GenBank*	NP_001356631.1	NM_003411.4
Pan troglodytes	Chimpanzee	Chordata/ Mammalia/ Primates/ Hominidae/ *Pan*	ZFY	NCBI GenPept - *NCBI GenBank*	XP_009443992.1	XM_009445712.3
Gorilla gorilla	Gorilla	Chordata/ Mammalia/ Primates/ Hominidae/ *Gorilla*	ZFY	NCBI GenPept - *NCBI GenBank*	Q52V16.1	AH014841.2
Macaca mulatta	Rhesus monkey	Chordata/ Mammalia/ Primates/ Cercopithecidae/ *Macaca*	ZFY	NCBI GenPept - *NCBI GenBank*	XP_014984082.1	XM_015128596.2
Trachypithecus francoisi	François' langur	Chordata/ Mammalia/ Primates/ Cercopithecidae/ *Trachypithecus*	ZFY	NCBI GenPept - *NCBI GenBank*	XP_033067617.1	XM_033211726.1
Papio anubis	Olive baboon	Chordata/ Mammalia/ Primates/ Cercopithecidae/ *Papio*	ZFY isoform X1	NCBI GenPept - *NCBI GenBank*	XP_031516968.1	XM_031661108.1
Chlorocebus sabaeus	Green monkey	Chordata/ Mammalia/ Primates/ Cercopithecidae/ *Chlorocebus*	ZFY isoform X1	NCBI GenPept - *NCBI GenBank*	XP_008017167.1	XM_008018975.1
Rhinopithecus roxellana	Golden snub-nosed monkey	Chordata/ Mammalia/ Primates/ Cercopithecidae/ Rhinopithecus	ZFY isoform X1	NCBI GenPept - *NCBI GenBank*	XP_030782172.1	XM_030926312.1
Hylobates moloch	Silvery gibbon	Chordata/ Mammalia/ Primates/ Hylobatidae/ *Hylobates*	ZFY isoform X1	NCBI GenPept - *NCBI GenBank*	XP_032612406.1	XM_032756515.1
Callithrix jacchus	White-tufted-ear marmoset	Chordata/ Mammalia/ Primates/ Callitrichidae/ *Callithrix*	ZFY	NCBI GenPept - *NCBI GenBank*	XP_035145821.1	FJ527008.1
Mus musculus	Mouse	Chordata/ Mammalia/ Rodentia/ Muridae/ *Mus*	Zfy1	NCBI GenPept - *NCBI GenBank*	P10925.3	NM_009570.4
			Zfy2		(P20662.2)	NM_009571.2
Rattus norvegicus	Brown rat	Chordata/ Mammalia/ Rodentia/ Muridae/ *Rattus*	Predicted ZFY2 isoform X1	NCBI GenPept - *NCBI GenBank*	XP_008771898.1	XM_017602438.1
Marmota marmota	Alpine marmot	Chordata/ Mammalia/ Rodentia/ Sciuridae/	ZFY	NCBI GenPept -	XP_015343506.1	XM_015488020.1

		Marmota		*NCBI GenBank*		
Bos taurus	Cattle	Chordata/ Mammalia/ Artiodactyla/ Bovidae/ *Bos*	ZFY	NCBI GenPept - *NCBI GenBank*	Q95LI3.1	NM_177491.1
Bison bison	American bison	Chordata/ Mammalia/ Artiodactyla/ Bovidae/ *Bison*	ZFY isoform X1	NCBI GenPept - *NCBI GenBank*	XP_010855418.1	XM_010857116.1
Capra hircus	Goat	Chordata/ Mammalia/ Artiodactyla/ Bovidae/ *Capra*	Predicted ZFY isoform X1	NCBI GenPept - *NCBI GenBank*	XP_017900383.1	XM_018044894.1
Cervus elaphus	Red deer	Chordata/ Mammalia/ Artiodactyla/ Cervidae/ *Cervus*	ZFY	NCBI GenPept - *NCBI GenBank*	AMY96563.1	KU041539.1
Odocoileus virginianus	White-tailed deer	Chordata/ Mammalia/ Artiodactyla/ Cervidae/ *Odocoileus*	ZFY isoform X1	NCBI GenPept - *NCBI GenBank*	XP_020759307.1	XM_020903648.1
Sus scrofa	Pig	Chordata/ Mammalia/ Artiodactyla/ Suidae/ *Sus*	ZFY	UniProtKB - *EMBL-EBI*	F1SPY3	FQ670201.4
Neophocaena asiaeorientalis	Narrow-ridged finless porpoise	Chordata/ Mammalia/ Artiodactyla/ Phoconidae/ *Neophocaena*	ZFY	NCBI GenPept - *NCBI GenBank*	XP_024612082.1	XM_024756314.1
Canis lupus	Dog	Chordata/ Mammalia/ Carnivora/ Canidae/ *Canis*	ZFY	NCBI GenPept - *NCBI GenBank*	AKI82174.1	JX964866.1[*]
Mustela erminea	Ermine	Chordata/ Mammalia/ Carnivora/ Mustelidae/ *Mustela*	ZFX-like isoform X1	NCBI GenPept - *NCBI GenBank*	XP_032187800.1	XM_032331909.1
Loxodonta africana	African savanna elephant	Chordata/ Mammalia/ Proboscidea/ Elephsntidae/ *Loxodonta*	ZFY	NCBI GenPept - *NCBI GenBank*	JAC06687.1	GATM01000012.1[*]
Equus caballus	Horse	Chordata/ Mammalia/ Perissodactyla/ Equidae/ *Equus*	Predicted ZFY	Obtained via (Jane *et al.*, 2018)	No_accession[†]	No_accession[†]
Ornithorhynchus anatinus	Platypus	Chordata/ Mammalia/ Monotremata/ Ornithorhynchidae/ *Ornithorhynchus*	ZFY isoform X2	NCBI GenPept - *NCBI GenBank*	XP_028935710.1	XM_029079877.1
Monodelphis domestica	Gray short-tailed opossum	Chordata/ Mammalia/ Didelphimorphia/ Didelphidae/ *Monodelphis*	ZFY	NCBI GenPept - *NCBI GenBank*	XP_016288863.1	XM_016433377.1
Gallus gallus	Chicken	Chordata/ Aves/ Galliformes/ Phasianidae/ *Gallus*	ZFX isoform X1	NCBI GenPept - *NCBI GenBank*	XP_015127980.1	XM_015272494.2
Xenopus laevis	African clawed toad	Chordata/ Amphibia/ Anura/ Pipidae/ *Xenopus*	ZFY1	NCBI GenPept - *NCBI GenBank*	Q01611.1	BC070611.1[*]

---

[*] cDNA was manually edited using the CDS range from respective database to create the ZFY cDNA sequence

[†] Horse ZFY cDNA sequence prediction published by (Jane *et al.*, 2018). This was derived by using hZFY NM_003411.4 cDNA sequence and pairwise aligning it with the horse Y chromosome and manually editing intron/exon boundaries. The horse protein sequence was derived via horse cDNA in silico translation.

Table 2. Fish* ZFY/ZFX *sequences. The table shows the different fish species ZFY/ZFX sequences available. The sequences were obtained from the databases indicated. The ***nucleotide/cDNA*** sequence database is in *Italic.*

Binomial Nomenclature	**Common Species Nomenclature**	**Taxonomic Classification**	**Protein Name**	**Database for obtaining Protein & *Nucleotide* Sequence**	**Database Protein Accession**	**Database Nucleotide Accession**
Collichthys lucidus	Big head croaker	Chordata/ Actinopterygii/ Perciformes/ Sciaenidae/ *Collichthys*	ZFY1	NCBI GenPept - *NCBI GenBank*	TKS65875.1	ML241175.1[‡]
Perca flavescens	Yellow perch	Chordata/ Actinopterygii/ Perciformes/ Percidae/ *Perca*	ZFY1-like isoform X1	NCBI GenPept - *NCBI GenBank*	XP_028451227.1	XM_028595426.1
Parambassis ranga	Indian glassy fish	Chordata/ Actinopterygii/ Perciformes/ Ambassidae/ *Parambassis*	ZFY1-like isoform X1	NCBI GenPept - *NCBI GenBank*	XP_028276673.1	XM_028420872.1
Larimichthys crocea	Large yellow croaker	Chordata/ Actinopterygii/ Perciformes/ Sciaenidae/ *Larimichthys*	ZFY isoform X1	NCBI GenPept - *NCBI GenBank*	XP_010749798.1	XM_010751496.3
Amphiprion ocellaris	Ocellaris clownfish	Chordata/ Actinopterygii/ Perciformes/ Pomacentridae/ *Amphiprion*	ZFY1-like isoform X1	NCBI GenPept - *NCBI GenBank*	XP_023133903.1	XM_023278135.1
Oncorhynchus tshawytscha	Chinook salmon	Chordata/ Actinopterygii/ Salmoniformes/ Salmonidae/ *Oncorhynchus*	ZFY1-like	NCBI GenPept - *NCBI GenBank*	XP_024253620.1	XM_024397852.1
Salvelinus alpinus	Arctic char	Chordata/ Actinopterygii/ Salmoniformes/ Salmonidae/ *Salvelinus*	ZFY1	NCBI GenPept - *NCBI GenBank*	XP_023843891.1	XM_023988123.1
Oncorhynchus kisutch	Coho salmon	Chordata/ Actinopterygii/ Salmoniformes/ Salmonidae/ *Oncorhynchus*	ZFY1-like	NCBI GenPept - *NCBI GenBank*	XP_020321060.1	XM_020464688.2
Maylandia zebra	Zebra mbuna	Chordata/ Actinopterygii/ Cichliformes/ Cichlidae/ *Maylandia*	ZFY1 isoform X1	NCBI GenPept - *NCBI GenBank*	XP_004564062.1	XM_004564005.5
Astatotilapia calliptera	Eastern river bream	Chordata/ Actinopterygii/ Cichliformes/ Cichlidae/ *Astatotilapia*	ZFY1-like isoform X1	NCBI GenPept - *NCBI GenBank*	XP_026038267.1	XM_026182482.1
Takifugu rubripes	Japanese puffer	Chordata/ Actinopterygii/ Tetraodointiformes/ Tetraodontidae/ *Takifugu*	ZFY isoform X1	NCBI GenPept - *NCBI GenBank*	XP_011609888.1	XM_011611586.2
Betta splendens	Siamese fighting fish	Chordata/ Actinopterygii/ Anabantiformes/ Osphronemidae/ *Betta*	ZFY isoform X1	NCBI GenPept - *NCBI GenBank*	XP_029029380.1	XM_029173547.1
Denticeps clupeoides	Denticle herring	Chordata/ Actinopterygii/ Clupeiformes/ Denticipitidae/ *Denticeps*	ZFY1-ike	NCBI GenPept - *NCBI GenBank*	XP_028839070.1	XM_028983237.1
Anabarilius grahami	Kanglang fish	Chordata/ Actinopterygii/ Cypriniformes/ Cyprinidae/	ZFY1	NCBI GenPept -	ROL53794.1	RJVU01007700.1[‡]

[‡] cDNA was manually edited using the CDS range from respective database to create the ZFY cDNA sequence

		Anabarilius		*NCBI GenBank*		
Seriola dorsalis	California yellowtail	Chordata/ Actinopterygii/ Carangiformes/ Carangidae/ *Seriola*	ZFY1-like isoform X1	NCBI GenPept - *NCBI GenBank*	XP_023277193.1	XM_023421425.1
Cynoglossus semilaevis	Tongue sole	Chordata/ Actinopterygii/ Pleuronectiformes/ Cynoglossidae/ *Cynoglossus*	ZFY1 isoform X1	NCBI GenPept - *NCBI GenBank*	XP_008331409.1	XM_008333187.3

The protein and cDNA (nucleotide) sequences were aligned using the program Molecular Evolutionary Genetic Analysis (MEGAX). The protein and nucleotide sequences were aligned using either the ClustalW or the MUSCLE alignment tools with default parameters as shown by **Table 3** and **Table 4**. As ClustalW alignments treated gaps better, we used this tool for the phylogeny analysis of our project but for completeness, we included the MUSCLE alignment parameters. To calculate the percentage identity of the sequences, we used the formula:

$$\%\ identity = \frac{no.\ of\ conserved\ sites\ from\ the\ multiple\ sequence\ alignment}{all\ of\ the\ sites\ of\ the\ ZFY\ sequence} \times 100$$

Subsequently, the resulting alignments were then used to create a model the protein and nucleotide alignments and the parameters were all the same for the protein sequence alignments, but the nucleotide sequence was different in that the selected nucleotide positions needed to be specified and the default parameter was applied. The most suitable substitution model exhibited the lowest Bayesian Information Criterion (BIC) score. Therefore, the lowest BIC was used to compute the suitable substitution model parameter during the construction of the phylogenetic tree.

Table 3. ClustalW/ ClustalW (Codons) alignment parameters. The table shows the parameters used for the alignment of the vertebrate ZFY/ZFX protein and nucleotide sequences to outline the conservation of certain amino acids and bases.

Alignment	
Pairwise Alignment	
Gap Opening Penalty	10.00
Gap Extension Penalty	0.10
Multiple Alignment	
Gap Opening Penalty	10.00
Gap Extension Penalty	0.20
Weight	
Protein Weight Matrix	Gonnet
Residue-specific Penalties	ON
Hydrophilic Penalties	ON
Gap Separation Matrix	4
End Gap Separation	OFF
Use Negative Matrix	OFF
Delay Divergent Cutoff	30
Keep Predefined Gap	Not selected

Table 4 .MUSCLE alignment parameters. The table exhibits default parameters used to create alignments by the MUSCLE alignment tool. This includes protein and nucleotide alignments, and the options are all the same except for the nucleotide MUSCLE alignment where the genetic code advanced option is present. [ ] indicates nucleotide specific parameters

Gap Penalties	
Gap Open	-2.90
Gap Extend	0.00
Hydrophobicity Multiplier	1.20
Memory/ Iterations	
Max Memory in MB	2048
Max Iterations	16
Advanced Options	
[Genetic Code]	[Standard]
Cluster Method (Iterations 1,2)	UPGMA
Cluster Method (Other Iterations)	UPGMA
Min Diag Length (Lambda)	24

For the phylogenetic analysis of the CLUSTALW aligned nucleotide and protein sequences, the tree used was the automatic setting which used the Neighbour-joining tree. The statistical method used was the Maximum Likelihood and the substitution type was dependent on whether the sequences inputted were protein or nucleotide sequence. The parameters for the data subset to use and system resource usage were default parameters, with all sites of the sequences being used including gaps, no site coverage cutoff, no branch swap filter and 3 threads used.

2.2 ZFY Phylogenetic Tree

Thereafter, maximum-likelihood method was used to construct series of land vertebrate ZFY phylogenetic trees, as it allows us to understand what the phylogenetic tree for ZFY proteins looks like. In addition, the Bootstrap method was used for the test of phylogeny and the number of Bootstrap replications was set to 100. There was a slight variation in the substitution model as the protein sequences used Jones-Taylor-Thornton model and the nucleotide sequence used the Kimura 2-parameter model. However, both methods used a discrete Gamma distribution (5 categories) to model evolutionary rate differences among sites. The phylogenetic tree was then rooted so that the outgroup species was highlighted.

2.3 ZFY 9aaTAD Motif and DNA binding site prediction

ZFY protein sequences were analysed by the ‘Nine Amino Acids Transactivation Domain 9aaTAD Prediction Tool’ (Piskáček, 2020) at *www.med.muni.cz/9aaTAD/index.php* . For putative transactivation domain analysis of mammalian transcription factors, the ‘Moderately stringent Pattern’ was the most suitable ‘[MDENQSTYG]{KRHCGP}[ILVFWM]{KRHCGP}{CGP}{CGP}[ILVFWM]{CGP}{CGP}’. The algorithm harnesses a refinement criteria RC1- RC12 which allows the elimination of false positive results that might be encountered, ensuring precision of the prediction. This is done by comparing the sequence of interest with the hydrophobic profiles of experimentally proven TADs and false positive results located out of region. However, the refinement criteria are ever changing as new orthologs of 9aa TADs and 9aa TADs are identified. Thus, the refinement criteria we used were as following:

Refinement Criteria	**Position Criteria**
1	Two staple hydrophobic positions (position **3** and **7**) with at least one hydrophobic amino acid neighbouring
2	
3	Hydrophilic amino acid (s) between the core hydrophobic residues
4	Pattern limits overall hydrophobic and hydrophilic amino acids (However, serine and threonine residues included in both criteria)
5	
6	Consecutive hydrophobic/hydrophilic residues limited
7	
8	Pattern containing NQRKH amino acids in particular region limited as usually suggestive of false positive prediction
9	
10	
11	Prohibition of helix breaking and bridging residues as 9aa TAD possesses predicted α-helical secondary structure
12	

Table 5. 9aa TAD motif algorithm refinement criteria. The protein sequence is analysed using the criteria above to predict sequences likely to be 9aa TAD motifs.

For the zinc finger domain prediction, we used the tool located on http://zf.princeton.edu/index.php. To determine which regions were zinc finger motifs/domains, the website uses hmmsearch program which incorporates the HMMER 2.3.2 protein sequence homology search software and an HMM profile for ZF-C2H2 family (Finn *et al.*, 2009; Persikov and Singh, 2011). The HMMER searches sequence databases for homologs of our protein sequence and makes alignments, allowing an accurate prediction of zinc finger domains. This used a zinc finger score (ZF) to determine which regions portrayed high scores as the threshold score for confidence is $\geq$17.7, defined by Pfam (Finn *et al.*, 2009).

Furthermore, zinc finger motif prediction was executed via analysing the binding profile for the domains using the F2+F3 union option on the B1H resources. This analysed each sequence which gave us the zinc finger domain sequence that was interpreted to bind DNA and predicted a target by producing a sequence logo and the nucleotide predicted to bind had the highest bit score. To find which four amino acids of the domain sequence bound to DNA, we used the known Krüppel-type zinc finger structure (Stubbs, Sun and Caetano-Anolles, 2011) and this gave us the four positions bind DNA which were the -1, 2, 3 and 6 relative to the zinc finger helix.

2.4 Transformation and plasmid minipreps

Backbone vector	Insert	Tag	Intended purpose
pcDNA3.1(+)	hZFY-long (full length)	HA	Isolation and purification within mammalian cell lines
	hZFY-short isoform		
	hZFY-long (full length)	eGFP	Detection and analysis of function within mammalian cell lines
	hZFY-short isoform		
pET-15b	hZFY-long acidic domain	Histidine	Bacterial expression Purification by metal affinity chromatography Transactivation and repression analysis of acidic domains Locating acidic domain binding partners
	hZFY-short acidic domain		
pFN26A (BIND)	hZFY-long acidic domain	Luciferase	Gene expression assays of acidic domains at transcriptional level by using luciferase reporter system (Gal4) in mammalian cell lines
	hZFY-short acidic domain		

Table 6. Plasmid vectors. The table shows plasmid vector maps for the 3 plasmid vectors which are pcDNA3.1+ (Invitrogen) (Cat. # **V79020**), pET15b (Novagen [EMD Millipore]) (**69661**) and pFN26A (Promega) (**E1380**) used to create 8 constructs containing 4 tags for initial bacterial transformation. Different ZFY genes (full length [with and without DNA binding domain]) and short variant [with and without DNA binding domain]) were inserted into the respective vectors.

Human *ZFY* (h*ZFY*) genes were inserted into vectors using various restriction enzymes which included: XhoI and XbaI for all pcDNA3.1(+) vector backbones, XhoI and BlpI for all pET15b vector backbones and AsiSI and BlpI for all pFN26A (BIND) vector backbones and ordered from the respective vendors mentioned in the figure legend of **Table 6.** 1µL of each plasmid was transformed into 50µL aliquots of T7 Express Competent *E. coli* (High Efficiency) (**C266I**) and successively heat-shocked at 42°C for 10 seconds and recovered on ice for 5 minutes. Subsequently, 950µL SOC media (**B9020S**) was added to the mixture and incubated at 37°C for 60 minutes in a Multitron Standard shaking incubator at 200 revolution per minute (rpm). The cells were then pelleted by centrifugation at 1500 x g for 1 minute and resuspended in 200 µL in SOC media. 100µL of the mixture was then loaded onto Lysogeny Broth agar (LB-agar) plates loaded with 100µg/mL ampicillin at 37°C overnight. Following, isolated *E. coli* colonies were used to create 3mL of LB-ampicillin inoculated starter cultures that were incubated overnight in a shaking incubator at 200rpm and 37°C. Subsequently, 500µL of the culture was used to create 1mL of ZFY glycerol stocks containing 500µL of 50% glycerol and were stored in cryovials at -80°C. The remainder of each *E. coli* starter culture were then used for the plasmid extraction using a QIAprep Spin Miniprep Kit (**27104**).

2.5 *E. coli* Growth

Glycerol stocks of *E. coli* expressing pET-15b plasmids (T7 promoter, N-terminal His tag and h*ZFY* long/ short AD) were used to 250mL starter cultures with LB growth media (40% tryptone, 40% sodium chloride (NaCl) and 20% yeast extract) and 100µg/mL ampicillin. Subsequently, the cultures were grown overnight and shaken at 37°C and 200rpm in a Multitron Standard shaking incubator. For the initiation of the bacterial culture growth, OD_{600} was measured in a Biomate 3S Spectrophotometer to calculate the volume of starter needed to start at OD_{600} of 0.1. This was calculated by dividing the absorbance of the 250mL starter culture by 0.1, and then use this value to divide by the volume of the LB growth media (1L).

2.6 ZFY Induction and Isolation

105mL of *E. coli* starter culture expressing hZFY-long and 104mL of hZFY-short expressing *E. coli* starter culture were added to traditional separate 1L LB-ampicillin (40% tryptone, 40% sodium chloride (NaCl), 20% yeast extract and

100µg/mL ampicillin), shaken at 200rpm until mid-log phase of OD_{600}= 0.7, and induced overnight with isopropyl β-d-thiogalactopyranoside (IPTG) (1µL/mL) at 30°C. Cells were harvested by centrifugation of the 1L solutions using a JA10 rotor Beckman Avanti J-25 centrifuge at 4°C for 30 minutes, sonicated for 10 minutes using a SoniPrep 150 (with 30 second on and off intervals) and ultracentrifuged in a Beckman ultracentrifuge for an hour in a 70Ti rotor at 70,000rpm and 4°C, and the pellets were frozen at -80°C and suspended in lysis buffer (50mM Tris, 300mM NaCl at pH8) with 1 µL/mL EDTA-free Pierce Protease and Phosphatase Inhibitor Mini Tablets (**A32965**). Thereafter, the supernatant was isolated by nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography **(88221**) using 20mM phosphate buffered saline (pH 7.4) with 10mM imidazole (equilibration buffer), 25mM imidazole (wash buffer) and eluted with 250mM imidazole (elution buffer).

2.7 *Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)*

To analyse the contents purified by the nickel-NTA affinity chromatography, a handcast Tris-glycine gel and precast Bis-Tris SDS PAGE gels were utilised.

The Tris-glycine gel was composed of a 15% resolving gel containing 40% acrylamide, 1.5M Tris-HCl at pH8.8, 10% SDS, 36% v/v DH_2O, 0.05% v/v TEMED (**1610800EDU**) and 10% APS solution (**1610700**), and a 5% stacking gel composed of 40% acrylamide, 0.5M Tris-HCl, 10% SDS, 63% v/v DH_2O, 0.1 v/v TEMED and 10% APS solution. To add on, the 1X Tris Glycine running buffer for the gel was composed of 0.0248M tris base, 0.19M glycine, 0.0035M v/v SDS solution final concentration, diluted from a 10X Tris Glycine buffer stock. Subsequently, a 2X Laemmli sample buffer stock was diluted with the sample (1:1 dilution) to 1X and the final concentration of the Laemmli sample buffer components were 2% SDS, 10% glycerol, 0.063M tris base, 5% β-mercaptoethanol and 0.002% bromophenol blue at pH 6.8 and heated to 85°C for 5 minutes in a Grant Instruments JB Academy unstirred water bath. The gel was submerged in the Tris-Glycine running buffer, and 5µL of protein ladder was added to the first lane and 20µL heated sample mixtures were loaded into the other lanes. The gel was run at 150V for 1hr and subsequently stained for 1hr by Coomassie blue (1.21mM Coomassie Brilliant Blue, 50% v/v methanol, 10% v/v glacial acetic acid, 40% v/v DH_2O) and destained overnight on a see-saw rocker with Coomassie destain (50% v/v DH2O, 40% methanol v/v, 10% v/v acetic acid).

The NuPAGE Bis-Tris gel (**NP0321PK2**) had a polyacrylamide of 4-12% and the 20X stock NuPAGE MOPS SDS running buffer [**NP0001**] was diluted with distilled water to 1X working solution with a final concentration 2.5mM MOPS, 2.5mM Tris Base, 0.005% SDS, 0.05mM EDTA at pH 7.7. Furthermore, the 4X stock NuPAGE

LDS Sample Buffer [**NP0007**] (35.25mM Tris base, 26.50mM Tris-HCl, 0.5% LDS,0.13mM EDTA, 0.06mM SERVA Blue G-250, 0.04mM phenol red at pH 8.5,) was diluted 1:3 sample buffer to sample dilution (5µL sample buffer and 15µL sample) and heated for 10 minutes to 70°C. The gel was submerged in the Bis-Tris running buffer and the wells were loaded with the heated sample mixtures and protein ladder. The runtime, voltage and visualisation were the same as the Tris-Glycine method

For visualisation, a G:Box F3 gel imaging system was used, and the relative mobility of the bands were compared to the PageRuler Prestained Protein Ladder (**26616**) which has 10 recombinant proteins ranging from 10kDa to 180kDa, and PageRuler Prestained Protein Ladder (**26619**), which has 9 recombinant proteins ranging from 10kDa to 250kDa.

2.8 Western blot analysis

For the separation process of the western blot, it was done using the tris-glycine gel described without Coomassie staining. For hZFY-long AD protein, the acrylamide gel was incubated in Bjerrum and Schafer-Nielsen transfer buffer (comprised of 48mM tris, 39mM glycine and 20% methanol) for 15 minutes and transferred onto Polyvinylidene fluoride (PVDF) membrane that was activated in 100% methanol for 1 minute and equilibrated in Bjerrum and Schafer-Nielsen transfer buffer for 15 minutes. Subsequently, the membrane and gel were sandwiched between equilibrated filter paper (also equilibrated in Bjerrum and Schafer-Nielsen transfer buffer for 15 minutes) on a Trans-Blot SD Semi-Dry Transfer Cell for 20 minutes at 15V. Following, the membrane was retrieved and blocked with 10mL of 5% bovine serum albumin (BSA) diluted in tris-buffered saline (20mM tris and 150mM NaCl) with 0.025% triton (TBST) for 60 minutes at room temperature on a platform shaker. Thereafter, the blocked membrane was incubated with an anti His-tag primary antibody overnight at 4°C. Successively, the PVDF membrane was washed five times with TBST for a total of 25 minutes, and the membrane was incubated for 120 minutes in 1:10,000 diluted HRP-conjugated secondary antibody with 5% BSA-TBST blocking buffer. Then, the membrane was washed a further 5 times in TBST and incubated for 5 minutes in 5mL total solution of Pierce enhanced chemiluminescence (ECL) western blotting Substrate (**32209**) for imaging in a G:Box F3 gel imaging system.

For hZFY-short AD, we equilibrated nitrocellulose membrane, gel, and filter paper separately in 50mL Towbin transfer buffer (25mM Tris, 190mM glycine and 20% methanol, pH 8.3) for 15 minutes. The gel and nitrocellulose were sandwiched between the equilibrated filter paper in Pierce Power Blotter for 5 minutes at 25V and 1.3A current, and the membrane with was briefly Ponceau stained to visualise if the transfer was successful. The membrane was then washed with 10 mL TBST (20mM tris and 150mM NaCl) with 0.025% triton (TBST) for 5 minutes to remove the stain and subsequently blocked in a 1% milk TBST solution for 60 minutes at room temperature on a platform shaker. Then, the membrane was incubated with 0.1% peroxidase-conjugated anti-his antibody in TBST for 60 minutes at room temperature. The antibody was removed, and the membrane was washed 5 times for a total of 25 minutes and detected using the same method as the PVDF membrane.

3 Results

3.1 ZFY exon length is conserved in majority of the land vertebrates

We sought to investigate ZFY proteins across as many species as we could access and 45 ZFY protein sequences in total (29 land vertebrate sequences and 16 fish sequences) were used for our investigation. Most animal species have one *ZFY* gene present, with the exception of mice as two *Zfy* genes are present. Although the majority of animal species have 7 coding exons, the exon lengths were not always the same length. Therefore, the length of the amino acids between the species varied as the shortest ZFY protein was 701 amino acids long (*T. rubripes*) and the longest ZFY protein was 814 amino acids long (*O. anatinus*). The consensus shown by land vertebrates was exons 1, 3, and 6 were highly conserved in length as **Table 7** showed the length of these exons were all the same excluding *X. laevis* exon 1. It is apparent *X. laevis* has one more codon in comparison to the rest of the land vertebrates.

Exons 4 and 7 were moderately conserved as they had varied lengths in some but not all of the land vertebrates. For example, exon 4 of *T. francoisi* had 44 codons and *N. asiaeorientalis* had 47 codons which were different to majority 48 codons for the rest of the land vertebrates. Although *N. asiaeorientalis* is technically not a land vertebrate, it was grouped in as it is considered a Cetartiodactyla. In addition, exon 7 of land vertebrates was typically 394 codons but *O. anatinus* had 102 codons and *G. gallus* had 395, and finally, *M. musculus, R. norvegicus* and *X. laevis* both exons 4 and 7 were shorter in length in comparison to the other land vertebrates. Furthermore, exons 2 and 5 were the least conserved in exonic length as the majority of the land vertebrate species had varying exon lengths. These differences in the exon lengths were suggestive of either truncation of some residues in some of the ZFY proteins due to a base deletion or insertions where the exon length was significantly larger for instance.

Table 7. Land vertebrates ZFY exon length. The table illustrates the exon size of the ZFY transcripts. There are typically 7 coding exons, and the length of base pairs represents the coding regions of the exons. There are 28 land vertebrates in total with 26 of the 28 species were mammals (*M. musculus* has two Zfy sequences) and the rest of the species were bird and amphibian. N/A represents 'Not Available' for species that do not have any information regarding the exons from databases.

	ZFY Coding Exon Length (bp)						
Species	**Exon 1**	**Exon 2**	**Exon 3**	**Exon 4**	**Exon 5**	**Exon 6**	**Exon 7**
H. sapiens	61	573	150	144	153	141	1184
P. troglodytes	61	573	150	144	153	141	1184
G. gorilla	61	573	150	144	153	141	1184
M. mulatta	61	573	150	144	153	141	1184
T. francoisi	61	572	150	132	153	141	1184
P. anubis	61	573	150	144	153	141	1184
C. sabaeus	61	573	150	144	153	141	1184
R. roxellana	61	573	150	144	153	141	1184
H. moloch	61	573	150	144	153	141	1184
C. jacchus	61	567	150	144	153	141	1184
M. musculus ZFY1	61	552	150	141	123	141	1181
M. musculus ZFY2	61	552	150	141	123	141	1166
R. norvegicus	61	552	150	141	123	141	1178
M. marmota	61	558	150	144	153	141	1184
B. taurus	61	570	150	144	156	141	1184
B. bison	61	570	150	144	156	141	1184
C. hircus	61	570	150	144	156	141	1184
C. elaphus	N/A	N/A	N/A	N/A	N/A	N/A	N/A
O. virginianus	61	570	150	144	153	141	1184
S. scrofa	61	561	150	144	156	141	1184
N. asiaeorientalis	61	570	150	141	156	141	1184
C. lupus	N/A	N/A	N/A	N/A	N/A	N/A	N/A
M. erminea	61	570	150	144	156	141	1184
L. africana	N/A	N/A	N/A	N/A	N/A	N/A	N/A
E. caballus	N/A	N/A	N/A	N/A	N/A	N/A	N/A
O. anatinus	61	606	150	144	135	141	1208
M. domestica	61	570	150	144	135	141	1184
G. gallus	61	567	150	144	135	141	1187
X. laevis	64	558	150	141	138	141	1517

Fish species on the other hand exhibited a different pattern in comparison to the land vertebrates as shown in **Table 8**, but some of the exon lengths were the same as land vertebrates. Exon 1 amongst the fish was highly conserved and in comparison, with most land vertebrates, it contained 1-2 additional codons. Moreover, exon 2 of fish species was less conserved and significantly shorter than land vertebrate exon 2 as they had nearly twice the number of base pairs. Therefore, for the entirety of this project, we focused exclusively on land vertebrate (and *N. asiaeorientalis*) ZFY protein and nucleotide sequences.

Exon 3 was shown to be the most conserved length wise in all species (fish species and land vertebrate species) as the exon had 150 base pairs. This showed that the exon encodes for exactly 50 codons throughout all ZFY proteins. In addition, exons 4 and 5 were shorter in fish species compared to the land vertebrates, but the difference in the exon length was not strikingly significant. Exons 6 and 7 of fish species were generally longer than land vertebrates. However, exon 6 was roughly 1-2 codons longer whereas exon 7 was significantly longer in fish than some of the land vertebrates by a few 100 base pairs.

Although the exon lengths highlighted some degree of conservation, this did not highlight differences in nucleotide or protein as there was likely gene evolution. Therefore, protein and nucleotide alignments we carried forward bioinformatic techniques to emphasise deviation or conservation of the ZFY sequences.

Table 8. Fish species ZFY exon length. The table illustrates the exonic length of fish ZFY and they have typically 7 coding exons with varying lengths. There are 16 fish species in total and for the fish with a missing exon , they have been indicated by the (-) symbol.

	ZFY Length (bp)						
Species	**Exon 1**	**Exon 2**	**Exon 3**	**Exon 4**	**Exon 5**	**Exon 6**	**Exon 7**
C. lucidus	67	315	150	129	135	144	1301
P. flavescens	67	330	150	129	135	144	1301
P. ranga	67	315	150	129	135	144	1301
L. crocea	67	315	150	129	135	144	1301
A. ocellaris	67	315	150	129	135	144	1301
O. tshawytscha	67	336	150	108	135	147	1280
S. alpinus	67	336	150	108	135	147	1280
O. kisutch	67	336	150	108	135	147	1280
M. zebra	67	318	150	129	135	144	1301
A. calliptera	67	318	150	129	135	144	1301
T. rubripes	67	315	150	129	-	144	1301
B. splendens	67	315	150	129	135	144	1301
D. clupeoides	67	333	150	120	135	144	1319
A. grahami	67	345	150	120	135	138	1307
S. dorsalis	67	315	150	129	135	144	1301
C. semilaevis	67	315	150	126	-	144	1337

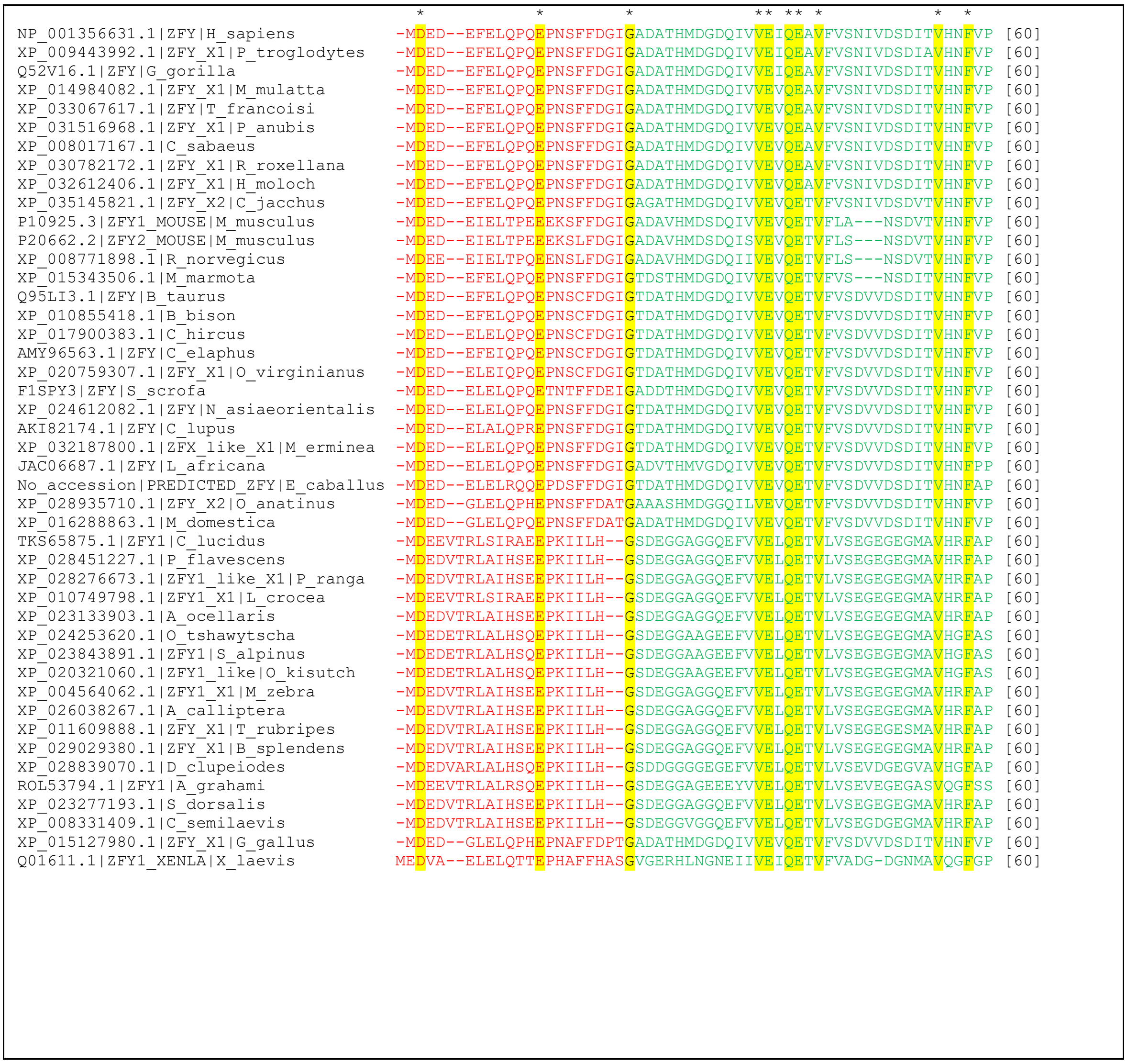
3.2 Vertebrates ZFY protein alignment and phylogeny

We carried out multiple sequence alignments to analyse the ZFY sequences. **Table 9** showed low conservation of the ZFY protein as only 284 (281 within the exon and 3 residues on splice junctions) amino acids in all of the vertebrates were conserved of the total 926 aligned sites. **Table 9** showed that exon 1 contained gap within the alignment sites. Upon observation, we found that exon 2 of rodents had a 3 amino acid gap between sites 47-49, and that fish species had an exon 2 which diverged highly in comparison to land vertebrates as large gaps occurred between sites 113-146, 153-170 and 191-223. Therefore, as stated in the previous section, the fish species were not used for the remainder of the bioinformatic analysis after this section.

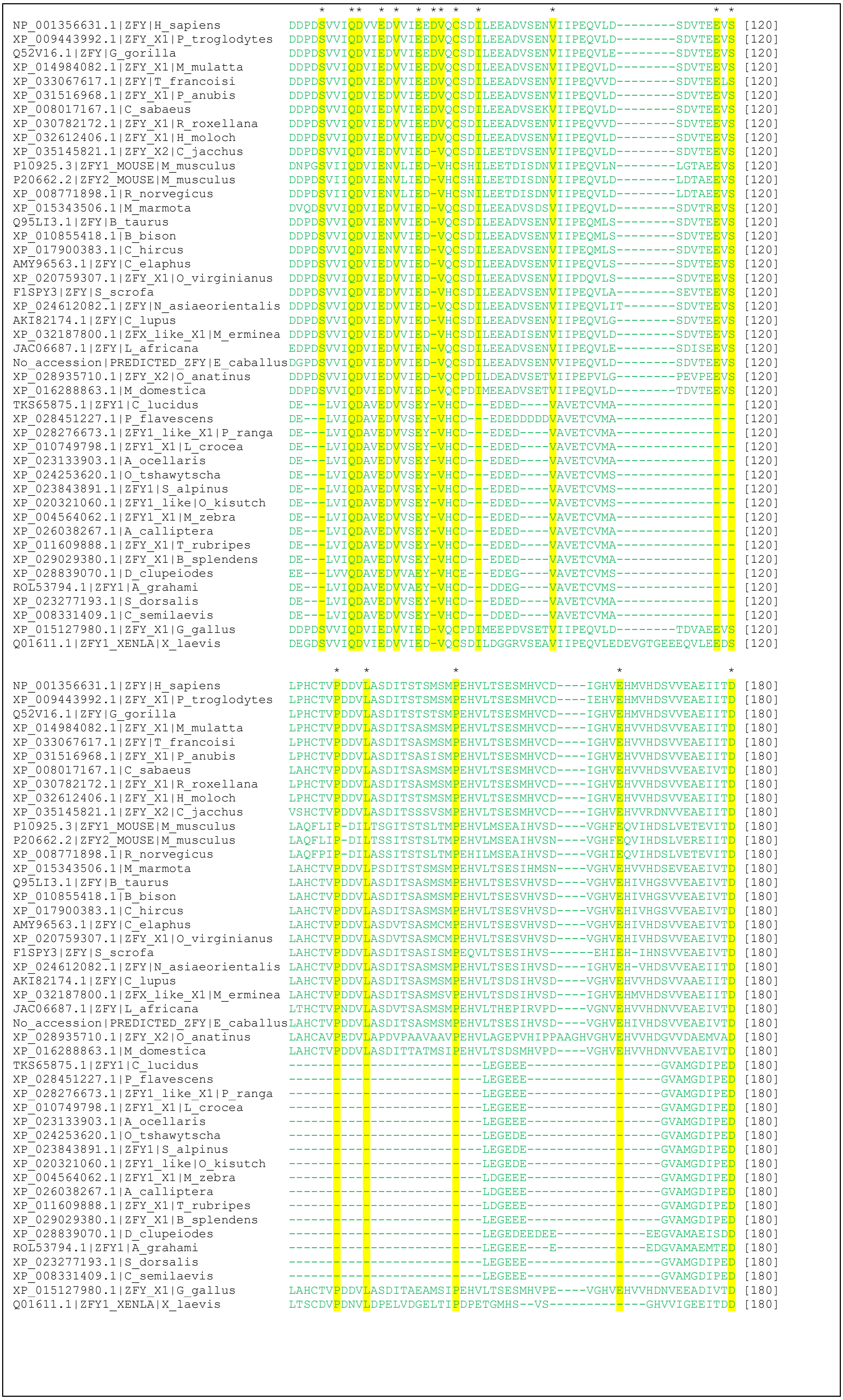
Although the length of exon 3 was highly conserved across the species as indicated by **Table 8** and **Table 9**, the amino acid sequences were not as well conserved across all species as the length as indicated by **Table 10**. Moreover, exon 4 in fish species had a gap between sites 298-302 and exon 5 had a gap in 3 of the 4 rodent sequences between sites 334-342, and also in most fish species between sites 352-359. Exon 6 contained a large stretch of basic residues between sites 449-464, likely to be a nuclear localisation sequence. The acidic activation domain lies between exon 2 and exon 6 and has multiple clusters of conserved amino acids usually 5 or more amino acids. In addition, within the acidic activation domain, most land vertebrates between sites 352-361 contained a stretch of alanine residues, likely to be polyalanine motifs and interestingly, marsupial, monotreme, bird and amphibian did not have these alanine clusters. However, this stretch was not well conserved in rodents, and not available in fish species as they had gaps present or different amino acids in this region, but smaller motifs were present between sites 380-384. This indicated that the stretch/

motif likely has some functional role within the placental land vertebrates. Therefore, within the acidic domain, we located a likely nuclear localisation sequence and a polyalanine motif in the majority of species.

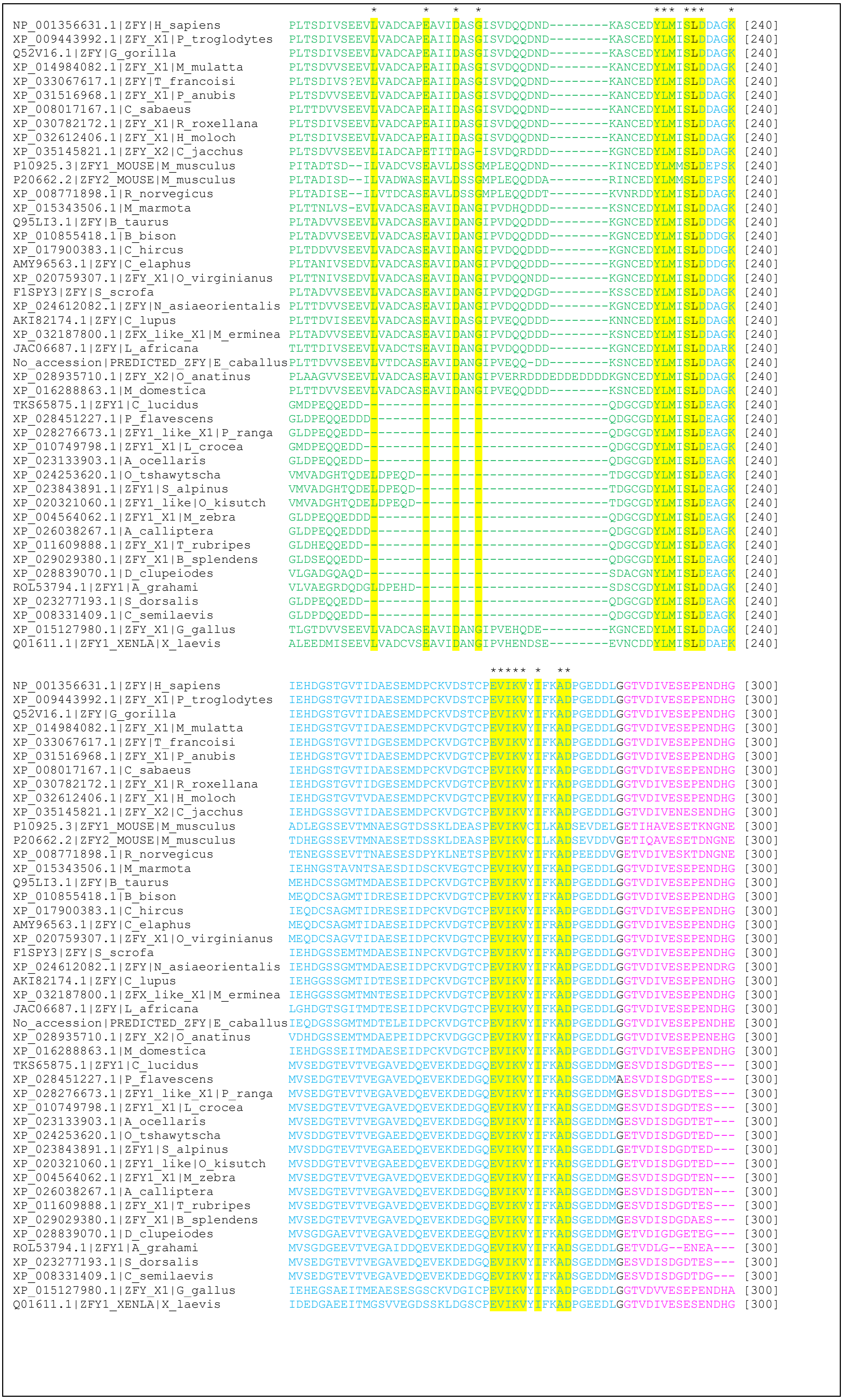
Lastly, exon 7 was the most highly conserved exon of ZFY and is where the zinc fingers of ZFY are located. This was suggestive of exon 7 having very important function in ZFY and would be studied further in sections ahead. In addition, Exon 7 of the fish sequences showed that fish have two inserts not observed in the land vertebrates as they had gaps between sites 584-600 and 809-842, and the inserts are rich in proline, serine, and lysine residues. Though we interpreted multiple sequence alignments to for sequence conservation, it was important we generated phylogenetic trees which allowed us to see closely related species.



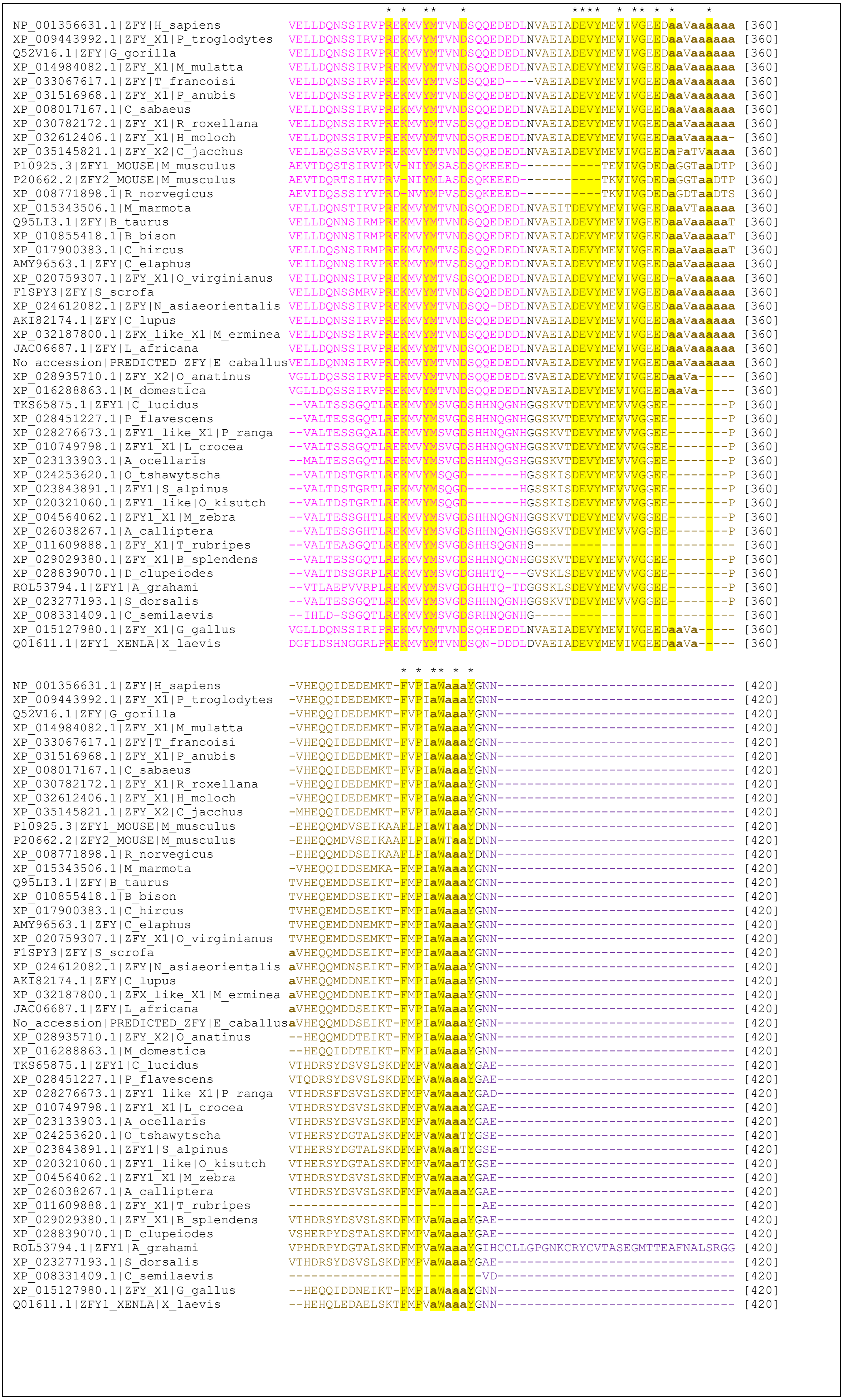
```
                                        *           *        *            ** ** *           *  *
NP_001356631.1|ZFY|H_sapiens          -MDED--EFELQPQEPNSFFDGIGADATHMDGDQIVVEIQEAVFVSNIVDSDITVHNFVP [60]
XP_009443992.1|ZFY_X1|P_troglodytes   -MDED--EFELQPQEPNSFFDGIGADATHMDGDQIVVEIQEAVFVSNIVDSDIAVHNFVP [60]
Q52V16.1|ZFY|G_gorilla                -MDED--EFELQPQEPNSFFDGIGADATHMDGDQIVVEIQEAVFVSNIVDSDITVHNFVP [60]
XP_014984082.1|ZFY_X1|M_mulatta       -MDED--EFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVP [60]
XP_033067617.1|ZFY|T_francoisi        -MDED--EFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVP [60]
XP_031516968.1|ZFY_X1|P_anubis        -MDED--EFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVP [60]
XP_008017167.1|C_sabaeus              -MDED--EFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVP [60]
XP_030782172.1|ZFY_X1|R_roxellana     -MDED--EFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVP [60]
XP_032612406.1|ZFY_X1|H_moloch        -MDED--EFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVP [60]
XP_035145821.1|ZFY_X2|C_jacchus       -MDED--EFELQPQEPNSFFDGIGAGATHMDGDQIVVEVQETVFVSNIVDSDVTVHNFVP [60]
P10925.3|ZFY1_MOUSE|M_musculus        -MDED--EIELTPEEEKSFFDGIGADAVHMDSDQIVVEVQETVFLA---NSDVTVHNFVP [60]
P20662.2|ZFY2_MOUSE|M_musculus        -MDED--EIELTPEEEKSLFDGIGADAVHMDSDQISVEVQETVFLS---NSDVTVHNFVP [60]
XP_008771898.1|R_norvegicus           -MDEE--EIELTPQEENSLFDGIGADAVHMDGDQIIVEVQETVFLS---NSDVTVHNFVP [60]
XP_015343506.1|M_marmota              -MDED--EFELQPQEPNSFFDGIGTDSTHMDGDQIVVEVQETVFVS---NSDITVHNFVP [60]
Q95LI3.1|ZFY|B_taurus                 -MDED--EFELQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVP [60]
XP_010855418.1|B_bison                -MDED--EFELQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVP [60]
XP_017900383.1|C_hircus               -MDED--ELELQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVP [60]
AMY96563.1|ZFY|C_elaphus              -MDED--EFEIQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVP [60]
XP_020759307.1|ZFY_X1|O_virginianus   -MDED--ELEIQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVP [60]
F1SPY3|ZFY|S_scrofa                   -MDED--ELELQPQETNTFFDEIGADDTHMDGDQIVVEVQETVFVSDVVDSDITVHNFVP [60]
XP_024612082.1|ZFY|N_asiaeorientalis  -MDED--ELELQPQEPNSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVP [60]
AKI82174.1|ZFY|C_lupus                -MDED--ELALQPREPNSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVP [60]
XP_032187800.1|ZFX_like_X1|M_erminea  -MDED--ELELQPQEPNSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVP [60]
JAC06687.1|ZFY|L_africana             -MDED--ELELQPQEPNSFFDGIGADVTHMVGDQIVVEVQETVFVSDVVDSDITVHNFPP [60]
No_accession|PREDICTED_ZFY|E_caballus -MDED--ELELRQQEPDSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFAP [60]
XP_028935710.1|ZFY_X2|O_anatinus      -MDED--GLELQPHEPNSFFDATGAAASHMDGGQILVEVQETVFVSDVVDSDITVHNFVP [60]
XP_016288863.1|M_domestica            -MDED--GLELQPQEPNSFFDATGADATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVP [60]
TKS65875.1|ZFY1|C_lucidus             -MDEEVTRLSIRAEEPKIILH--GSDEGGAGGQEFVVELQETVLVSEGEGEGMAVHRFAP [60]
XP_028451227.1|P_flavescens           -MDEDVTRLAIHSEEPKIILH--GSDEGGAGGQEFVVELQETVLVSEGEGEGMAVHRFAP [60]
XP_028276673.1|ZFY1_like_X1|P_ranga   -MDEDVTRLAIHSEEPKIILH--GSDEGGAGGQEFVVELQETVLVSEGEGEGMAVHRFAP [60]
XP_010749798.1|ZFY1_X1|L_crocea       -MDEEVTRLSIRAEEPKIILH--GSDEGGAGGQEFVVELQETVLVSEGEGEGMAVHRFAP [60]
XP_023133903.1|A_ocellaris            -MDEDVTRLAIHSEEPKIILH--GSDEGGAGGQEFVVELQETVLVSEGEGEGMAVHRFAP [60]
XP_024253620.1|O_tshawytscha          -MDEDETRLALHSQEPKIILH--GSDEGGAAGEEFVVELQETVLVSEGEGEGMAVHGFAS [60]
XP_023843891.1|ZFY1|S_alpinus         -MDEDETRLALHSQEPKIILH--GSDEGGAAGEEFVVELQETVLVSEGEGEGMAVHGFAS [60]
XP_020321060.1|ZFY1_like|O_kisutch    -MDEDETRLALHSQEPKIILH--GSDEGGAAGEEFVVELQETVLVSEGEGEGMAVHGFAS [60]
XP_004564062.1|ZFY1_X1|M_zebra        -MDEDVTRLAIHSEEPKIILH--GSDEGGAGGQEFVVELQETVLVSEGEGEGMAVHRFAP [60]
XP_026038267.1|A_calliptera           -MDEDVTRLAIHSEEPKIILH--GSDEGGAGGQEFVVELQETVLVSEGEGEGMAVHRFAP [60]
XP_011609888.1|ZFY_X1|T_rubripes      -MDEDVTRLAIHSEEPKIILH--GSDEGGAGGQEFVVELQETVLVSEGEGESMAVHRFAP [60]
XP_029029380.1|ZFY_X1|B_splendens     -MDEDVTRLAIHSEEPKIILH--GSDEGGAGGQEFVVELQETVLVSEGEGEGMAVHRFAP [60]
XP_028839070.1|D_clupeiodes           -MDEDVARLALHSQEPKIILH--GSDDGGGGEGEFVVELQETVLVSEVDGEGVAVHGFAP [60]
ROL53794.1|ZFY1|A_grahami             -MDEEVTRLALRSQEPKIILH--GSDEGGAGEEEYVVELQETVLVSEVEGEGASVQGFSS [60]
XP_023277193.1|S_dorsalis             -MDEDVTRLAIHSEEPKIILH--GSDEGGAGGQEFVVELQETVLVSEGEGEGMAVHRFAP [60]
XP_008331409.1|C_semilaevis           -MDEDVTRLAIHSEEPKIILH--GSDEGGVGGQEFVVELQETVLVSEGDGEGMAVHRFAP [60]
XP_015127980.1|ZFY_X1|G_gallus        -MDED--GLELQPHEPNAFFDPTGADATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVP [60]
Q01611.1|ZFY1_XENLA|X_laevis          MEDVA--ELELQTTEPHAFFHASGVGERHLNGNEIIVEIQETVFVADG-DGNMAVQGFGP [60]
```



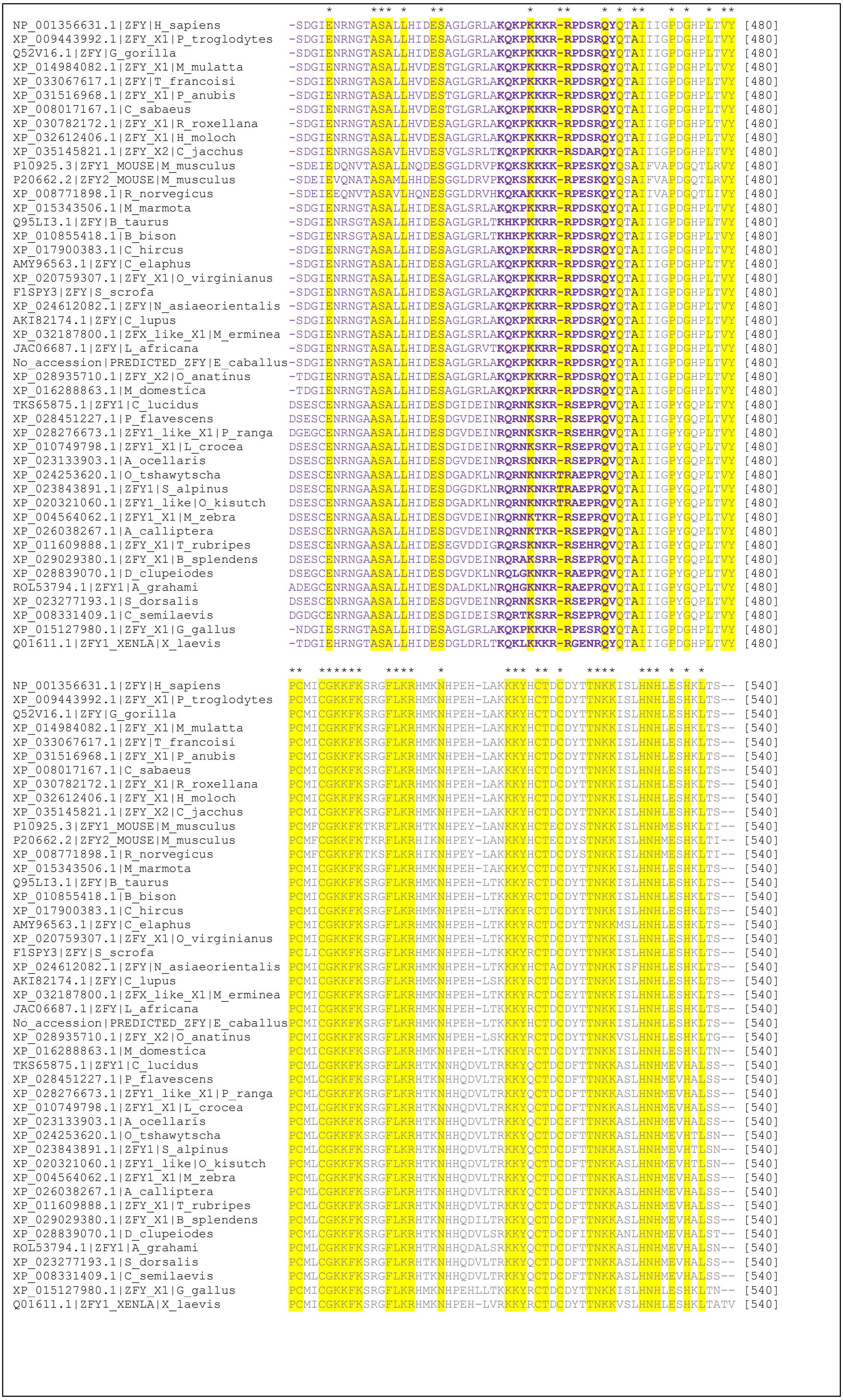
```
                                         *   **  * *  * ** *  *         *                     * *
NP_001356631.1|ZFY|H_sapiens         DDPDSVVIQDVVEDVVIEEDVQCSDILEEADVSENVIIPEQVLD--------SDVTEEVS [120]
XP_009443992.1|ZFY_X1|P_troglodytes  DDPDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD--------SDVTEEVS [120]
Q52V16.1|ZFY|G_gorilla               DDPDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLE--------SDVTEEVS [120]
XP_014984082.1|ZFY_X1|M_mulatta      DDPDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD--------SDVTEEVS [120]
XP_033067617.1|ZFY|T_francoisi       DDPDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVVD--------SDVTEELS [120]
XP_031516968.1|ZFY_X1|P_anubis       DDPDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD--------SDVTEEVS [120]
XP_008017167.1|C_sabaeus             DDPDSVVIQDVIEDVVIEEDVQCSDILEEADVSEKVIIPEQVLD--------SDVTEEVS [120]
XP_030782172.1|ZFY_X1|R_roxellana    DDPDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVVD--------SDVTEEVS [120]
XP_032612406.1|ZFY_X1|H_moloch       DDPDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD--------SDVTEEVS [120]
XP_035145821.1|ZFY_X2|C_jacchus      DDPDSVVIQDVIEDVVIED-VQCSDILEETDVSENVIIPEQVLD--------SDVTEEVS [120]
P10925.3|ZFY1_MOUSE|M_musculus       DNPGSVIIQDVIENVLIED-VHCSHILEETDISDNVIIPEQVLN--------LGTAEEVS [120]
P20662.2|ZFY2_MOUSE|M_musculus       DDPDSVIIQDVIENVLIED-VHCSHILEETDISDNVIIPEQVLD--------LDTAEEVS [120]
XP_008771898.1|R_norvegicus          DDPDSVIIQDVIENVLIED-VHCSNILEETDISDNVIIPEQVLD--------LDTAEEVS [120]
XP_015343506.1|M_marmota             DVQDSVVIQDVIEDVVIED-VQCSDILEEADVSDSVIIPEQVLD--------SDVTREVS [120]
Q95LI3.1|ZFY|B_taurus                DDPDSVVIQDVIENVVIED-VQCSDILEEADVSENVIIPEQMLS--------SDVTEEVS [120]
XP_010855418.1|B_bison               DDPDSVVIQDVIENVVIED-VQCSDILEEADVSENVIIPEQMLS--------SDVTEEVS [120]
XP_017900383.1|C_hircus              DDPDSVVIQDVIENVVIED-VQCSDILEEADVSENVIIPEQMLS--------SDVTEEVS [120]
AMY96563.1|ZFY|C_elaphus             DDPDSVVIQDVIEDVVIED-VQCSDILEEADVSENVIIPEQVLS--------SDVTEEVS [120]
XP_020759307.1|ZFY_X1|O_virginianus  DDPDSVVIQDVIEDVVIED-VQCSDILEEADVSENVIIPDQVLS--------SDVTEEVS [120]
F1SPY3|ZFY|S_scrofa                  DDPDSVVIQDVIEDVVIED-VHCSDILEEADVSENVIIPEQVLA--------SEVTEEVS [120]
XP_024612082.1|ZFY|N_asiaeorientalis DDPDSVVIQDVIEDVVIED-VQCSDILEEADVSENVIIPEQVLIT-------SDVTEEVS [120]
AKI82174.1|ZFY|C_lupus               DDPDSVVIQDVIEDVVIED-VHCSDILEEADVSENVIIPEQVLG--------SDVTEEVS [120]
XP_032187800.1|ZFX_like_X1|M_erminea DDPDSVVIQDVIEDVVIED-VHCSDILEEADISENVIIPEQVLD--------SDVTEEVS [120]
JAC06687.1|ZFY|L_africana            EDPDSVVIQDVIEDVVIEN-VQCSDILEEADVSENVIIPEQVLE--------SDISEEVS [120]
No_accession|PREDICTED_ZFY|E_caballusDGPDSVVIQDVIEDVVIED-VQCSDILEEADVSENVVIPEQVLD--------SDVTEEVS [120]
XP_028935710.1|ZFY_X2|O_anatinus     DDPDSVVIQDVIEDVVIED-VQCPDILDEADVSETVIIPEPVLG--------PEVPEEVS [120]
XP_016288863.1|M_domestica           DDPDSVVIQDVIEDVVIED-VQCPDIMEEADVSETVIIPEQVLD--------TDVTEEVS [120]
TKS65875.1|ZFY1|C_lucidus            DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMA---------------- [120]
XP_028451227.1|P_flavescens          DE---LVIQDAVEDVVSEY-VHCD---EDEDDDDVAVETCVMA---------------- [120]
XP_028276673.1|ZFY1_like_X1|P_ranga  DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMA---------------- [120]
XP_010749798.1|ZFY1_X1|L_crocea      DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMA---------------- [120]
XP_023133903.1|A_ocellaris           DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMA---------------- [120]
XP_024253620.1|O_tshawytscha         DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMS---------------- [120]
XP_023843891.1|ZFY1|S_alpinus        DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMS---------------- [120]
XP_020321060.1|ZFY1_like|O_kisutch   DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMS---------------- [120]
XP_004564062.1|ZFY1_X1|M_zebra       DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMA---------------- [120]
XP_026038267.1|A_calliptera          DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMA---------------- [120]
XP_011609888.1|ZFY_X1|T_rubripes     DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMA---------------- [120]
XP_029029380.1|ZFY_X1|B_splendens    DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMA---------------- [120]
XP_028839070.1|D_clupeiodes          EE---LVVQDAVEDVVAEY-VHCE---EDEG----VAVETCVMS---------------- [120]
ROL53794.1|ZFY1|A_grahami            DE---LVIQDAVEDVVAEY-VHCD---DDEG----VAVETCVMS---------------- [120]
XP_023277193.1|S_dorsalis            DE---LVIQDAVEDVVSEY-VHCD---EDED----VAVETCVMA---------------- [120]
XP_008331409.1|C_semilaevis          DE---LVIQDAVEDVVSEY-VHCD---DDED----VAVETCVMA---------------- [120]
XP_015127980.1|ZFY_X1|G_gallus       DDPDSVVIQDVIEDVVIED-VQCPDIMEEPDVSETVIIPEQVLD--------TDVAEEVS [120]
Q01611.1|ZFY1_XENLA|X_laevis         DEGDSVVIQDVIEDVVIED-VQCSDILDGGRVSEAVIIPEQVLEDEVGTGEEEQVLEEDS [120]

                                           *   *           *                     *              *
NP_001356631.1|ZFY|H_sapiens         LPHCTVPDDVLASDITSTSMSMPEHVLTSESMHVCD----IGHVEHMVHDSVVEAEIITD [180]
XP_009443992.1|ZFY_X1|P_troglodytes  LPHCTVPDDVLASDITSTSMSMPEHVLTSESMHVCD----IEHVEHMVHDSVVEAEIITD [180]
Q52V16.1|ZFY|G_gorilla               LPHCTVPDDVLASDITSTSTSMPEHVLTSESMHVCD----IGHVEHMVHDSVVEAEIITD [180]
XP_014984082.1|ZFY_X1|M_mulatta      LPHCTVPDDVLASDITSASMSMPEHVLTSESMHVCD----IGHVEHVVHDSVVEAEIITD [180]
XP_033067617.1|ZFY|T_francoisi       LPHCTVPDDVLASDITSASMSMPEHVLTSESMHVCD----IGHVEHVVHDSVVEAEIITD [180]
XP_031516968.1|ZFY_X1|P_anubis       LPHCTVPDDVLASDITSASISMPEHVLTSESMHVCD----IGHVEHVVHDSVVEAEIITD [180]
XP_008017167.1|C_sabaeus             LAHCTVPDDVLASDITSASMSMPEHVLTSESMHVCD----IGHVEHVVHDSVVEAEIVTD [180]
XP_030782172.1|ZFY_X1|R_roxellana    LPHCTVPDDVLASDITSASMSMPEHVLTSESMHVCD----IGHVEHVVHDSVVEAEIITD [180]
XP_032612406.1|ZFY_X1|H_moloch       LPHCTVPDDVLASDITSTSMSMPEHVLTSESMHVCD----IGHVEHVVHDSVVEAEIITD [180]
XP_035145821.1|ZFY_X2|C_jacchus      VSHCTVPDDVLASDITSSSVSMPEHVLTSESMHVCD----IGHVEHVVRDNVVEAEIITD [180]
P10925.3|ZFY1_MOUSE|M_musculus       LAQFLIP-DILTSGITSTSLTMPEHVLMSEAIHVSD----VGHFEQVIHDSLVETEVITD [180]
P20662.2|ZFY2_MOUSE|M_musculus       LAQFLIP-DILTSSITSTSLTMPEHVLMSEAIHVSN----VGHFEQVIHDSLVEREIITD [180]
XP_008771898.1|R_norvegicus          LAQFPIP-DILASSITSTSLTMPEHILMSEAIHVSD----VGHIEQVIHDSLVETEVITD [180]
XP_015343506.1|M_marmota             LAHCTVPDDVLPSDITSTSMSMPEHVLTSESIHMSN----VGHVEHVVHDSEVEAEIVTD [180]
Q95LI3.1|ZFY|B_taurus                LAHCTVPDDVLASDITSASMSMPEHVLTSESVHVSD----VGHVEHIVHGSVVEAEIVTD [180]
XP_010855418.1|B_bison               LAHCTVPDDVLASDITSASMSMPEHVLTSESVHVSD----VGHVEHIVHGSVVEAEIVTD [180]
XP_017900383.1|C_hircus              LAHCTVPDDVLASDITSASMSMPEHVLTSESVHVSD----VGHVEHIVHGSVVEAEIVTD [180]
AMY96563.1|ZFY|C_elaphus             LAHCTVPDDVLASDVTSASMCMPEHVLTSESVHVSD----VGHVEHIVHDSVVEAEIVTD [180]
XP_020759307.1|ZFY_X1|O_virginianus  LAHCTVPDDVLASDVTSASMCMPEHVLTSESVHVSD----VGHVEHIVHDSVVEAEIVTD [180]
F1SPY3|ZFY|S_scrofa                  LAHCTVPDDVLASDITSASISMPEQVLTSESIHVS------EHIEH-IHNSVVEAEIVTD [180]
XP_024612082.1|ZFY|N_asiaeorientalis LAHCTVPDDVLASDITSASMSMPEHVLTSESIHVSD----IGHVEH-VHDSVVEAEIITD [180]
AKI82174.1|ZFY|C_lupus               LAHCTVPDDVLASDITSASMSMPEHVLTSDSIHVSD----VGHVEHVVHDSVVAAEIITD [180]
XP_032187800.1|ZFX_like_X1|M_erminea LAHCTVPDDVLASDITSASMSVPEHVLTSDSIHVSD----IGHVEHMVHDSVVEAEIITD [180]
JAC06687.1|ZFY|L_africana            LTHCTVPNDVLASDVTSASMSMPEHVLTHEPIRVPD----VGNVEHVVHDNVVEAEIVTD [180]
No_accession|PREDICTED_ZFY|E_caballusLAHCTVPDDVLASDITSASMSMPEHVLTSESIHVSD----VGHVEHIVHDSVVEAEIVTD [180]
XP_028935710.1|ZFY_X2|O_anatinus     LAHCAVPEDVLAPDVPAAVAAVPEHVLAGEPVHIPPAAGHVGHVEHVVHDGVVDAEMVAD [180]
XP_016288863.1|M_domestica           LAHCTVPDDVLASDITTATMSIPEHVLTSDSMHVPD----VGHVEHVVHDNVVEAEIVTD [180]
TKS65875.1|ZFY1|C_lucidus            --------------------------LEGEEE------------------GVAMGDIPED [180]
XP_028451227.1|P_flavescens          --------------------------LEGEEE------------------GVAMGDIPED [180]
XP_028276673.1|ZFY1_like_X1|P_ranga  --------------------------LEGEEE------------------GVAMGDIPED [180]
XP_010749798.1|ZFY1_X1|L_crocea      --------------------------LEGEEE------------------GVAMGDIPED [180]
XP_023133903.1|A_ocellaris           --------------------------LEGEEE------------------GVAMGDIPED [180]
XP_024253620.1|O_tshawytscha         --------------------------LEGEDE------------------GVAMGDIPED [180]
XP_023843891.1|ZFY1|S_alpinus        --------------------------LEGEDE------------------GVAMGDIPED [180]
XP_020321060.1|ZFY1_like|O_kisutch   --------------------------LEGEDE------------------GVAMGDIPED [180]
XP_004564062.1|ZFY1_X1|M_zebra       --------------------------LDGEEE------------------GVAMGDIPED [180]
XP_026038267.1|A_calliptera          --------------------------LDGEEE------------------GVAMGDIPED [180]
XP_011609888.1|ZFY_X1|T_rubripes     --------------------------LDGEEE------------------GVAMGDIPED [180]
XP_029029380.1|ZFY_X1|B_splendens    --------------------------LEGEEE------------------GVAMGDIPED [180]
XP_028839070.1|D_clupeiodes          --------------------------LEGEDEEDEE------------EEGVAMAEISDD [180]
ROL53794.1|ZFY1|A_grahami            --------------------------LEGEEE---E------------EDGVAMAEMTED [180]
XP_023277193.1|S_dorsalis            --------------------------LEGEEE------------------GVAMGDIPED [180]
XP_008331409.1|C_semilaevis          --------------------------LEGEEE------------------GVAMGDIPED [180]
XP_015127980.1|ZFY_X1|G_gallus       LAHCTVPDDVLASDITAEAMSIPEHVLTSESMHVPE----VGHVEHVVHDNVEEADIVTD [180]
Q01611.1|ZFY1_XENLA|X_laevis         LTSCDVPDNVLDPELVDGELTIPDPETGMHS--VS-------------GHVVIGEEITDD [180]
```



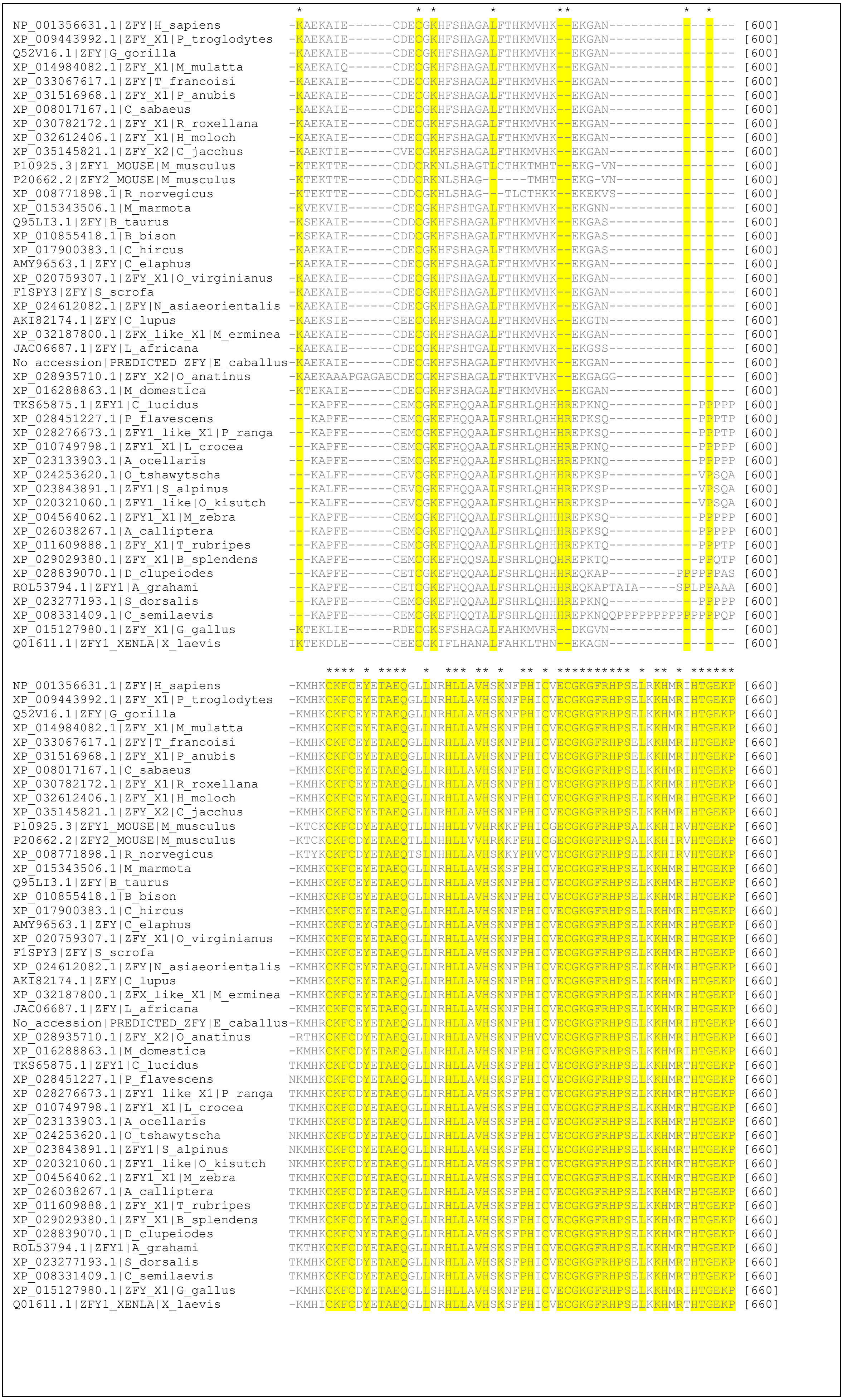
```
                                                *      *   *  *                       *** ***   *
NP_001356631.1|ZFY|H_sapiens         PLTSDIVSEEVLVADCAPEAIIDASGISVDQQDND--------KASCEDYLMISLDDAGK [240]
XP_009443992.1|ZFY_X1|P_troglodytes  PLTSDIVSEEVLVADCAPEAIIDASGISVDQQDND--------KASCEDYLMISLDDAGK [240]
Q52V16.1|ZFY|G_gorilla               PLTSDIVSEEVLVADCAPEAIIDASGISVDQQDND--------KASCEDYLMISLDDAGK [240]
XP_014984082.1|ZFY_X1|M_mulatta      PLTSDVVSEEVLVADCAPEAIIDASGISVDQQDND--------KANCEDYLMISLDDAGK [240]
XP_033067617.1|ZFY|T_francoisi       PLTSDIVS?EVLVADCAPEAIIDASGISVDQQDND--------KANCEDYLMISLDDAGK [240]
XP_031516968.1|ZFY_X1|P_anubis       PLTSDVVSEEVLVADCAPEAIIDASGISVDQQDND--------KANCEDYLMISLDDAGK [240]
XP_008017167.1|C_sabaeus             PLTTDVVSEEVLVADCAPEAIIDASGISVDQQDND--------KANCEDYLMISLDDAGK [240]
XP_030782172.1|ZFY_X1|R_roxellana    PLTSDIVSEEVLVADCAPEAIIDASGISVDQQDND--------KANCEDYLMISLDDAGK [240]
XP_032612406.1|ZFY_X1|H_moloch       PLTSDIVSEEVLVADCAPEAIIDASGISVDQQDND--------KANCEDYLMISLDDAGK [240]
XP_035145821.1|ZFY_X2|C_jacchus      PLTSDVVSEEVLIADCAPETITDAG-ISVDQRDDD--------KGNCEDYLMISLDDAGK [240]
P10925.3|ZFY1_MOUSE|M_musculus       PITADTSD--ILVADCVSEAVLDSSGMPLEQQDND--------KINCEDYLMMSLDEPSK [240]
P20662.2|ZFY2_MOUSE|M_musculus       PLTADISD--ILVADWASEAVLDSSGMPLEQQDDA--------RINCEDYLMMSLDEPSK [240]
XP_008771898.1|R_norvegicus          PLTADISE--ILVTDCASEAVLDSSGMPLEQQDDT--------KVNRDDYLMISLDDAGK [240]
XP_015343506.1|M_marmota             PLTTNLVS-EVLVADCASEAVIDANGIPVDHQDDD--------KSNCEDYLMISLDDAGK [240]
Q95LI3.1|ZFY|B_taurus                PLTADVVSEEVLVADCASEAVIDANGIPVDQQDDD--------KGNCEDYLMISLDDDGK [240]
XP_010855418.1|B_bison               PLTADVVSEEVLVADCASEAVIDANGIPVDQQDDD--------KGNCEDYLMISLDDDGK [240]
XP_017900383.1|C_hircus              PLTDDVVSEEVLVADCASEAVIDANGIPVDQQDDD--------KGNCEDYLMISLDDDGK [240]
AMY96563.1|ZFY|C_elaphus             PLTANIVSEDVLVADCASEAVIDANGIPVDQQDDD--------KGNCEDYLMISLDDDGK [240]
XP_020759307.1|ZFY_X1|O_virginianus  PLTTNIVSEDVLVADCASEAVIDANGIPVDQQNDD--------KGNCEDYLMISLDDDGK [240]
F1SPY3|ZFY|S_scrofa                  PLTADVVSEEVLVADCASEAVIDANGIPVDQQDGD--------KSSCEDYLMISLDDAGK [240]
XP_024612082.1|ZFY|N_asiaeorientalis PLTTDVVSEEVLVADCASEAVIDANGIPVDQQDDD--------KGNCEDYLMISLDDAGK [240]
AKI82174.1|ZFY|C_lupus               PLTTDVISEEVLVADCASEAVIDASGIPVEQQDDD--------KNNCEDYLMISLDDAGK [240]
XP_032187800.1|ZFX_like_X1|M_erminea PLTADVVSEEVLVADCASEAVIDANGIPVDQQDDD--------KSNCEDYLMISLDDAGK [240]
JAC06687.1|ZFY|L_africana            TLTTDIVSEEVLVADCTSEAVIDANGIPVDQQDDD--------KGNCEDYLMISLDDARK [240]
No_accession|PREDICTED_ZFY|E_caballusPLTTDVVSEEVLVTDCASEAVIDANGIPVEQQ-DD--------KSNCEDYLMISLDDAGK [240]
XP_028935710.1|ZFY_X2|O_anatinus     PLAAGVVSEEVLVADCASEAVIDANGIPVERRDDDEDDEDDDDKGNCEDYLMISLDDAGK [240]
XP_016288863.1|M_domestica           PLTTDVVSEEVLVADCASEAVIDANGIPVEQQDDD--------KSNCEDYLMISLDDAGK [240]
TKS65875.1|ZFY1|C_lucidus            GMDPEQQEDD---------------------------------QDGCGDYLMISLDEAGK [240]
XP_028451227.1|P_flavescens          GLDPEQQEDDD--------------------------------QDGCGDYLMISLDEAGK [240]
XP_028276673.1|ZFY1_like_X1|P_ranga  GLDPEQQEDD---------------------------------QDGCGDYLMISLDEAGK [240]
XP_010749798.1|ZFY1_X1|L_crocea      GMDPEQQEDD---------------------------------QDGCGDYLMISLDEAGK [240]
XP_023133903.1|A_ocellaris           GLDPEQQEDD---------------------------------QDGCGDYLMISLDEAGK [240]
XP_024253620.1|O_tshawytscha         VMVADGHTQDELDPEQD--------------------------TDGCGDYLMISLDEAGK [240]
XP_023843891.1|ZFY1|S_alpinus        VMVADGHTQDELDPEQD--------------------------TDGCGDYLMISLDEAGK [240]
XP_020321060.1|ZFY1_like|O_kisutch   VMVADGHTQDELDPEQD--------------------------TDGCGDYLMISLDEAGK [240]
XP_004564062.1|ZFY1_X1|M_zebra       GLDPEQQEDDD--------------------------------QDGCGDYLMISLDEAGK [240]
XP_026038267.1|A_calliptera          GLDPEQQEDDD--------------------------------QDGCGDYLMISLDEAGK [240]
XP_011609888.1|ZFY_X1|T_rubripes     GLDHEQQEDD---------------------------------QDGCGDYLMISLDEAGK [240]
XP_029029380.1|ZFY_X1|B_splendens    GLDSEQQEDD---------------------------------QDGCGDYLMISLDEAGK [240]
XP_028839070.1|D_clupeiodes          VLGADGQAQD---------------------------------SDACGNYLMISLDEAGK [240]
ROL53794.1|ZFY1|A_grahami            VLVAEGRDQDGLDPEHD--------------------------SDSCGDYLMISLDDAGK [240]
XP_023277193.1|S_dorsalis            GLDPEQQEDD---------------------------------QDGCGDYLMISLDEAGK [240]
XP_008331409.1|C_semilaevis          GLDPDQQEDD---------------------------------QDGCGDYLMISLDEAGK [240]
XP_015127980.1|ZFY_X1|G_gallus       TLGTDVVSEEVLVADCASEAVIDANGIPVEHQDE---------KGNCEDYLMISLDDAGK [240]
Q01611.1|ZFY1_XENLA|X_laevis         ALEEDMISEEVLVADCVSEAVIDANGIPVHENDSE--------EVNCDDYLMISLDDAEK [240]

                                                                ***** *  **
NP_001356631.1|ZFY|H_sapiens         IEHDGSTGVTIDAESEMDPCKVDSTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_009443992.1|ZFY_X1|P_troglodytes  IEHDGSTGVTIDAESEMDPCKVDSTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
Q52V16.1|ZFY|G_gorilla               IEHDGSTGVTIDAESEMDPCKVDSTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_014984082.1|ZFY_X1|M_mulatta      IEHDGSTGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_033067617.1|ZFY|T_francoisi       IEHDGSTGVTIDGESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_031516968.1|ZFY_X1|P_anubis       IEHDGSTGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_008017167.1|C_sabaeus             IEHDGSTGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_030782172.1|ZFY_X1|R_roxellana    IEHDGSTGVTIDGESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_032612406.1|ZFY_X1|H_moloch       IEHDGSTGVTVDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_035145821.1|ZFY_X2|C_jacchus      IEHDGSSGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVENESENDHG [300]
P10925.3|ZFY1_MOUSE|M_musculus       ADLEGSSEVTMNAESGTDSSKLDEASPEVIKVCILKADSEVDELGETIHAVESETKNGNE [300]
P20662.2|ZFY2_MOUSE|M_musculus       TDHEGSSEVTMNAESETDSSKLDEASPEVIKVCILKADSEVDDVGETIQAVESETDNGNE [300]
XP_008771898.1|R_norvegicus          TENEGSSEVTTNAESESDPYKLNETSPEVIKVYIFKADPEEDDVGETVDIVESKTDNGNE [300]
XP_015343506.1|M_marmota             IEHNGSTAVNTSAESDIDSCKVEGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
Q95LI3.1|ZFY|B_taurus                MEHDCSSGMTMDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_010855418.1|B_bison               MEQDCSAGMTIDRESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_017900383.1|C_hircus              IEQDCSAGMTIDRESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
AMY96563.1|ZFY|C_elaphus             MEQDCSAGMTIDAESEIDPCKVDGTCPEVIKVYIFRADPGEDDLGGTVDIVESEPENDHG [300]
XP_020759307.1|ZFY_X1|O_virginianus  MEQDCSAGVTIDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
F1SPY3|ZFY|S_scrofa                  IEHDGSSEMTMDAESEINPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_024612082.1|ZFY|N_asiaeorientalis IEHDGSSGMTMDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDRG [300]
AKI82174.1|ZFY|C_lupus               IEHGGSSGMTIDTESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
XP_032187800.1|ZFX_like_X1|M_erminea IEHGGSSGMTMNTESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
JAC06687.1|ZFY|L_africana            LGHDGTSGITMDTESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
No_accession|PREDICTED_ZFY|E_caballusIEQDGSSGMTMDTELEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHE [300]
XP_028935710.1|ZFY_X2|O_anatinus     VDHDGSSEMTMDAEPEIDPCKVDGGCPEVIKVYIFKADPGEDDLGGTVDIVESEPENEHG [300]
XP_016288863.1|M_domestica           IEHDGSSEITMDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHG [300]
TKS65875.1|ZFY1|C_lucidus            MVSEDGTEVTVEGAVEDQEVEKDEDGQEVIKVYIFKADSGEDDMGESVDISDGDTES--- [300]
XP_028451227.1|P_flavescens          MVSEDGTEVTVEGAVEDQEVEKDEDGQEVIKVYIFKADSGEDDMAESVDISDGDTES--- [300]
XP_028276673.1|ZFY1_like_X1|P_ranga  MVSEDGTEVTVEGAVEDQEVEKDEDGQEVIKVYIFKADSGEDDMGESVDISDGDTES--- [300]
XP_010749798.1|ZFY1_X1|L_crocea      MVSEDGTEVTVEGAVEDQEVEKDEDGQEVIKVYIFKADSGEDDMGESVDISDGDTES--- [300]
XP_023133903.1|A_ocellaris           MVSEDGTEVTVEGAVEDQEVEKDEDGQEVIKVYIFKADSGEDDMGESVDISDGDTET--- [300]
XP_024253620.1|O_tshawytscha         MVSDDGTEVTVEGAEEDQEVEKDEDGQEVIKVYIFKADSGEDDLGETVDISDGDTED--- [300]
XP_023843891.1|ZFY1|S_alpinus        MVSDDGTEVTVEGAEEDQEVEKDEDGQEVIKVYIFKADSGEDDLGETVDISDGDTED--- [300]
XP_020321060.1|ZFY1_like|O_kisutch   MVSDDGTEVTVEGAEEDQEVEKDEDGQEVIKVYIFKADSGEDDLGETVDISDGDTED--- [300]
XP_004564062.1|ZFY1_X1|M_zebra       MVSEDGTEVTVEGAVEDQEVEKDEDGQEVIKVYIFKADSGEDDMGESVDISDGDTEN--- [300]
XP_026038267.1|A_calliptera          MVSEDGTEVTVEGAVEDQEVEKDEDGQEVIKVYIFKADSGEDDMGESVDISDGDTEN--- [300]
XP_011609888.1|ZFY_X1|T_rubripes     MVSEDGTEVTVEGAVEDQEVEKDEDGQEVIKVYIFKADSGEDDMGESVDISDGDTES--- [300]
XP_029029380.1|ZFY_X1|B_splendens    MVSEDGTEVTVEGAVEDQEVEKDEDGQEVIKVYIFKADSGEDDMGESVDISDGDAES--- [300]
XP_028839070.1|D_clupeiodes          MVSGDGAEVTVEGAVEDQEVEKDEEGQEVIKVYIFKADSGEDDMGETVDIGDGETEG--- [300]
ROL53794.1|ZFY1|A_grahami            MVSGDGEEVTVEGAIDDQEVEKDEDGQEVIKVYIFKADSGEDDLGETVDLG--ENEA--- [300]
XP_023277193.1|S_dorsalis            MVSEDGTEVTVEGAVEDQEVEKDEDGQEVIKVYIFKADSGEDDMGESVDISDGDTES--- [300]
XP_008331409.1|C_semilaevis          MVSEDGTEVTVEGAVEDQEVEKDEDGQEVIKVYIFKADSGEDDMGESVDISDGDTDG--- [300]
XP_015127980.1|ZFY_X1|G_gallus       IEHEGSAEITMEAESESGSCKVDGICPEVIKVYIFKADPGEDDLGGTVDVVESEPENDHA [300]
Q01611.1|ZFY1_XENLA|X_laevis         IDEDGAEEITMGSVVEGDSSKLDGSCPEVIKVYIFKADPGEEDLGGTVDIVESESENDHG [300]
```



```
                                                  * *  **   *              ****  * ** * *    *
NP_001356631.1|ZFY|H_sapiens         VELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
XP_009443992.1|ZFY_X1|P_troglodytes  VELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
Q52V16.1|ZFY|G_gorilla               VELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
XP_014984082.1|ZFY_X1|M_mulatta      VELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
XP_033067617.1|ZFY|T_francoisi       VELLDQNSSIRVPREKMVYMTVSDSQQED----VAEIADEVYMEVIVGEEDaaVaaaaaa [360]
XP_031516968.1|ZFY_X1|P_anubis       VELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
XP_008017167.1|C_sabaeus             VELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
XP_030782172.1|ZFY_X1|R_roxellana    VELLDQNSSIRVPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
XP_032612406.1|ZFY_X1|H_moloch       VELLDQNSSIRVPREKMVYMTVNDSQREDEDLNVAEIADEVYMEVIVGEEDaaVaaaaa- [360]
XP_035145821.1|ZFY_X2|C_jacchus      VELLEQSSSVRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaPaTVaaaa [360]
P10925.3|ZFY1_MOUSE|M_musculus       AEVTDQSTSIRVPRV-NIYMSASDSQKEEED-----------TEVIVGDEDaGGTaaDTP [360]
P20662.2|ZFY2_MOUSE|M_musculus       AEVTDQRTSIHVPRV-NIYMLASDSQKEEED-----------TKVIVGDEDaGGTaaDTP [360]
XP_008771898.1|R_norvegicus          AEVIDQSSSIYVPRD-NVYMPVSDSQKEEED-----------TKVIVGDEDaGDTaaDTS [360]
XP_015343506.1|M_marmota             VELLDQNSTIRVPREKMVYMTVNDSQQEDEDLNVAEITDEVYMEVIVGEEDaaVTaaaaa [360]
Q95LI3.1|ZFY|B_taurus                VELLDQNNSIRMPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaT [360]
XP_010855418.1|B_bison               VELLDQNNSIRMPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaT [360]
XP_017900383.1|C_hircus              VELLDQSNSIRMPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaT [360]
AMY96563.1|ZFY|C_elaphus             VEILDQNNSIRVPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
XP_020759307.1|ZFY_X1|O_virginianus  VEILDQNNSIRVPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEED-aVaaaaaa [360]
F1SPY3|ZFY|S_scrofa                  VELLDQNSSMRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
XP_024612082.1|ZFY|N_asiaeorientalis VELLDQNSSIRVPREKMVYMTVNDSQQ-DEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
AKI82174.1|ZFY|C_lupus               VELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
XP_032187800.1|ZFX_like_X1|M_erminea VELLDQNSSIRVPREKMVYMTVNDSQQEDDDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
JAC06687.1|ZFY|L_africana            VELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
No_accession|PREDICTED_ZFY|E_caballusVELLDQNNSIRVPRDKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa [360]
XP_028935710.1|ZFY_X2|O_anatinus     VGLLDQSSSIRVPREKMVYMTVNDSQQEDEDLSVAEIADEVYMEVIVGEEDaaVa----- [360]
XP_016288863.1|M_domestica           VGLLDQSSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVa----- [360]
TKS65875.1|ZFY1|C_lucidus            --VALTSSSGQTLREKMVYMSVGDSHHNQGNHGGSKVTDEVYMEVVVGGEE--------P [360]
XP_028451227.1|P_flavescens          --VALTESSGQTLREKMVYMSVGDSHHNQGNHGGSKVTDEVYMEVVVGGEE--------P [360]
XP_028276673.1|ZFY1_like_X1|P_ranga  --VALTESSGQALREKMVYMSVGDSHHNQGNHGGSKVTDEVYMEVVVGGEE--------P [360]
XP_010749798.1|ZFY1_X1|L_crocea      --VALTSSSGQTLREKMVYMSVGDSHHNQGNHGGSKVTDEVYMEVVVGGEE--------P [360]
XP_023133903.1|A_ocellaris           --MALTESSGQTLREKMVYMSVGDSHHNQGSHGGSKVTDEVYMEVVVGGEE--------P [360]
XP_024253620.1|O_tshawytscha         --VALTDSTGRTLREKMVYMSQGD-------HGSSKISDEVYMEVVVGGEE--------P [360]
XP_023843891.1|ZFY1|S_alpinus        --VALTDSTGRTLREKMVYMSQGD-------HGSSKISDEVYMEVVVGGEE--------P [360]
XP_020321060.1|ZFY1_like|O_kisutch   --VALTDSTGRTLREKMVYMSQGD-------HGSSKISDEVYMEVVVGGEE--------P [360]
XP_004564062.1|ZFY1_X1|M_zebra       --VALTESSGHTLREKMVYMSVGDSHHNQGNHGGSKVTDEVYMEVVVGGEE--------P [360]
XP_026038267.1|A_calliptera          --VALTESSGHTLREKMVYMSVGDSHHNQGNHGGSKVTDEVYMEVVVGGEE--------P [360]
XP_011609888.1|ZFY_X1|T_rubripes     --VALTEASGQTLREKMVYMSVGDSHHSQGNHS--------------------------- [360]
XP_029029380.1|ZFY_X1|B_splendens    --VALTESSGQTLREKMVYMSVGDSHHNQGNHGGSKVTDEVYMEVVVGGEE--------P [360]
XP_028839070.1|D_clupeiodes          --VALTDSSGRPLREKMVYMSVGDGHHTQ---GVSKLSDEVYMEVVVGGEE--------P [360]
ROL53794.1|ZFY1|A_grahami            --VTLAEPVVRPLREKMVYMSVGDGHHTQ-TDGGSKLSDEVYMEVVVGGEE--------P [360]
XP_023277193.1|S_dorsalis            --VALTESSGQTLREKMVYMSVGDSHHNQGNHGGSKVTDEVYMEVVVGGEE--------P [360]
XP_008331409.1|C_semilaevis          --IHLD-SSGQTLREKMVYMSVGDSRHNQGNHG--------------------------- [360]
XP_015127980.1|ZFY_X1|G_gallus       VGLLDQNSSIRIPREKMVYMTVNDSQHEDEDLNVAEIADEVYMEVIVGEEDaaVa----- [360]
Q01611.1|ZFY1_XENLA|X_laevis         DGFLDSHNGGRLPREKMVYMTVNDSQN-DDDLDVAEIADEVYMEVIVGEEDaaVa----- [360]

                                                    * * ** * *
NP_001356631.1|ZFY|H_sapiens         -VHEQQIDEDEMKT-FVPIaWaaaYGNN-------------------------------- [420]
XP_009443992.1|ZFY_X1|P_troglodytes  -VHEQQIDEDEMKT-FVPIaWaaaYGNN-------------------------------- [420]
Q52V16.1|ZFY|G_gorilla               -VHEQQIDEDEMKT-FVPIaWaaaYGNN-------------------------------- [420]
XP_014984082.1|ZFY_X1|M_mulatta      -VHEQQIDEDEMKT-FVPIaWaaaYGNN-------------------------------- [420]
XP_033067617.1|ZFY|T_francoisi       -VHEQQIDEDEMKT-FVPIaWaaaYGNN-------------------------------- [420]
XP_031516968.1|ZFY_X1|P_anubis       -VHEQQIDEDEMKT-FVPIaWaaaYGNN-------------------------------- [420]
XP_008017167.1|C_sabaeus             -VHEQQIDEDEMKT-FVPIaWaaaYGNN-------------------------------- [420]
XP_030782172.1|ZFY_X1|R_roxellana    -VHEQQIDEDEMKT-FVPIaWaaaYGNN-------------------------------- [420]
XP_032612406.1|ZFY_X1|H_moloch       -VHEQQIDEDEMKT-FVPIaWaaaYGNN-------------------------------- [420]
XP_035145821.1|ZFY_X2|C_jacchus      -MHEQQIDEDEMKT-FVPIaWaaaYGNN-------------------------------- [420]
P10925.3|ZFY1_MOUSE|M_musculus       -EHEQQMDVSEIKAAFLPIaWTaaYDNN-------------------------------- [420]
P20662.2|ZFY2_MOUSE|M_musculus       -EHEQQMDVSEIKAAFLPIaWTaaYDNN-------------------------------- [420]
XP_008771898.1|R_norvegicus          -EHEQQMDDSEIKAAFLPIaWaaaYDNN-------------------------------- [420]
XP_015343506.1|M_marmota             -VHEQQIDDSEMKA-FMPIaWaaaYGNN-------------------------------- [420]
Q95LI3.1|ZFY|B_taurus                TVHEQEMDDSEIKT-FMPIaWaaaYGNN-------------------------------- [420]
XP_010855418.1|B_bison               TVHEQEMDDSEIKT-FMPIaWaaaYGNN-------------------------------- [420]
XP_017900383.1|C_hircus              TVHEQEMDDSEIKT-FMPIaWaaaYGNN-------------------------------- [420]
AMY96563.1|ZFY|C_elaphus             TVHEQEMDDNEMKT-FMPIaWaaaYGNN-------------------------------- [420]
XP_020759307.1|ZFY_X1|O_virginianus  TVHEQEMDDSEMKT-FMPIaWaaaYGNN-------------------------------- [420]
F1SPY3|ZFY|S_scrofa                  aVHEQQMDDSEIKT-FMPIaWaaaYGNN-------------------------------- [420]
XP_024612082.1|ZFY|N_asiaeorientalis aVHEQQMDNSEIKT-FMPIaWaaaYGNN-------------------------------- [420]
AKI82174.1|ZFY|C_lupus               aVHEQQMDDNEIKT-FMPIaWaaaYGNN-------------------------------- [420]
XP_032187800.1|ZFX_like_X1|M_erminea aVHEQQMDDNEIKT-FMPIaWaaaYGNN-------------------------------- [420]
JAC06687.1|ZFY|L_africana            aVHEQQMDDSEIKT-FVPIaWaaaYGNN-------------------------------- [420]
No_accession|PREDICTED_ZFY|E_caballusaVHEQQMDDSEIKT-FMPIaWaaaYGNN-------------------------------- [420]
XP_028935710.1|ZFY_X2|O_anatinus     --HEQQMDDTEIKT-FMPIaWaaaYGNN-------------------------------- [420]
XP_016288863.1|M_domestica           --HEQQIDDTEIKT-FMPIaWaaaYGNN-------------------------------- [420]
TKS65875.1|ZFY1|C_lucidus            VTHDRSYDSVSLSKDFMPVaWaaaYGAE-------------------------------- [420]
XP_028451227.1|P_flavescens          VTQDRSYDSVSLSKDFMPVaWaaaYGAE-------------------------------- [420]
XP_028276673.1|ZFY1_like_X1|P_ranga  VTHDRSFDSVSLSKDFMPVaWaaaYGAD-------------------------------- [420]
XP_010749798.1|ZFY1_X1|L_crocea      VTHDRSYDSVSLSKDFMPVaWaaaYGAE-------------------------------- [420]
XP_023133903.1|A_ocellaris           VTHDRSYDSVSLSKDFMPVaWaaaYGAE-------------------------------- [420]
XP_024253620.1|O_tshawytscha         VTHERSYDGTALSKDFMPVaWaaTYGSE-------------------------------- [420]
XP_023843891.1|ZFY1|S_alpinus        VTHERSYDGTALSKDFMPVaWaaTYGSE-------------------------------- [420]
XP_020321060.1|ZFY1_like|O_kisutch   VTHERSYDGTALSKDFMPVaWaaTYGSE-------------------------------- [420]
XP_004564062.1|ZFY1_X1|M_zebra       VTHDRSYDSVSLSKDFMPVaWaaaYGAE-------------------------------- [420]
XP_026038267.1|A_calliptera          VTHDRSYDSVSLSKDFMPVaWaaaYGAE-------------------------------- [420]
XP_011609888.1|ZFY_X1|T_rubripes     --------------------------AE-------------------------------- [420]
XP_029029380.1|ZFY_X1|B_splendens    VTHDRSYDSVSLSKDFMPVaWaaaYGAE-------------------------------- [420]
XP_028839070.1|D_clupeiodes          VSHERPYDSTALSKDFMPVaWaaaYGAE-------------------------------- [420]
ROL53794.1|ZFY1|A_grahami            VPHDRPYDGTALSKDFMPVaWaaaYGIHCCLLGPGNKCRYCVTASEGMTTEAFNALSRGG [420]
XP_023277193.1|S_dorsalis            VTHDRSYDSVSLSKDFMPVaWaaaYGAE-------------------------------- [420]
XP_008331409.1|C_semilaevis          --------------------------VD-------------------------------- [420]
XP_015127980.1|ZFY_X1|G_gallus       --HEQQIDDNEIKT-FMPIaWaaaYGNN-------------------------------- [420]
Q01611.1|ZFY1_XENLA|X_laevis         --HEHQLEDAELSKTFMPVaWaaaYGNN-------------------------------- [420]
```



```
                                          *     *** *   **           *   **    * * **   * *  * **
NP_001356631.1|ZFY|H_sapiens         -SDGIENRNGTASALLHIDESAGLGRLAKQKPKKKR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_009443992.1|ZFY_X1|P_troglodytes  -SDGIENRNGTASALLHIDESAGLGRLAKQKPKKKR-RPDSRQYQTAIIIGPDGHPLTVY [480]
Q52V16.1|ZFY|G_gorilla               -SDGIENRNGTASALLHIDESAGLGRLAKQKPKKKR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_014984082.1|ZFY_X1|M_mulatta      -SDGIENRNGTASALLHVDESTGLGRLAKQKPKKKR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_033067617.1|ZFY|T_francoisi       -SDGIENRNGTASALLHVDESAGLGRLAKQKPKKKR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_031516968.1|ZFY_X1|P_anubis       -SDGIENRNGTASALLHVDESTGLGRLAKQKPKKKR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_008017167.1|C_sabaeus             -SDGIENRNGTASALLHVDESTGLGRLAKQKPKKKR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_030782172.1|ZFY_X1|R_roxellana    -SDGIENRNGTASALLHVDESAGLGRLAKQKPKKKR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_032612406.1|ZFY_X1|H_moloch       -SDGIENRNGTASALLHVDESAGLGRLAKQKPKKKR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_035145821.1|ZFY_X2|C_jacchus      -SDGIENRNGSASAVLHVDESVGLSRLTKQKPKKKR-RSDARQYQTAIIIGPDGHPLTVY [480]
P10925.3|ZFY1_MOUSE|M_musculus       -SDEIEDQNVTASALLNQDESGGLDRVPKQKSKKKK-RPESKQYQSAIFVAPDGQTLRVY [480]
P20662.2|ZFY2_MOUSE|M_musculus       -SDEIEVQNATASAMLHHDESGGLDRVPKQKSKKKK-RPESKQYQSAIFVAPDGQTLRVY [480]
XP_008771898.1|R_norvegicus          -SDEIEEQNVTASAVLHQNESGGLDRVHKQKAKKKK-RPESKQYQTAIIVAPDGQTLIVY [480]
XP_015343506.1|M_marmota             -SDGIENRNGTASALLHIDESAGLSRLAKQKPKKRR-RPDSKQYQTAIIIGPDGHPLTVY [480]
Q95LI3.1|ZFY|B_taurus                -SDGIENRSGTASALLHIDESAGLGRLTKHKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_010855418.1|B_bison               -SDGIENRSGTASALLHIDESAGLGRLTKHKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_017900383.1|C_hircus              -SDGIENRNGTASALLHIDESAGLGRLAKQKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
AMY96563.1|ZFY|C_elaphus             -SDGIENRNGTASALLHIDESAGLGRLAKQKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_020759307.1|ZFY_X1|O_virginianus  -SDGIENRNGTASALLHIDESAGLGRLAKQKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
F1SPY3|ZFY|S_scrofa                  -SDGIENRNGTASALLHIDESAGLGRLAKQKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_024612082.1|ZFY|N_asiaeorientalis -SDGIENRNGTASALLHIDESAGLGRLAKQKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
AKI82174.1|ZFY|C_lupus               -SDGIENRNGTASALLHIDESAGLGRLAKQKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_032187800.1|ZFX_like_X1|M_erminea -SDGIENRNGTASALLHIDESAGLSRLAKQKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
JAC06687.1|ZFY|L_africana            -SDGIENRNGTASALLHIDESAGLGRVTKQKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
No_accession|PREDICTED_ZFY|E_caballus-SDGIENRNGTASALLHIDESAGLGRLAKQKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_028935710.1|ZFY_X2|O_anatinus     -TDGIENRNGTASALLHIDESAGLGRLAKQKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
XP_016288863.1|M_domestica           -TDGIENRNGTASALLHIDESAGLGRLAKQKPKKRR-RPDSRQYQTAIIIGPDGHPLTVY [480]
TKS65875.1|ZFY1|C_lucidus            DSESCENRNGAASALLHIDESDGIDEINRQRNKSKR-RSEPRQVQTAIIIGPYGQPLTVY [480]
XP_028451227.1|P_flavescens          DSESCENRNGAASALLHIDESDGVDEINRQRNKSRR-RSEPRQVQTAIIIGPYGQPLTVY [480]
XP_028276673.1|ZFY1_like_X1|P_ranga  DGEGCENRNGAASALLHIDESDGVDEINRQRNKSKR-RSEHRQVQTAIIIGPYGQPLTVY [480]
XP_010749798.1|ZFY1_X1|L_crocea      DSESCENRNGAASALLHIDESDGIDEINRQRNKSKR-RSEPRQVQTAIIIGPYGQPLTVY [480]
XP_023133903.1|A_ocellaris           DSESCENRNGAASALLHIDESDGVDEINRQRSKNKR-RSEPRQVQTAIIIGPYGQPLTVY [480]
XP_024253620.1|O_tshawytscha         DSESCENRNGAASALLHIDESDGADKLNRQRNKNKRTRAEPRQVQTAIIIGPYGQPLTVY [480]
XP_023843891.1|ZFY1|S_alpinus        DSESCENRNGAASALLHIDESDGGDKLNRQRNKNKRTRAEPRQVQTAIIIGPYGQPLTVY [480]
XP_020321060.1|ZFY1_like|O_kisutch   DSESCENRNGAASALLHIDESDGADKLNRQRNKNKRTRAEPRQVQTAIIIGPYGQPLTVY [480]
XP_004564062.1|ZFY1_X1|M_zebra       DSESCENRNGAASALLHIDESDGVDEINRQRNKTKR-RSEPRQVQTAIIIGPYGQPLTVY [480]
XP_026038267.1|A_calliptera          DSESCENRNGAASALLHIDESDGVDEINRQRNKTKR-RSEPRQVQTAIIIGPYGQPLTVY [480]
XP_011609888.1|ZFY_X1|T_rubripes     DSESCENRNGAASALLHIDESEGVDDIGRQRSKNKR-RSEHRQVQTAIIIGPYGQPLTVY [480]
XP_029029380.1|ZFY_X1|B_splendens    DSESCENRNGAASALLHIDESDGVDEINRQRAKSRR-RSEPRQVQTAIIIGPYGQPLTVY [480]
XP_028839070.1|D_clupeiodes          DSEGCENRNGAASALLHIDESDGVDKLNRQLGKNKR-RAEPRQVQTAIIIGPYGQPLTVY [480]
ROL53794.1|ZFY1|A_grahami            ADEGCENRNGAASALLHIDESDALDKLNRQHGKNKR-RAEPRQVQTAIIIGPYGQPLTVY [480]
XP_023277193.1|S_dorsalis            DSESCENRNGAASALLHIDESDGVDEINRQRNKSKR-RSEPRQVQTAIIIGPYGQPLTVY [480]
XP_008331409.1|C_semilaevis          DGDGCENRNGAASALLHIDESDGIDEISRQRTKSRR-RSEPRQVQTAIIIGPYGQPLTVY [480]
XP_015127980.1|ZFY_X1|G_gallus       -NDGIESRNGTASALLHIDESAGLGRLAKQKPKKKR-RPESRQYQTAIIIGPDGHPLTVY [480]
Q01611.1|ZFY1_XENLA|X_laevis         -TDGIEHRNGTASALLHIDESDGLDRLTKQKLKKKR-RGENRQYQTAIIIGPDGHPLTVY [480]

                                     **  ******   ****   *        *** ** *   ****   *** * * *
NP_001356631.1|ZFY|H_sapiens         PCMICGKKFKSRGFLKRHMKNHPEH-LAKKKYHCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_009443992.1|ZFY_X1|P_troglodytes  PCMICGKKFKSRGFLKRHMKNHPEH-LAKKKYHCTDCDYTTNKKISLHNHLESHKLTS-- [540]
Q52V16.1|ZFY|G_gorilla               PCMICGKKFKSRGFLKRHMKNHPEH-LAKKKYHCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_014984082.1|ZFY_X1|M_mulatta      PCMICGKKFKSRGFLKRHMKNHPEH-LAKKKYHCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_033067617.1|ZFY|T_francoisi       PCMICGKKFKSRGFLKRHMKNHPEH-LAKKKYHCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_031516968.1|ZFY_X1|P_anubis       PCMICGKKFKSRGFLKRHMKNHPEH-LAKKKYHCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_008017167.1|C_sabaeus             PCMICGKKFKSRGFLKRHMKNHPEH-LAKKKYHCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_030782172.1|ZFY_X1|R_roxellana    PCMICGKKFKSRGFLKRHMKNHPEH-LAKKKYHCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_032612406.1|ZFY_X1|H_moloch       PCMICGKKFKSRGFLKRHMKNHPEH-LAKKKYHCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_035145821.1|ZFY_X2|C_jacchus      PCMICGKKFKSRGFLKRHMKNHPEH-LAKKKYHCTDCDYTTNKKISLHNHLESHKLTS-- [540]
P10925.3|ZFY1_MOUSE|M_musculus       PCMFCGKKFKTKRFLKRHTKNHPEY-LANKKYHCTECDYSTNKKISLHNHMESHKLTI-- [540]
P20662.2|ZFY2_MOUSE|M_musculus       PCMFCGKKFKTKRFLKRHIKNHPEY-LANKKYHCTECDYSTNKKISLHNHMESHKLTI-- [540]
XP_008771898.1|R_norvegicus          PCMFCGKKFKTKSFLKRHIKNHPEY-LAKKKYHCTDCDYSTNKKISLHNHMESHKLTI-- [540]
XP_015343506.1|M_marmota             PCMICGKKFKSRGFLKRHMKNHPEH-IAKKKYCCTDCDYTTNKKISLHNHLESHKLTS-- [540]
Q95LI3.1|ZFY|B_taurus                PCMICGKKFKSRGFLKRHMKNHPEH-LTKKKYRCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_010855418.1|B_bison               PCMICGKKFKSRGFLKRHMKNHPEH-LTKKKYRCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_017900383.1|C_hircus              PCMICGKKFKSRGFLKRHMKNHPEH-LTKKKYRCTDCDYTTNKKISLHNHLESHKLTS-- [540]
AMY96563.1|ZFY|C_elaphus             PCMICGKKFKSRGFLKRHMKNHPEH-LTKKKYRCTDCDYTTNKKMSLHNHLESHKLTS-- [540]
XP_020759307.1|ZFY_X1|O_virginianus  PCMICGKKFKSRGFLKRHMKNHPEH-LTKKKYRCTDCDYTTNKKISLHNHLESHKLTS-- [540]
F1SPY3|ZFY|S_scrofa                  PCLICGKKFKSRGFLKRHMKNHPEH-LTKKKYRCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_024612082.1|ZFY|N_asiaeorientalis PCMICGKKFKSRGFLKRHMKNHPEH-LTKKKYHCTACDYTTNKKISFHNHLESHKLTS-- [540]
AKI82174.1|ZFY|C_lupus               PCMICGKKFKSRGFLKRHMKNHPEH-LSKKKYRCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_032187800.1|ZFX_like_X1|M_erminea PCMICGKKFKSRGFLKRHMKNHPEH-LTKKKYRCTDCEYTTNKKISLHNHLESHKLTS-- [540]
JAC06687.1|ZFY|L_africana            PCMICGKKFKSRGFLKRHMKNHPEH-LTKKKYRCTDCDYTTNKKISLHNHLESHKLTS-- [540]
No_accession|PREDICTED_ZFY|E_caballusPCMICGKKFKSRGFLKRHMKNHPEH-LTKKKYHCTDCDYTTNKKISLHNHLESHKLTS-- [540]
XP_028935710.1|ZFY_X2|O_anatinus     PCMICGKKFKSRGFLKRHMKNHPEH-LSKKKYRCTDCDYTTNKKVSLHNHLESHKLTG-- [540]
XP_016288863.1|M_domestica           PCMICGKKFKSRGFLKRHMKNHPEH-LTKKKYRCTDCDYTTNKKISLHNHLESHKLTN-- [540]
TKS65875.1|ZFY1|C_lucidus            PCMLCGKKFKSRGFLKRHTKNNHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHALSS-- [540]
XP_028451227.1|P_flavescens          PCMLCGKKFKSRGFLKRHTKNHHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHALSS-- [540]
XP_028276673.1|ZFY1_like_X1|P_ranga  PCMLCGKKFKSRGFLKRHTKNHHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHALSS-- [540]
XP_010749798.1|ZFY1_X1|L_crocea      PCMLCGKKFKSRGFLKRHTKNNHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHALSS-- [540]
XP_023133903.1|A_ocellaris           PCMLCGKKFKSRGFLKRHTKNHHQDVLTRKKYQCTDCEFTTNKKASLHNHMEVHALSS-- [540]
XP_024253620.1|O_tshawytscha         PCMLCGKKFKSRGFLKRHTKNHHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHTLSN-- [540]
XP_023843891.1|ZFY1|S_alpinus        PCMLCGKKFKSRGFLKRHTKNHHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHTLSN-- [540]
XP_020321060.1|ZFY1_like|O_kisutch   PCMLCGKKFKSRGFLKRHTKNHHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHTLSN-- [540]
XP_004564062.1|ZFY1_X1|M_zebra       PCMLCGKKFKSRGFLKRHTKNHHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHALSS-- [540]
XP_026038267.1|A_calliptera          PCMLCGKKFKSRGFLKRHTKNHHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHALSS-- [540]
XP_011609888.1|ZFY_X1|T_rubripes     PCMLCGKKFKSRGFLKRHTKNHHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHALSS-- [540]
XP_029029380.1|ZFY_X1|B_splendens    PCMLCGKKFKSRGFLKRHTKNHHQDILTRKKYQCTDCDFTTNKKASLHNHMEVHALSS-- [540]
XP_028839070.1|D_clupeiodes          PCMLCGKKFKSRGFLKRHTKNHHQDVLSRKKYQCTDCDFITNKKANLHNHMEVHALST-- [540]
ROL53794.1|ZFY1|A_grahami            PCMLCGKKFKSRGFLKRHTRNHHQDALSRKKYQCTDCDFTTNKKASLHNHMEVHALSN-- [540]
XP_023277193.1|S_dorsalis            PCMLCGKKFKSRGFLKRHTKNHHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHALSS-- [540]
XP_008331409.1|C_semilaevis          PCMLCGKKFKSRGFLKRHTKNHHQDVLTRKKYQCTDCDFTTNKKASLHNHMEVHALSS-- [540]
XP_015127980.1|ZFY_X1|G_gallus       PCMICGKKFKSRGFLKRHMKNHPEHLLTKKKYRCTDCDYTTNKKISLHNHLESHKLTN-- [540]
Q01611.1|ZFY1_XENLA|X_laevis         PCMICGKKFKSRGFLKRHMKNHPEH-LVRKKYRCTDCDYTTNKKVSLHNHLESHKLTATV [540]
```



```
                                       *               * *       *        **               *  *
NP_001356631.1|ZFY|H_sapiens         -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
XP_009443992.1|ZFY_X1|P_troglodytes  -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
Q52V16.1|ZFY|G_gorilla               -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
XP_014984082.1|ZFY_X1|M_mulatta      -KAEKAIQ------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
XP_033067617.1|ZFY|T_francoisi       -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
XP_031516968.1|ZFY_X1|P_anubis       -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
XP_008017167.1|C_sabaeus             -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
XP_030782172.1|ZFY_X1|R_roxellana    -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
XP_032612406.1|ZFY_X1|H_moloch       -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
XP_035145821.1|ZFY_X2|C_jacchus      -KAEKTIE------CVECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
P10925.3|ZFY1_MOUSE|M_musculus       -KTEKTTE------CDDCRKNLSHAGTLCTHKTMHT--EKG-VN---------------- [600]
P20662.2|ZFY2_MOUSE|M_musculus       -KTEKTTE------CDDCRKNLSHAG------TMHT--EKG-VN---------------- [600]
XP_008771898.1|R_norvegicus          -KTEKTTE------CDDCGKHLSHAG---TLCTHKK--EKEKVS---------------- [600]
XP_015343506.1|M_marmota             -KVEKVIE------CDECGKHFSHTGALFTHKMVHK--EKGNN----------------- [600]
Q95LI3.1|ZFY|B_taurus                -KSEKAIE------CDDCGKHFSHAGALFTHKMVHK--EKGAS----------------- [600]
XP_010855418.1|B_bison               -KSEKAIE------CDDCGKHFSHAGALFTHKMVHK--EKGAS----------------- [600]
XP_017900383.1|C_hircus              -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAS----------------- [600]
AMY96563.1|ZFY|C_elaphus             -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
XP_020759307.1|ZFY_X1|O_virginianus  -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
F1SPY3|ZFY|S_scrofa                  -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
XP_024612082.1|ZFY|N_asiaeorientalis -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
AKI82174.1|ZFY|C_lupus               -KAEKSIE------CEECGKHFSHAGALFTHKMVHK--EKGTN----------------- [600]
XP_032187800.1|ZFX_like_X1|M_erminea -KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
JAC06687.1|ZFY|L_africana            -KAEKAIE------CDECGKHFSHTGALFTHKMVHK--EKGSS----------------- [600]
No_accession|PREDICTED_ZFY|E_caballus-KAEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
XP_028935710.1|ZFY_X2|O_anatinus     -KAEKAAAPGAGAECDECGKHFSHAGALFTHKTVHK--EKGAGG---------------- [600]
XP_016288863.1|M_domestica           -KTEKAIE------CDECGKHFSHAGALFTHKMVHK--EKGAN----------------- [600]
TKS65875.1|ZFY1|C_lucidus            ---KAPFE------CEMCGKEFHQQAALFSHRLQHHHREPKNQ------------PPPPP [600]
XP_028451227.1|P_flavescens          ---KAPFE------CEMCGKEFHQQAALFSHRLQHHHREPKSQ------------PPPTP [600]
XP_028276673.1|ZFY1_like_X1|P_ranga  ---KAPFE------CEMCGKEFHQQAALFSHRLQHHHREPKSQ------------PPPTP [600]
XP_010749798.1|ZFY1_X1|L_crocea      ---KAPFE------CEMCGKEFHQQAALFSHRLQHHHREPKNQ------------PPPPP [600]
XP_023133903.1|A_ocellaris           ---KAPFE------CEMCGKEFHQQAALFSHRLQHHHREPKNQ------------PPPPP [600]
XP_024253620.1|O_tshawytscha         ---KALFE------CEVCGKEFHQQAALFSHRLQHHHREPKSP------------VPSQA [600]
XP_023843891.1|ZFY1|S_alpinus        ---KALFE------CEVCGKEFHQQAALFSHRLQHHHREPKSP------------VPSQA [600]
XP_020321060.1|ZFY1_like|O_kisutch   ---KALFE------CEVCGKEFHQQAALFSHRLQHHHREPKSP------------VPSQA [600]
XP_004564062.1|ZFY1_X1|M_zebra       ---KAPFE------CEMCGKEFHQQAALFSHRLQHHHREPKSQ------------PPPPP [600]
XP_026038267.1|A_calliptera          ---KAPFE------CEMCGKEFHQQAALFSHRLQHHHREPKSQ------------PPPPP [600]
XP_011609888.1|ZFY_X1|T_rubripes     ---KAPFE------CEMCGKEFHQQAALFSHRLQHHHREPKTQ------------PPPTP [600]
XP_029029380.1|ZFY_X1|B_splendens    ---KAPFE------CEMCGKEFHQQSALFSHRLQHQHREPKTQ------------PPQTP [600]
XP_028839070.1|D_clupeiodes          ---KAPFE------CETCGKEFHQQAALFSHRLQHHHREQKAP---------PPPPPPAS [600]
ROL53794.1|ZFY1|A_grahami            ---KAPFE------CETCGKEFHQQAALFSHRLQHHHREQKAPTAIA-----SPLPPAAA [600]
XP_023277193.1|S_dorsalis            ---KAPFE------CEMCGKEFHQQAALFSHRLQHHHREPKNQ------------PPPPP [600]
XP_008331409.1|C_semilaevis          ---KAPFE------CEMCGKEFHQQTALFSHRLQHHHREPKNQQPPPPPPPPPPPPPPQP [600]
XP_015127980.1|ZFY_X1|G_gallus       -KTEKLIE------RDECGKSFSHAGALFAHKMVHR--DKGVN----------------- [600]
Q01611.1|ZFY1_XENLA|X_laevis         IKTEKDLE------CEECGKIFLHANALFAHKLTHN--EKAGN----------------- [600]

                                          **** * ****  *  *** ** *  ** * ********** * ** * ******
NP_001356631.1|ZFY|H_sapiens         -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELRKHMRIHTGEKP [660]
XP_009443992.1|ZFY_X1|P_troglodytes  -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
Q52V16.1|ZFY|G_gorilla               -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_014984082.1|ZFY_X1|M_mulatta      -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_033067617.1|ZFY|T_francoisi       -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_031516968.1|ZFY_X1|P_anubis       -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_008017167.1|C_sabaeus             -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_030782172.1|ZFY_X1|R_roxellana    -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_032612406.1|ZFY_X1|H_moloch       -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_035145821.1|ZFY_X2|C_jacchus      -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
P10925.3|ZFY1_MOUSE|M_musculus       -KTCKCKFCDYETAEQTLLNHHLLVVHRKKFPHICGECGKGFRHPSALKKHIRVHTGEKP [660]
P20662.2|ZFY2_MOUSE|M_musculus       -KTCKCKFCDYETAEQTLLNHHLLVVHRKKFPHICGECGKGFRHPSALKKHIRVHTGEKP [660]
XP_008771898.1|R_norvegicus          -KTYKCKFCDYETAEQTSLNHHLLAVHSKKYPHVCVECGKGFRHPSELKKHIRVHTGEKP [660]
XP_015343506.1|M_marmota             -KMHKCKFCEYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
Q95LI3.1|ZFY|B_taurus                -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_010855418.1|B_bison               -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_017900383.1|C_hircus              -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELRKHMRIHTGEKP [660]
AMY96563.1|ZFY|C_elaphus             -KMHKCKFCEYGTAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_020759307.1|ZFY_X1|O_virginianus  -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
F1SPY3|ZFY|S_scrofa                  -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_024612082.1|ZFY|N_asiaeorientalis -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
AKI82174.1|ZFY|C_lupus               -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_032187800.1|ZFX_like_X1|M_erminea -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
JAC06687.1|ZFY|L_africana            -KMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
No_accession|PREDICTED_ZFY|E_caballus-KMHRCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_028935710.1|ZFY_X2|O_anatinus     -RTHKCKFCDYETAEQGLLNRHLLAVHSKNFPHVCVECGKGFRHPSELKKHMRIHTGEKP [660]
XP_016288863.1|M_domestica           -KMHKCKFCDYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
TKS65875.1|ZFY1|C_lucidus            TKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_028451227.1|P_flavescens          NKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_028276673.1|ZFY1_like_X1|P_ranga  TKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_010749798.1|ZFY1_X1|L_crocea      TKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_023133903.1|A_ocellaris           TKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_024253620.1|O_tshawytscha         NKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_023843891.1|ZFY1|S_alpinus        NKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_020321060.1|ZFY1_like|O_kisutch   NKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_004564062.1|ZFY1_X1|M_zebra       TKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_026038267.1|A_calliptera          TKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_011609888.1|ZFY_X1|T_rubripes     TKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_029029380.1|ZFY_X1|B_splendens    TKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_028839070.1|D_clupeiodes          TKMHKCKFCNYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
ROL53794.1|ZFY1|A_grahami            TKTHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_023277193.1|S_dorsalis            TKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_008331409.1|C_semilaevis          TKMHKCKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
XP_015127980.1|ZFY_X1|G_gallus       -KMHKCKFCDYETAEQGLLSHHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTGEKP [660]
Q01611.1|ZFY1_XENLA|X_laevis         -KMHICKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTGEKP [660]
```

```
                                     * *  * * ********** *  *       *  *  **        *   *
NP_001356631.1|ZFY|H_sapiens         YQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ESKTHQ [720]
XP_009443992.1|ZFY_X1|P_troglodytes  YQCQYCEYRSADSSNLKTHIKTKHSKEMPLKCDICLLTFSDTKEVQQHTLVHQ-ESKTHQ [720]
Q52V16.1|ZFY|G_gorilla               YQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ESKTHQ [720]
XP_014984082.1|ZFY_X1|M_mulatta      YQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ENKTHQ [720]
XP_033067617.1|ZFY|T_francoisi       YQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ENKTHQ [720]
XP_031516968.1|ZFY_X1|P_anubis       YQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ESKTHQ [720]
XP_008017167.1|C_sabaeus             YQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHALIHQ-ESKTHQ [720]
XP_030782172.1|ZFY_X1|R_roxellana    YQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQHHTLVHQ-ENRTHQ [720]
XP_032612406.1|ZFY_X1|H_moloch       YQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ESKTHQ [720]
XP_035145821.1|ZFY_X2|C_jacchus      YQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ESRTHQ [720]
P10925.3|ZFY1_MOUSE|M_musculus       YECQYCEYKSADSSNLKTHIKSKHSKEIPLKCGICLLTFSDNKEAQQHAVLHQ-ESRTHQ [720]
P20662.2|ZFY2_MOUSE|M_musculus       YECQYCEYKSADSSNLKTHIKSKHSKEIPLKCGICLLTFSDTKEAQQHAVLHQ-ESRTHQ [720]
XP_008771898.1|R_norvegicus          YQCQYCEYKSADSSNLKTHIKTKHSKDIPLKCGICLMTFSDTKEAQQHALIHQ-ENRTHQ [720]
XP_015343506.1|M_marmota             YQCQYCEYRSADSSNLKTHVKTKHTKEMPFKCDICLLTFSDTKEVQQHALIHQ-ESKTHQ [720]
Q95LI3.1|ZFY|B_taurus                YQCQYCEYRSADSSNLKTHVKTKHSKEMSFKCDICLLTFSDTKEVQQHALIHQ-ESKTHQ [720]
XP_010855418.1|B_bison               YQCQYCEYRSADSSNLKTHVKTKHSKEMSFKCDICLLTFSDTKEVQQHALIHQ-ESKTHQ [720]
XP_017900383.1|C_hircus              YQCQYCEYRSADSSNLKTHVKTKHSKEMSFKCDICLLTFSDTKEVQQHALIHQ-ESKTHQ [720]
AMY96563.1|ZFY|C_elaphus             YRCQYCEYRSADSSNLKTHVKTKHSKEMSFKCDICLLTFSDTKEVQQHALIHQ-ESKTHQ [720]
XP_020759307.1|ZFY_X1|O_virginianus  YQCQYCEYRSADSSNLKTHVKTKHSKEMSFNCDICLLTFSDTKEVQQHALIHQ-ESKTHQ [720]
F1SPY3|ZFY|S_scrofa                  YQCQYCEYRSADSSNLKTHVKTKHSKEMPFKCDICLLTFSDTKDVQQHALIHQ-ESKTHQ [720]
XP_024612082.1|ZFY|N_asiaeorientalis YQCQYCEYRSADSSNLKTHVKTKHSKEMPFKCDSCLLTFSDTKEVQQHALIHQ-ESKTHQ [720]
AKI82174.1|ZFY|C_lupus               YQCQYCEYRSADSSNLKTHVKTKHSKEMPFKCDICLLTFSDTKEVQQHAVIHQ-ESKTHQ [720]
XP_032187800.1|ZFX_like_X1|M_erminea YQCQYCEYRSADSSNLKTHVKTKHSKEMPFKCDICLLTFSDTKEVQQHALIHQ-ESKTHQ [720]
JAC06687.1|ZFY|L_africana            YQCQYCEYRSADSSNLKTHVKTKHSKEMPYRCDICLLTFSDTKEVQQHAVIHQ-ESKTHQ [720]
No_accession|PREDICTED_ZFY|E_caballusYHCQYCEYRSADSSNLKTHVKTKHSKEMPFKCDICLLTFSDTKEVQQHGLIHQ-ESKTHQ [720]
XP_028935710.1|ZFY_X2|O_anatinus     YQCQFCPYRSADSSNLKTHVKTKHSKETPFRCEACPLTFADPKELQQHALLHHQESRAHQ [720]
XP_016288863.1|M_domestica           YQCQYCEYRSADSSNLKTHVKTKHSKEMPFKCEICLLTFSDTKEVQQHALIHQ-ESKTHQ [720]
TKS65875.1|ZFY1|C_lucidus            YSCLYCDYKSADSSNLKTHIKTKHSKEMPYKCERCFQTFAEEDELMQHGLTHE-ENKTHH [720]
XP_028451227.1|P_flavescens          YSCLYCDYKSADSSNLKTHIKTKHSKEMPYKCERCFQTFAEEEELMQHGLTHE-ENKTHH [720]
XP_028276673.1|ZFY1_like_X1|P_ranga  YSCLYCDYKSADSSNLKTHIKTKHSKEMPYKCERCFQTFAEEEELMQHGLTHE-ENKTHH [720]
XP_010749798.1|ZFY1_X1|L_crocea      YSCLYCDYKSADSSNLKTHIKTKHSKEMPYKCERCFQTFAEEDELMQHGLTHE-ENKTHH [720]
XP_023133903.1|A_ocellaris           YSCLYCDYKSADSSNLKTHIKTKHSKEMPYKCERCFQTFAEEEELMQHGLTHE-ENKTHH [720]
XP_024253620.1|O_tshawytscha         YSCLYCDYKSADSSNLKTHVKTKHSKEMPFKCERCFQTFAEEEELLQHGLTHE-EAKTHQ [720]
XP_023843891.1|ZFY1|S_alpinus        YSCLYCDYKSADSSNLKTHVKTNHSKEMPFKCDRCFQTFAEEEELLQHGLTHE-EAKTQ? [720]
XP_020321060.1|ZFY1_like|O_kisutch   YSCLYCDYKSADSSNLKTHVKTKHSKEMPFKCERCFQTFAEEEELLQHGLTHE-EAKTHQ [720]
XP_004564062.1|ZFY1_X1|M_zebra       YSCLYCDYKSADSSNLKTHIKTKHSKEMPYKCERCFQTFAEEEELMQHGLTHE-ENKTHH [720]
XP_026038267.1|A_calliptera          YSCLYCDYKSADSSNLKTHIKTKHSKEMPYKCERCFQTFAEEEELMQHGLTHE-ENKTHH [720]
XP_011609888.1|ZFY_X1|T_rubripes     YSCLYCDYKSADSSNLKTHIKTKHSKEMPYKCERCFQTFAEEEELMQHGLTHE-ENKTHL [720]
XP_029029380.1|ZFY_X1|B_splendens    YSCLYCDYKSADSSNLKTHIKTKHSKEMPYKCERCFQTFAEEEELIQHGLTHE-ENKTHH [720]
XP_028839070.1|D_clupeiodes          YSCLYCDYKSADSSNLKTHVKTKHSREMPYKCERCCQTFSEEEELAQHATTHE-EARTHQ [720]
ROL53794.1|ZFY1|A_grahami            YSCLYCDYKSADSSNLKTHVKTKHSRELPFRCERCGQTFAEEDELTQHAATHE-DARGHQ [720]
XP_023277193.1|S_dorsalis            YSCLYCDYKSADSSNLKTHIKTKHSKEMPYKCERCFQTFAEEEELMQHGLTHE-ENKTHH [720]
XP_008331409.1|C_semilaevis          YSCLYCDYKSADSSNLKTHIKTKHSKELPYKCERCFQTFAEEEELIQHGLTHE-ENKTHH [720]
XP_015127980.1|ZFY_X1|G_gallus       YQCQYCEYRSADSSNLKTHVKTKHSKETSSKCDICFQTFSDTKELQQHTLMHQ-ESKTHQ [720]
Q01611.1|ZFY1_XENLA|X_laevis         YLCQYCDYRSADSSNLKTHVKTKHSKEMPFKCDICLQTFTDSKDLQEHAILHQ-ESKNHQ [720]

                                     *  * * ********** ****** *****  * ************   *   * *****
NP_001356631.1|ZFY|H_sapiens         CLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQCRH [780]
XP_009443992.1|ZFY_X1|P_troglodytes  CLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQCRH [780]
Q52V16.1|ZFY|G_gorilla               CLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQCRH [780]
XP_014984082.1|ZFY_X1|M_mulatta      CLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQCRH [780]
XP_033067617.1|ZFY|T_francoisi       CLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQCRH [780]
XP_031516968.1|ZFY_X1|P_anubis       CLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQCRH [780]
XP_008017167.1|C_sabaeus             CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
XP_030782172.1|ZFY_X1|R_roxellana    CLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQCRH [780]
XP_032612406.1|ZFY_X1|H_moloch       CLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQCRH [780]
XP_035145821.1|ZFY_X2|C_jacchus      CLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCDKGFHRPSELKKHVAVHKGKKMHQCRH [780]
P10925.3|ZFY1_MOUSE|M_musculus       CSHCNHKSSNSSDLKRHIISVHTKAYPHKCDMCSKGFHRPSELKKHVATHKSKKMHQCRH [780]
P20662.2|ZFY2_MOUSE|M_musculus       CSHCNHKSSNSSDLKRHIISVHTKAYPHKCDMCSKGFHRPSELKKHVATHKSKKMHQCRH [780]
XP_008771898.1|R_norvegicus          CSYCNHKSSNSSDLKRHIISVHTKDYPHKCDMCSKGFHRPSELKKHVATHKSKKMHQCRH [780]
XP_015343506.1|M_marmota             CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
Q95LI3.1|ZFY|B_taurus                CVHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
XP_010855418.1|B_bison               CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
XP_017900383.1|C_hircus              CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
AMY96563.1|ZFY|C_elaphus             CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
XP_020759307.1|ZFY_X1|O_virginianus  CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
F1SPY3|ZFY|S_scrofa                  CLHCDHKSSNSSDLKRHIISVHTKDYPHKCEMCEKGFHRPSELKKHVAAHKGKKMHQCRH [780]
XP_024612082.1|ZFY|N_asiaeorientalis CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
AKI82174.1|ZFY|C_lupus               CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
XP_032187800.1|ZFX_like_X1|M_erminea CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
JAC06687.1|ZFY|L_africana            CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
No_accession|PREDICTED_ZFY|E_caballusCLHCDHKSSNSSDLKRHIISVHTKDYPHKCEMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
XP_028935710.1|ZFY_X2|O_anatinus     CLHCDHKSSNSSDLKRHVISVHTKDYPHKCDTCDKGFHRPSELKKHAAAHRGRKLHQCRH [780]
XP_016288863.1|M_domestica           CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQCRH [780]
TKS65875.1|ZFY1|C_lucidus            CAHCDHKSSNSSDLKRHIISVHTKDYPHKCAICGKGFHRPSELKKHSLSHRTKKLHQCRH [780]
XP_028451227.1|P_flavescens          CAHCDHKSSNSSDLKRHIISVHTKDYPHKCAICGKGFHRPSELKKHSVSHRTKKLHQCRH [780]
XP_028276673.1|ZFY1_like_X1|P_ranga  CAHCDHKSSNSSDLKRHIISVHTKDYPHKCAVCGKGFHRPSELKKHSASHRTKKLHQCRH [780]
XP_010749798.1|ZFY1_X1|L_crocea      CAHCDHKSSNSSDLKRHIISVHTKDYPHKCAICGKGFHRPSELKKHSLSHRTKKLHQCRH [780]
XP_023133903.1|A_ocellaris           CAHCDHKSSNSSDLKRHIISVHTKDYPHKCAVCGKGFHRPSELKKHSVSHRTKKLHQCRH [780]
XP_024253620.1|O_tshawytscha         CAHCEHKSSNSSDLKRHIISVHTKDYPHKCAVCDKGFHRPSELKKHSASHRAKKLHQCRH [780]
XP_023843891.1|ZFY1|S_alpinus        CAQCEHKSSNSSDLKRHIISVHTKDYPHKCAVCDKGFHRPSELKKHSASHRAKKLHQCRH [780]
XP_020321060.1|ZFY1_like|O_kisutch   CAHCEHKSSNSSDLKRHIISVHTKDYPHKCAVCDKGFHRPSELKKHSASHRAKKLHQCRH [780]
XP_004564062.1|ZFY1_X1|M_zebra       CAHCDHKSSNSSDLKRHIISVHTKDYPHKCAVCGKGFHRPSELKKHSVSHRTKKVHQCRH [780]
XP_026038267.1|A_calliptera          CAHCDHKSSNSSDLKRHIISVHTKDYPHKCAVCGKGFHRPSELKKHSVSHRTKKVHQCRH [780]
XP_011609888.1|ZFY_X1|T_rubripes     CAHCDHKSSNSSDLKRHIISVHTKDYPHKCAVCGKGFHRPSELKKHSMSHRTKKLHQCRH [780]
XP_029029380.1|ZFY_X1|B_splendens    CAHCDHKSSNSSDLKRHIISVHTKDYPHKCAVCGKGFHRPSELKKHSVAHRTKKLHQCRH [780]
XP_028839070.1|D_clupeiodes          CAHCDHRSSNSSDLKRHIISVHTKDYPHKCAICGKGFHRPSELKKHSAAHRTKKLHQCRH [780]
ROL53794.1|ZFY1|A_grahami            CSHCEHRSSNSSDLKRHVISVHTKDYPHKCAICGKGFHRPSELKKHSASHKAKKLHQCRH [780]
XP_023277193.1|S_dorsalis            CAHCDHKSSNSSDLKRHIISVHTKDYPHKCAVCGKGFHRPSELKKHSVSHRTKKLHQCRH [780]
XP_008331409.1|C_semilaevis          CAHCDHKSSNSSDLKRHIISVHTKDYPHKCAVCGKGFHRPSELKKHSVAHRTKKLHQCRH [780]
XP_015127980.1|ZFY_X1|G_gallus       CLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKLHQCRH [780]
Q01611.1|ZFY1_XENLA|X_laevis         CLHCDHKSSNSSDLKRHIISVHTKDYPHKCEVCEKGFHRPSELKKHEAAHKGKKMHQCRH [780]
```

```
                                     *     *** ** **** * *                         *
NP_001356631.1|ZFY|H_sapiens         CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_009443992.1|ZFY_X1|P_troglodytes  CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
Q52V16.1|ZFY|G_gorilla               CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_014984082.1|ZFY_X1|M_mulatta      CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_033067617.1|ZFY|T_francoisi       CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_031516968.1|ZFY_X1|P_anubis       CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_008017167.1|C_sabaeus             CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_030782172.1|ZFY_X1|R_roxellana    CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_032612406.1|ZFY_X1|H_moloch       CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_035145821.1|ZFY_X2|C_jacchus      CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
P10925.3|ZFY1_MOUSE|M_musculus       CDFNSPDPFLLSHHILSAHTKNVPFKCK-------------------------------- [840]
P20662.2|ZFY2_MOUSE|M_musculus       CDFKSPDPFLLSHHILSAHTKNVPFKCK-------------------------------- [840]
XP_008771898.1|R_norvegicus          CDFKSPDPFLLSRHILSVHTKNVPFKCK-------------------------------- [840]
XP_015343506.1|M_marmota             CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
Q95LI3.1|ZFY|B_taurus                CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_010855418.1|B_bison               CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_017900383.1|C_hircus              CDFKIADPFVLSRHILSIHTKDLPFRCK-------------------------------- [840]
AMY96563.1|ZFY|C_elaphus             CDFRIADPFVLSRHILSVHAKDLPFRCK-------------------------------- [840]
XP_020759307.1|ZFY_X1|O_virginianus  CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
F1SPY3|ZFY|S_scrofa                  CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_024612082.1|ZFY|N_asiaeorientalis CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
AKI82174.1|ZFY|C_lupus               CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_032187800.1|ZFX_like_X1|M_erminea CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
JAC06687.1|ZFY|L_africana            CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
No_accession|PREDICTED_ZFY|E_caballusCDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_028935710.1|ZFY_X2|O_anatinus     CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
XP_016288863.1|M_domestica           CDFKIADPFVLSRHILSVHTKDLPFRCK-------------------------------- [840]
TKS65875.1|ZFY1|C_lucidus            CNFKIADPFVLSRHILSVHTKEQQASPE---KSEAKRTETHTPVATPKKSGPSGSSSSGP [840]
XP_028451227.1|P_flavescens          CNFKIADPFVLSRHILSVHTKEQQASPE---KSESKRTETHAPVVTPKKSAPSASSASGP [840]
XP_028276673.1|ZFY1_like_X1|P_ranga  CNFKIADPFVLSRHILSVHTKEQQASPE---KSETKRTETHTPAVTPKKSAPSGSSSAGP [840]
XP_010749798.1|ZFY1_X1|L_crocea      CNFKIADPFVLSRHILSVHTKEQQASPE---KSEAKRTETHTPVATPKKSGPSGSSSSGP [840]
XP_023133903.1|A_ocellaris           CNFKIADPFVLSRHILSVHTKEQQASPE---KSEAKRTETHTPVATPKKSAPSGSSASGP [840]
XP_024253620.1|O_tshawytscha         CNFKIADPFVLSRHILSVHTKEQQASPE------KNGSKRTLLGPSPASASASAPVAKKQ [840]
XP_023843891.1|ZFY1|S_alpinus        CNFKIADPFVLSRHILSVHTKEQQASPE------KNGAKRTLLGSSPASASASAPVAKKQ [840]
XP_020321060.1|ZFY1_like|O_kisutch   CNFKIADPFVLSRHILSVHTKEQQASPE------KNGSKRTLLGPSPASASASAPVAKKQ [840]
XP_004564062.1|ZFY1_X1|M_zebra       CNFKIADPFVLSRHILSVHTKEQQASPE---KSETKRTETHTPAVTPKKSGPSGSSSSGP [840]
XP_026038267.1|A_calliptera          CNFKIADPFVLSRHILSVHTKEQQASPE---KSETKRTETHTPAVTPKKSGPSGSSSSGP [840]
XP_011609888.1|ZFY_X1|T_rubripes     CNYKNADPFVLSRHILSVHTKEQQASPE---KTEVKRTETHTPVATPKKAASTASGASNA [840]
XP_029029380.1|ZFY_X1|B_splendens    CNFKIADPFVLSRHILSVHTKEQQASPE---KSEPKRTETHTSVATPKKSAASGSSSSAP [840]
XP_028839070.1|D_clupeiodes          CNFKIADPFVLSRHILSVHTKEPQQQPPPQLQAQQQPAQQPAQPPPPPPQAPVEKSAPKR [840]
ROL53794.1|ZFY1|A_grahami            CNFKIADPFILSRHILSVHTKE-QQQPP------QQPTPPTSESPPPQ-APPTEKSATKR [840]
XP_023277193.1|S_dorsalis            CNFKIADPFVLSRHILSVHTKEQQASPE---KSEGKRTETHTPVVTPKKSAPSGSSGAGP [840]
XP_008331409.1|C_semilaevis          CNFKNADPFVLSRHILSVHTKEQQSPPE---KSESKRTETHSAPATPKKSAPGGSSGAAP [840]
XP_015127980.1|ZFY_X1|G_gallus       CDFKIADPFILSRHILSVHTKDLPFRCK-------------------------------- [840]
Q01611.1|ZFY1_XENLA|X_laevis         CEFHIADPFVLSRHILSVHTKELPYRCK-------------------------------- [840]

                                                               *** ** *   *******************
NP_001356631.1|ZFY|H_sapiens         --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_009443992.1|ZFY_X1|P_troglodytes  --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
Q52V16.1|ZFY|G_gorilla               --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_014984082.1|ZFY_X1|M_mulatta      --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_033067617.1|ZFY|T_francoisi       --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_031516968.1|ZFY_X1|P_anubis       --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_008017167.1|C_sabaeus             --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_030782172.1|ZFY_X1|R_roxellana    --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_032612406.1|ZFY_X1|H_moloch       --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_035145821.1|ZFY_X2|C_jacchus      --RCRKGFNQQNELKKHMKTHSGRKVYQCEYCEYNTTDASGFKRHVISIHTKDYPHRCEY [900]
P10925.3|ZFY1_MOUSE|M_musculus       --RCKKEFQQQCELQTHMKTHSSRKVYQCEYCEYSTKDASGFKRHVISIHTKDYPHSCDF [900]
P20662.2|ZFY2_MOUSE|M_musculus       --RCKKEFQQQCELQTHMKTHSSRKVYQCEYCEYSTKDASGFKRHVISIHTKDYPHRCDF [900]
XP_008771898.1|R_norvegicus          --RCKKGFRQQCELQKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_015343506.1|M_marmota             --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
Q95LI3.1|ZFY|B_taurus                --RCKKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_010855418.1|B_bison               --RCKKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_017900383.1|C_hircus              --RCKKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
AMY96563.1|ZFY|C_elaphus             --RCKKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_020759307.1|ZFY_X1|O_virginianus  --RCKKGFRQQNDLKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
F1SPY3|ZFY|S_scrofa                  --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_024612082.1|ZFY|N_asiaeorientalis --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
AKI82174.1|ZFY|C_lupus               --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRSS- [900]
XP_032187800.1|ZFX_like_X1|M_erminea --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
JAC06687.1|ZFY|L_africana            --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEH [900]
No_accession|PREDICTED_ZFY|E_caballus--RCRKGFRQQTELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
XP_028935710.1|ZFY_X2|O_anatinus     --RCRKGFRQQGELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCDF [900]
XP_016288863.1|M_domestica           --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
TKS65875.1|ZFY1|C_lucidus            PGRVSAASLASSVTVVIGKGQKERRIYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEI [900]
XP_028451227.1|P_flavescens          PGRVSAASLASSVTVVIGKGQKERRIYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEI [900]
XP_028276673.1|ZFY1_like_X1|P_ranga  PARVSAASLASSVTVVIGKGQKERRIYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEI [900]
XP_010749798.1|ZFY1_X1|L_crocea      PGRVSAASLASSVTVVIGKGQKERRIYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEI [900]
XP_023133903.1|A_ocellaris           PARVSAASLASSVTVVIGKGQKERRIYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEI [900]
XP_024253620.1|O_tshawytscha         VLVPGASSS----AAGLATGPRERRVYQCQYCDYSSGDASGFKRHVISIHTKDYPHRCEI [900]
XP_023843891.1|ZFY1|S_alpinus        VLVPGASSS----AAGLATGPRERRVYQCQYCDYSSGDASGFKRHVISIHTKDYPHRCEI [900]
XP_020321060.1|ZFY1_like|O_kisutch   VLVPGASSS----AAGLATGPRERRVYQCQYCDYSSGDASGFKRHVISIHTKDYPHRCEI [900]
XP_004564062.1|ZFY1_X1|M_zebra       PARVSAASLASSVTVVIGKGQKERRIYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEI [900]
XP_026038267.1|A_calliptera          PARVSAASLASSVTVVIGKGQKERRIYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEI [900]
XP_011609888.1|ZFY_X1|T_rubripes     KGRVSAASLASSVTVVIGKGQKERRIYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEI [900]
XP_029029380.1|ZFY_X1|B_splendens    PARVSAASLASSVTVVIGKGQKERRIYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEI [900]
XP_028839070.1|D_clupeiodes          TLGGGAAGKKGGGGGGGAKGHRERRVYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEF [900]
ROL53794.1|ZFY1|A_grahami            TIGAAQLAAPSVKKAGGSKGPRERRVYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCQY [900]
XP_023277193.1|S_dorsalis            PARVSAASLASSVTVVIGKGQKERRIYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEI [900]
XP_008331409.1|C_semilaevis          AVRVSAASLASSVTVVVGKGQKERRIYQCQYCDYSTGDASGFKRHVISIHTKDYPHRCEI [900]
XP_015127980.1|ZFY_X1|G_gallus       --RCRKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEY [900]
Q01611.1|ZFY1_XENLA|X_laevis         --RCKKGFRQQIELKKHMKTHSGKKVYQCEYCEYNTTDASGFKRHVISIHTKDYPHRCDY [900]
```

```
                                     * ************   ***
NP_001356631.1|ZFY|H_sapiens         CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
XP_009443992.1|ZFY_X1|P_troglodytes  CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
Q52V16.1|ZFY|G_gorilla               CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
XP_014984082.1|ZFY_X1|M_mulatta      CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
XP_033067617.1|ZFY|T_francoisi       CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
XP_031516968.1|ZFY_X1|P_anubis       CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
XP_008017167.1|C_sabaeus             CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
XP_030782172.1|ZFY_X1|R_roxellana    CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
XP_032612406.1|ZFY_X1|H_moloch       CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
XP_035145821.1|ZFY_X2|C_jacchus      CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
P10925.3|ZFY1_MOUSE|M_musculus       CKKGFRRPSEKNQHIMRHHK-VGLP- [926]
P20662.2|ZFY2_MOUSE|M_musculus       CKKGFRRPSEKNQHIMRHHKEVGLA- [926]
XP_008771898.1|R_norvegicus          CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
XP_015343506.1|M_marmota             CKKGFRRPSEKNQHIMRHHKEVSLS- [926]
Q95LI3.1|ZFY|B_taurus                CKKGFRRPSEKNQHITRHHKEVGLP- [926]
XP_010855418.1|B_bison               CKKGFRRPSEKNQHITRHHKEVGLP- [926]
XP_017900383.1|C_hircus              CKKGFRRPSEKNQHITRHHKEVGLP- [926]
AMY96563.1|ZFY|C_elaphus             CKKGFRRPSEKNQHITRHHKEVGLP- [926]
XP_020759307.1|ZFY_X1|O_virginianus  CKKGFRRPSEKNQHITRHHKEVGLP- [926]
F1SPY3|ZFY|S_scrofa                  CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
XP_024612082.1|ZFY|N_asiaeorientalis CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
AKI82174.1|ZFY|C_lupus               -------------------------- [926]
XP_032187800.1|ZFX_like_X1|M_erminea CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
JAC06687.1|ZFY|L_africana            CKKGFRRPSEKNQHIMRHHKEVGLP- [926]
No_accession|PREDICTED_ZFY|E_caballusCKKGFRRPSEKNQHIMRHHKEVGLP- [926]
XP_028935710.1|ZFY_X2|O_anatinus     CKKGFRRPSEKNQHIMRHHKDLGLP- [926]
XP_016288863.1|M_domestica           CKKGFRRPSEKNQHIMRHHKDVGLP- [926]
TKS65875.1|ZFY1|C_lucidus            CSKGFRRPSEKNQHIMRHHKDVVQAE [926]
XP_028451227.1|P_flavescens          CSKGFRRPSEKNQHIMRHHKDVVQAD [926]
XP_028276673.1|ZFY1_like_X1|P_ranga  CSKGFRRPSEKNQHIMRHHKDVVQAE [926]
XP_010749798.1|ZFY1_X1|L_crocea      CSKGFRRPSEKNQHIMRHHKDVVQAE [926]
XP_023133903.1|A_ocellaris           CSKGFRRPSEKNQHIMRHHKDVVQAE [926]
XP_024253620.1|O_tshawytscha         CSKGFRRPSEKNQHIARHHKDLVQAE [926]
XP_023843891.1|ZFY1|S_alpinus        CSKGFRRPSEKNQHIARHHKDLVQAE [926]
XP_020321060.1|ZFY1_like|O_kisutch   CSKGFRRPSEKNQHIARHHKDLVQAE [926]
XP_004564062.1|ZFY1_X1|M_zebra       CSKGFRRPSEKNQHIMRHHKDVVQAE [926]
XP_026038267.1|A_calliptera          CSKGFRRPSEKNQHIMRHHKDVVQAE [926]
XP_011609888.1|ZFY_X1|T_rubripes     CSKGFRRPSEKNQHIMRHHKDVVQTD [926]
XP_029029380.1|ZFY_X1|B_splendens    CSKGFRRPSEKNQHIMRHHKDVVQTE [926]
XP_028839070.1|D_clupeiodes          CSKGFRRPSEKNQHIMRHHKDMVQAE [926]
ROL53794.1|ZFY1|A_grahami            CSKGFRRPSEKNQHIMRHHKDIVPAE [926]
XP_023277193.1|S_dorsalis            CSKGFRRPSEKNQHIMRHHKDVVQAE [926]
XP_008331409.1|C_semilaevis          CSKGFRRPSEKNQHIMRHHKDVVQTE [926]
XP_015127980.1|ZFY_X1|G_gallus       CKKGFRRPSEKNQHIMRHHKDVGLP- [926]
Q01611.1|ZFY1_XENLA|X_laevis         CKKGFRRPSEKNQHTLKHHKEASLM- [926]
```

Table 9. Multiple sequence alignment of the primary sequence of vertebrate ZFY. The ZFY primary sequences were aligned using the MEGA-X programme's ClustalW alignment tool and used to analyse percentage identities. The conserved amino acids are represented by the YELLOW highlight and the (*) symbol above the conserved amino acid(s). Where there was no conservation, there are no symbols or highlights present. Each exon has exon colour coding with RED representing Exon 1, GREEN representing Exon 2, CYAN representing Exon 3, MAGENTA representing Exon 4, BROWN representing Exon 5, PURPLE representing Exon 6 and GREY representing Exon 7. The BLACK represents amino acids across a splice junction The nuclear localization sequence (NLS) is labelled in **UPPERCASE BOLD** text between residues 449-464. The polyalanine motif is represented by **LOWERCASE BOLD** alanine residues .

Vertebrate Exons	Percentage Identity (%) of sequence alignment	Number of conserved sites of sequence alignment	Number of Parsimony informative sites of sequence alignment
Exon 1	8.7	2	17
Exon 2	15.7	33	134
Exon 3	20.4	10	34
Exon 4	10.6	5	41
Exon 5	30.8	16	33
Exon 6	16.3	13	31
Exon 7	44.0	202	192

Table 10. Table demonstrating the percentage identity, conserved sites, and parsimony informative site of the vertebrate species. The table shows the percentage identity, calculated using the conserved sites and the total number of sites in aligned sequence (by exon).

Phylogenetic tree (**Figure 8**) analysis showed primate ZFY proteins were closely related, more specifically *H. sapiens* and *P. troglodytes* as they had an identical branch length. **Figure 8** also showed that carnivores and artiodactyls were closely related to each other. However, rodents had unusual phylogeny as they were phylogenetically discordant as *M. musculus* and *R. norvegicus* were expected to be grouped with *M. marmota* but instead were dispersed in between the artiodactyls and monotremes. Furthermore, the rest of the mammalian, bird and amphibian species indicated that they were distantly related to the primates as the branch lengths progressively get longer. The outgroup species (fish species) were distantly related to the land vertebrates, indicating evolutionary events possibly occurred resulting in the long branches. Therefore, the next step of the sequence analysis was investigating the land vertebrates and excluding fish species as they were very distantly related to the land vertebrates.

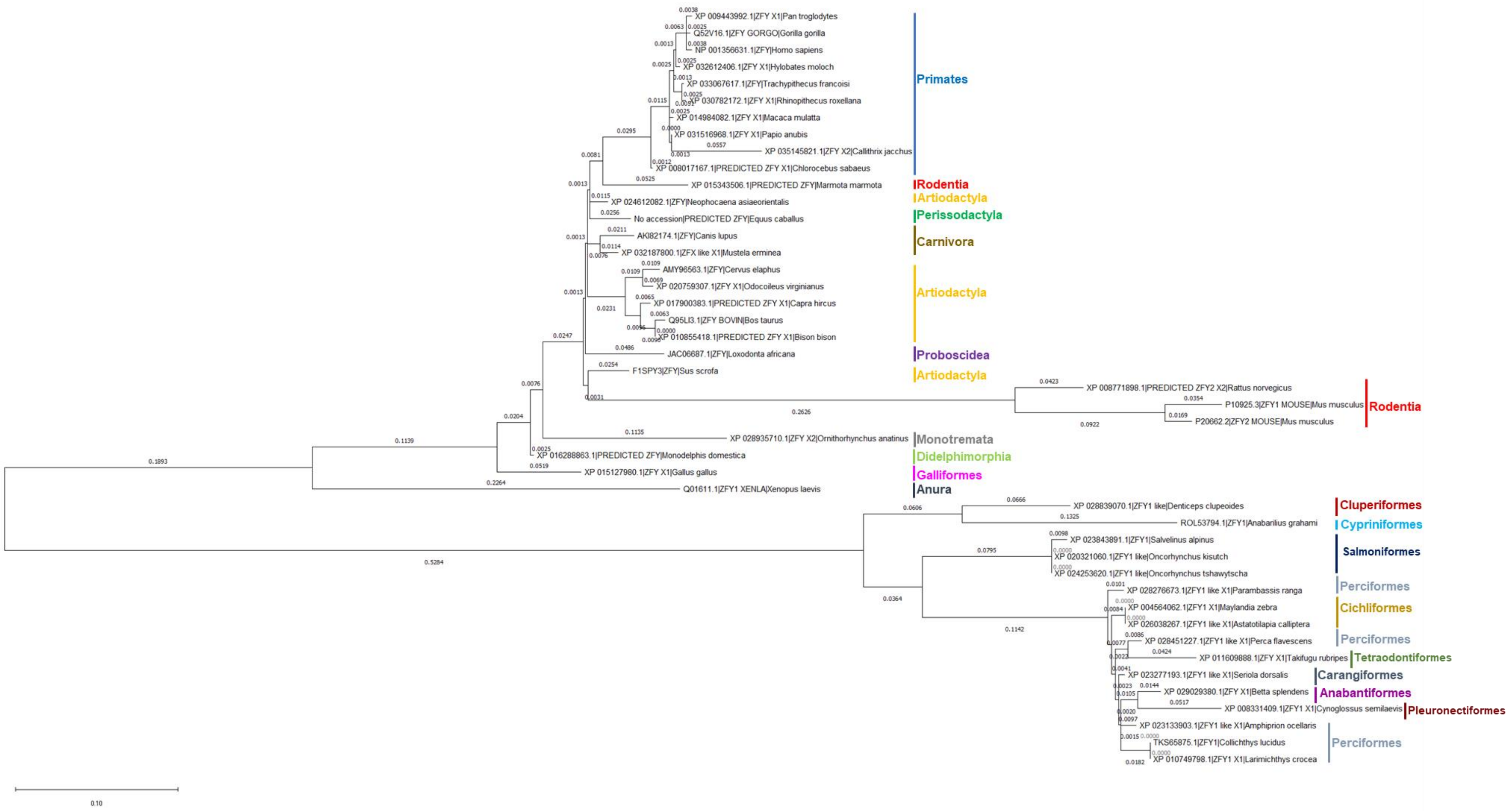
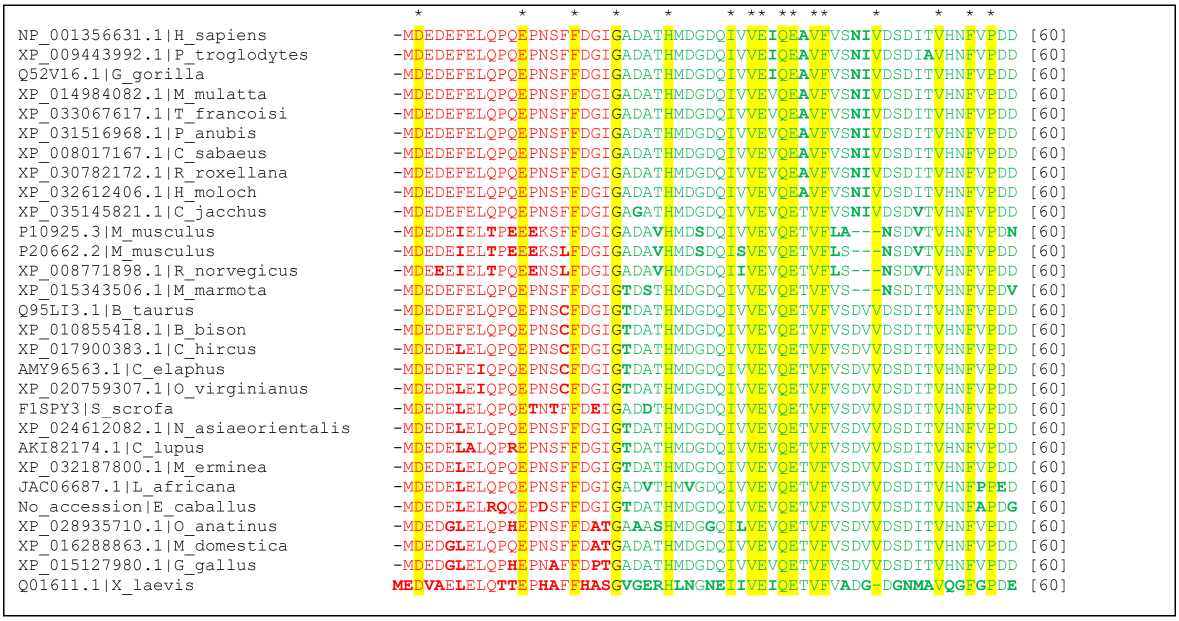

0.0038
XP 009443992.1|ZFY X1|Pan troglodytes
0.0063
0.0025
Q52V16.1|ZFY GORGO|Gorilla gorilla
0.0013
0.0038
NP 001356631.1|ZFY|Homo sapiens
0.0025
0.0025
XP 032612406.1|ZFY X1|Hylobates moloch
0.0013
XP 033067617.1|ZFY|Trachypithecus francoisi
0.0025
0.0115
0.0051
XP 030782172.1|ZFY X1|Rhinopithecus roxellana
0.0025
XP 014984082.1|ZFY X1|Macaca mulatta
0.0295
0.0000
XP 031516968.1|ZFY X1|Papio anubis
0.0557
XP 035145821.1|ZFY X2|Callithrix jacchus
0.0081
0.0013
0.0012
XP 008017167.1|PREDICTED ZFY X1|Chlorocebus sabaeus
Primates
0.0525
0.0013
XP 015343506.1|PREDICTED ZFY|Marmota marmota
Rodentia
0.0115
XP 024612082.1|ZFY|Neophocaena asiaeorientalis
Artiodactyla
0.0256
No accession|PREDICTED ZFY|Equus caballus
Perissodactyla
0.0013
0.0211
AKI82174.1|ZFY|Canis lupus
0.0114
0.0076
XP 032187800.1|ZFX like X1|Mustela erminea
Carnivora
0.0109
0.0109
AMY96563.1|ZFY|Cervus elaphus
0.0069
XP 020759307.1|ZFY X1|Odocoileus virginianus
0.0013
0.0065
XP 017900383.1|PREDICTED ZFY X1|Capra hircus
0.0231
0.0063
Q95LI3.1|ZFY BOVIN|Bos taurus
0.0096
0.0000
0.0098
XP 010855418.1|PREDICTED ZFY X1|Bison bison
Artiodactyla
0.0247
0.0486
JAC06687.1|ZFY|Loxodonta africana
Proboscidea
0.0254
F1SPY3|ZFY|Sus scrofa
Artiodactyla
0.0076
0.0031
0.0423
XP 008771898.1|PREDICTED ZFY2 X2|Rattus norvegicus
0.0354
P10925.3|ZFY1 MOUSE|Mus musculus
0.2626
0.0169
P20662.2|ZFY2 MOUSE|Mus musculus
0.0922
Rodentia
0.0204
0.1135
XP 028935710.1|ZFY X2|Ornithorhynchus anatinus
Monotremata
0.1139
0.0025
XP 016288863.1|PREDICTED ZFY|Monodelphis domestica
Didelphimorphia
0.1893
0.0519
XP 015127980.1|ZFY X1|Gallus gallus
Galliformes
0.2264
Q01611.1|ZFY1 XENLA|Xenopus laevis
Anura
0.0666
XP 028839070.1|ZFY1 like|Denticeps clupeoides
Cluperiformes
0.0606
0.1325
ROL53794.1|ZFY1|Anabarilius grahami
Cypriniformes
0.0098
XP 023843891.1|ZFY1|Salvelinus alpinus
0.0795
0.0000
XP 020321060.1|ZFY1 like|Oncorhynchus kisutch
0.5284
0.0000
XP 024253620.1|ZFY1 like|Oncorhynchus tshawytscha
Salmoniformes
0.0364
0.0101
XP 028276673.1|ZFY1 like X1|Parambassis ranga
Perciformes
0.0000
0.0084
XP 004564062.1|ZFY1 X1|Maylandia zebra
0.0000
XP 026038267.1|ZFY1 like X1|Astatotilapia calliptera
Cichliformes
0.1142
0.0086
0.0077
XP 028451227.1|ZFY1 like X1|Perca flavescens
Perciformes
0.0424
0.0023
XP 011609888.1|ZFY X1|Takifugu rubripes
Tetraodontiformes
0.0041
XP 023277193.1|ZFY1 like X1|Seriola dorsalis
Carangiformes
0.0023
0.0144
0.0105
XP 029029380.1|ZFY X1|Betta splendens
Anabantiformes
0.0517
0.0020
XP 008331409.1|ZFY1 X1|Cynoglossus semilaevis
Pleuronectiformes
0.0097
XP 023133903.1|ZFY1 like X1|Amphiprion ocellaris
0.0015
0.0000
TKS65875.1|ZFY1|Collichthys lucidus
0.0000
0.0182
XP 010749798.1|ZFY1 X1|Larimichthys crocea
Perciformes
0.10

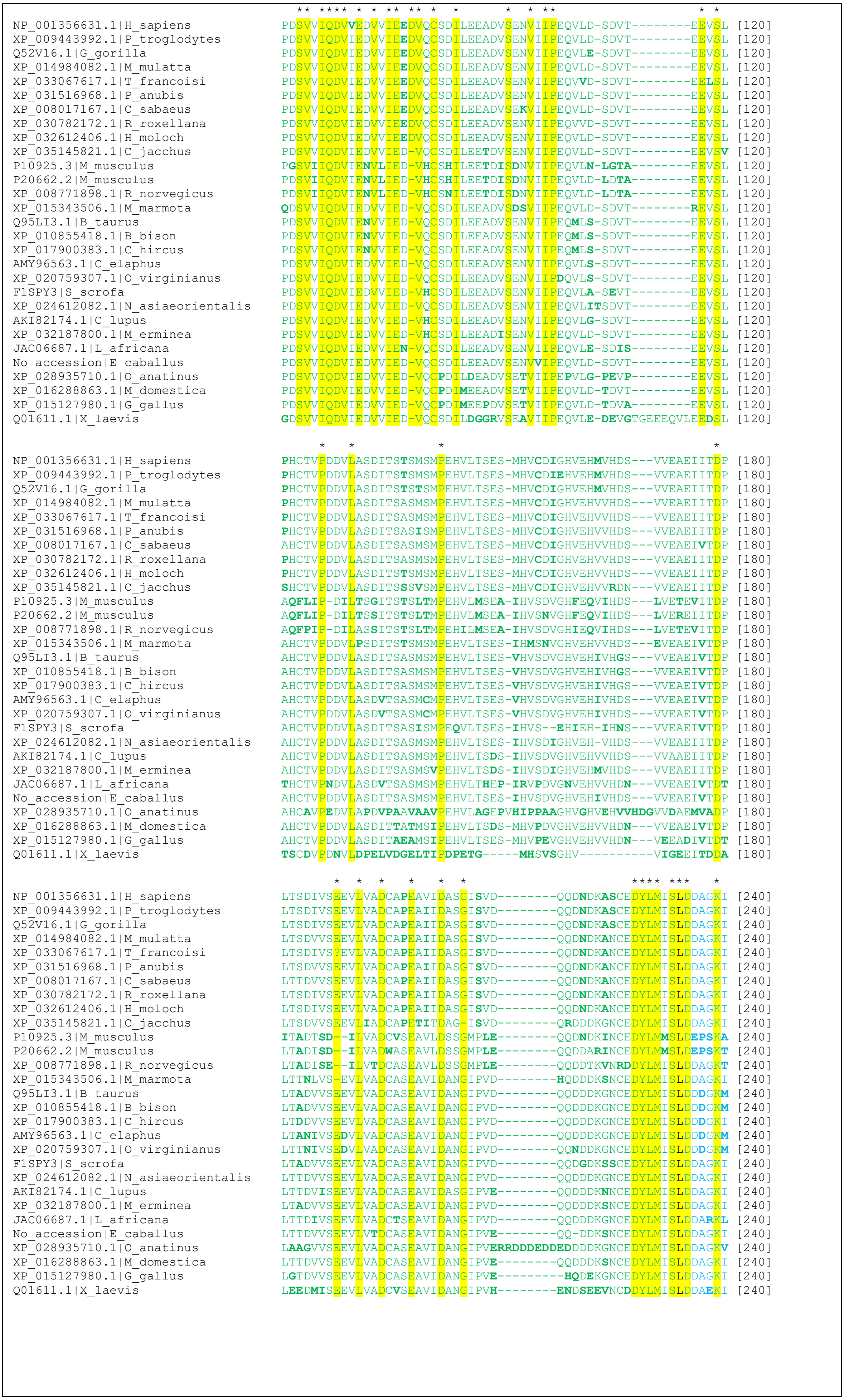
Figure 8. Maximum likelihood (Bootstrap) tree of the ZFY and ZFX multiple sequence alignment shown in Table 12. The tree inferred by Maximum Likelihood method and JTT matrix-based model shows relationships of the most closely related ZFY species and the most distantly related ZFY species. The species were grouped taxonomically with placental mammalian species all grouped together along with marsupials, monotremes and amphibians and the outgroup species were the fish. The branches were grouped into the orders as indicated by the various coloured text and labelling. The scale bar corresponds to 0.10 estimated amino acid substitutions per site.

3.3 Land vertebrates ZFY protein alignment and phylogeny

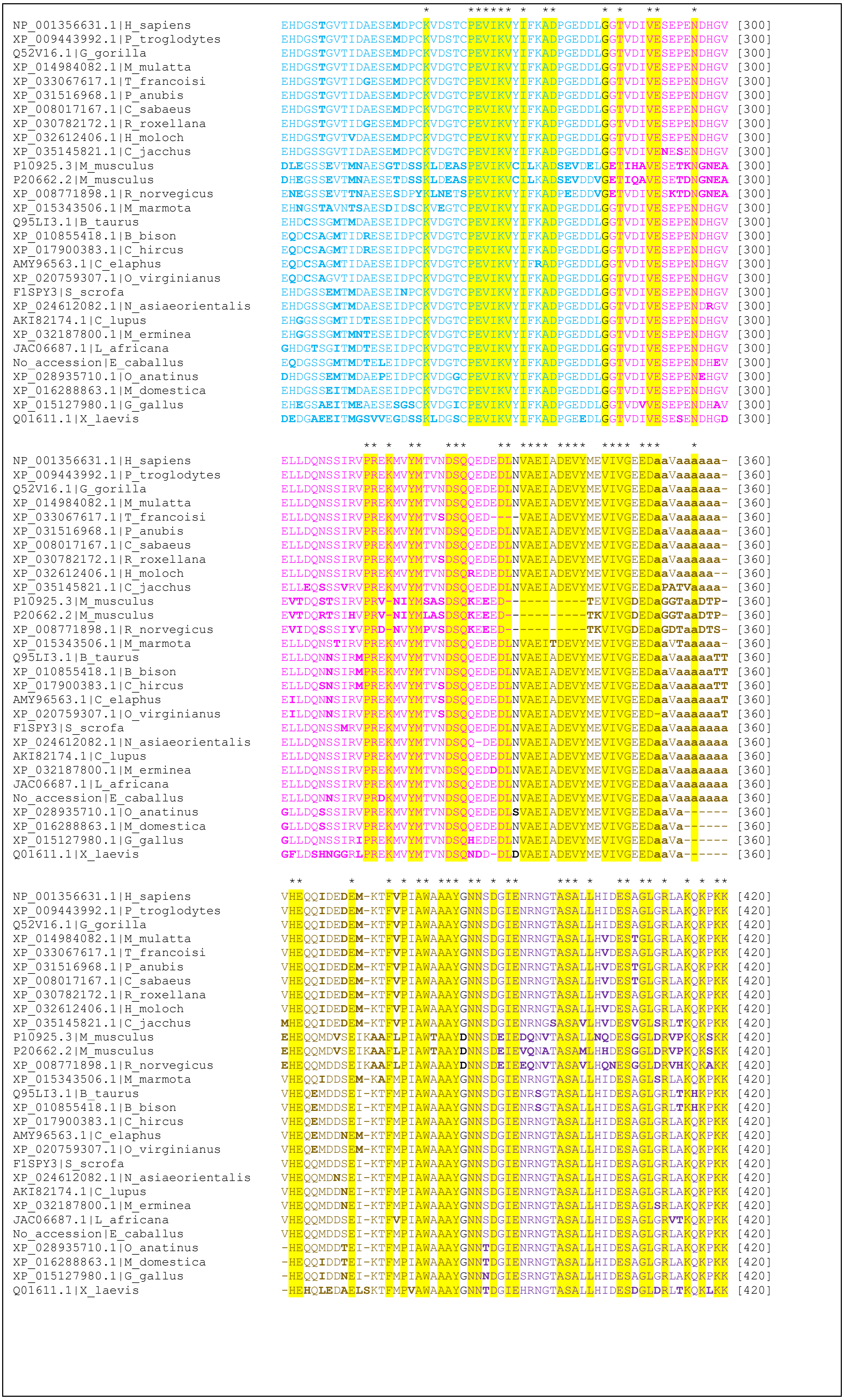
When the alignments were performed on land vertebrates alone (and *N. asiaeorientalis*), the number of conserved amino acids increased from 284 (as mentioned in sections before) to 375 amino acids (371 within the exons and 4 splice junction residues conserved), which showed that the fish ZFY proteins were distantly related as the majority of the differences arose from fish species. Firstly, the alignment showed that there was an increase in the percentage identity of the ZFY protein of land vertebrates and Cetartiodactyla shown by **Table 11** and **Table 12**. For exon 2, there was roughly 7% increase in comparison to when the fish species were included in the alignment (**Table 11**), and for exon 3 there was a 4% increase. In addition, the increase in conservation became more significant as the increase was approximately 18% for exon 4, 20% for exon 5 and 21% exon 6, which highlighted the high divergence of these exons in fish species. Lastly, exon 7 had a 17% increase in percentage identity, and this outlined that exon 7 possibly had a very significant role as the conservation is very high.



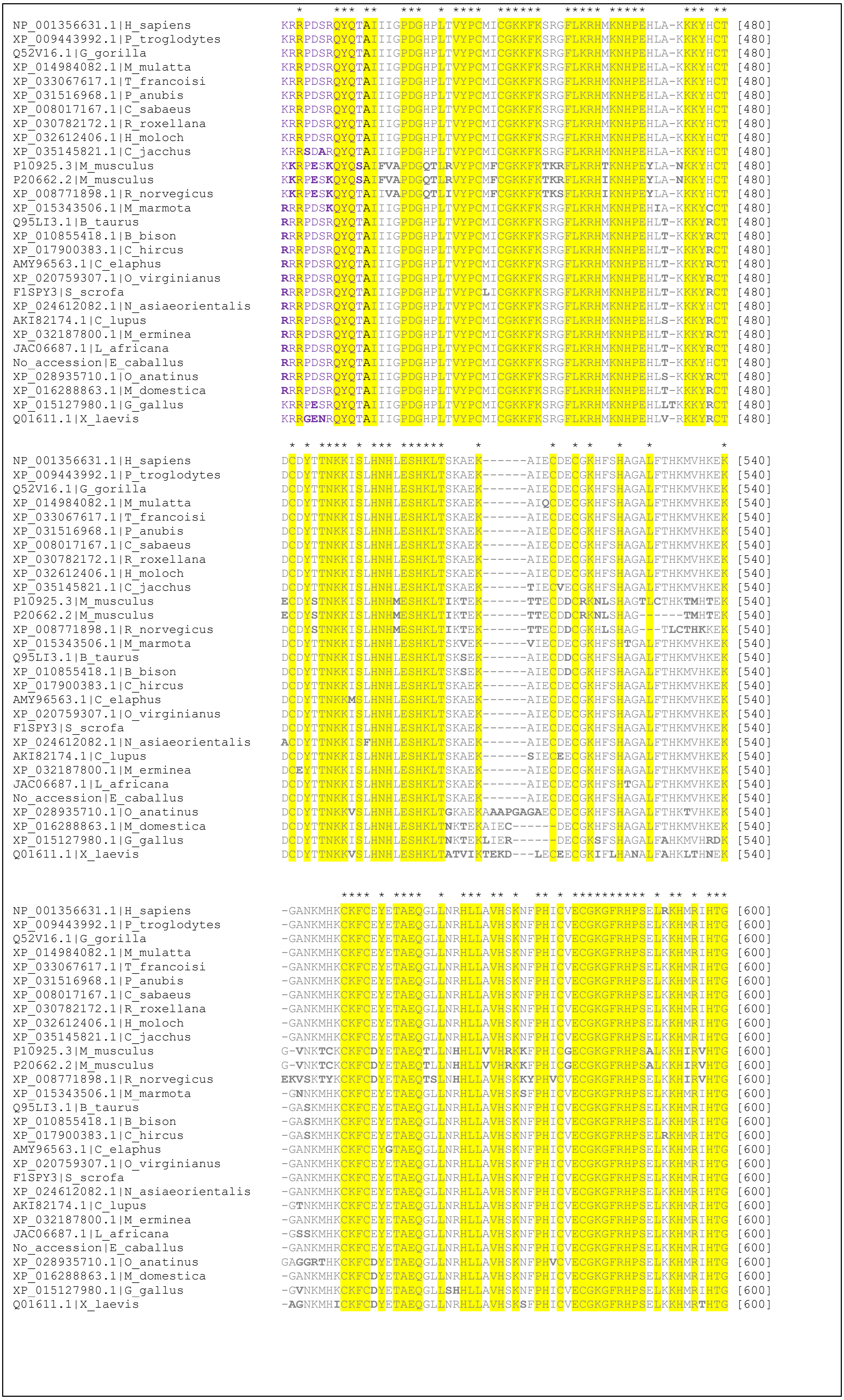
```
                                    *         *    *   *    *     * ** ** **    *      *  * *
NP_001356631.1|H_sapiens          -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEIQEAVFVSNIVDSDITVHNFVPDD [60]
XP_009443992.1|P_troglodytes      -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEIQEAVFVSNIVDSDIAVHNFVPDD [60]
Q52V16.1|G_gorilla                -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEIQEAVFVSNIVDSDITVHNFVPDD [60]
XP_014984082.1|M_mulatta          -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_033067617.1|T_francoisi        -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_031516968.1|P_anubis           -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_008017167.1|C_sabaeus          -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_030782172.1|R_roxellana        -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_032612406.1|H_moloch           -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_035145821.1|C_jacchus          -MDEDEFELQPQEPNSFFDGIGAGATHMDGDQIVVEVQETVFVSNIVDSDVTVHNFVPDD [60]
P10925.3|M_musculus               -MDEDEIELTPEEEKSFFDGIGADAVHMDSDQIVVEVQETVFLA---NSDVTVHNFVPDN [60]
P20662.2|M_musculus               -MDEDEIELTPEEEKSLFDGIGADAVHMDSDQISVEVQETVFLS---NSDVTVHNFVPDD [60]
XP_008771898.1|R_norvegicus       -MDEEEIELTPQEENSLFDGIGADAVHMDGDQIIVEVQETVFLS---NSDVTVHNFVPDD [60]
XP_015343506.1|M_marmota          -MDEDEFELQPQEPNSFFDGIGTDSTHMDGDQIVVEVQETVFVS---NSDITVHNFVPDV [60]
Q95LI3.1|B_taurus                 -MDEDEFELQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_010855418.1|B_bison            -MDEDEFELQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_017900383.1|C_hircus           -MDEDELELQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
AMY96563.1|C_elaphus              -MDEDEFEIQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_020759307.1|O_virginianus      -MDEDELEIQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
F1SPY3|S_scrofa                   -MDEDELELQPQETNTFFDEIGADDTHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_024612082.1|N_asiaeorientalis  -MDEDELELQPQEPNSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
AKI82174.1|C_lupus                -MDEDELALQPREPNSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_032187800.1|M_erminea          -MDEDELELQPQEPNSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
JAC06687.1|L_africana             -MDEDELELQPQEPNSFFDGIGADVTHMVGDQIVVEVQETVFVSDVVDSDITVHNFPPED [60]
No_accession|E_caballus           -MDEDELELRQQEPDSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFAPDG [60]
XP_028935710.1|O_anatinus         -MDEDGLELQPHEPNSFFDATGAAASHMDGGQILVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_016288863.1|M_domestica        -MDEDGLELQPQEPNSFFDATGADATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_015127980.1|G_gallus           -MDEDGLELQPHEPNAFFDPTGADATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
Q01611.1|X_laevis                 MEDVAELELQTTEPHAFFHASGVGERHLNGNEIIVEIQETVFVADG-DGNMAVQGFGPDE [60]
```



```
                                      ** **** * * ** ** *  *      *  * **                   * *
NP_001356631.1|H_sapiens            PDSVVIQDVVEDVVIEEDVQCSDILEEADVSENVIIPEQVLD-SDVT--------EEVSL [120]
XP_009443992.1|P_troglodytes        PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD-SDVT--------EEVSL [120]
Q52V16.1|G_gorilla                  PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLE-SDVT--------EEVSL [120]
XP_014984082.1|M_mulatta            PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD-SDVT--------EEVSL [120]
XP_033067617.1|T_francoisi          PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVVD-SDVT--------EELSL [120]
XP_031516968.1|P_anubis             PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD-SDVT--------EEVSL [120]
XP_008017167.1|C_sabaeus            PDSVVIQDVIEDVVIEEDVQCSDILEEADVSEKVIIPEQVLD-SDVT--------EEVSL [120]
XP_030782172.1|R_roxellana          PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVVD-SDVT--------EEVSL [120]
XP_032612406.1|H_moloch             PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD-SDVT--------EEVSL [120]
XP_035145821.1|C_jacchus            PDSVVIQDVIEDVVIED-VQCSDILEETDVSENVIIPEQVLD-SDVT--------EEVSV [120]
P10925.3|M_musculus                 PGSVIIQDVIENVLIED-VHCSHILEETDISDNVIIPEQVLN-LGTA--------EEVSL [120]
P20662.2|M_musculus                 PDSVIIQDVIENVLIED-VHCSHILEETDISDNVIIPEQVLD-LDTA--------EEVSL [120]
XP_008771898.1|R_norvegicus         PDSVIIQDVIENVLIED-VHCSNILEETDISDNVIIPEQVLD-LDTA--------EEVSL [120]
XP_015343506.1|M_marmota            QDSVVIQDVIEDVVIED-VQCSDILEEADVSDSVIIPEQVLD-SDVT--------REVSL [120]
Q95LI3.1|B_taurus                   PDSVVIQDVIENVVIED-VQCSDILEEADVSENVIIPEQMLS-SDVT--------EEVSL [120]
XP_010855418.1|B_bison              PDSVVIQDVIENVVIED-VQCSDILEEADVSENVIIPEQMLS-SDVT--------EEVSL [120]
XP_017900383.1|C_hircus             PDSVVIQDVIENVVIED-VQCSDILEEADVSENVIIPEQMLS-SDVT--------EEVSL [120]
AMY96563.1|C_elaphus                PDSVVIQDVIEDVVIED-VQCSDILEEADVSENVIIPEQVLS-SDVT--------EEVSL [120]
XP_020759307.1|O_virginianus        PDSVVIQDVIEDVVIED-VQCSDILEEADVSENVIIPDQVLS-SDVT--------EEVSL [120]
F1SPY3|S_scrofa                     PDSVVIQDVIEDVVIED-VHCSDILEEADVSENVIIPEQVLA-SEVT--------EEVSL [120]
XP_024612082.1|N_asiaeorientalis    PDSVVIQDVIEDVVIED-VQCSDILEEADVSENVIIPEQVLITSDVT--------EEVSL [120]
AKI82174.1|C_lupus                  PDSVVIQDVIEDVVIED-VHCSDILEEADVSENVIIPEQVLG-SDVT--------EEVSL [120]
XP_032187800.1|M_erminea            PDSVVIQDVIEDVVIED-VHCSDILEEADISENVIIPEQVLD-SDVT--------EEVSL [120]
JAC06687.1|L_africana               PDSVVIQDVIEDVVIEN-VQCSDILEEADVSENVIIPEQVLE-SDIS--------EEVSL [120]
No_accession|E_caballus             PDSVVIQDVIEDVVIED-VQCSDILEEADVSENVVIPEQVLD-SDVT--------EEVSL [120]
XP_028935710.1|O_anatinus           PDSVVIQDVIEDVVIED-VQCPDILDEADVSETVIIPEPVLG-PEVP--------EEVSL [120]
XP_016288863.1|M_domestica          PDSVVIQDVIEDVVIED-VQCPDIMEEADVSETVIIPEQVLD-TDVT--------EEVSL [120]
XP_015127980.1|G_gallus             PDSVVIQDVIEDVVIED-VQCPDIMEEPDVSETVIIPEQVLD-TDVA--------EEVSL [120]
Q01611.1|X_laevis                   GDSVVIQDVIEDVVIED-VQCSDILDGGRVSEAVIIPEQVLE-DEVGTGEEEQVLEEDSL [120]

                                         *   *           *                                    *
NP_001356631.1|H_sapiens            PHCTVPDDVLASDITSTSMSMPEHVLTSES-MHVCDIGHVEHMVHDS---VVEAEIITDP [180]
XP_009443992.1|P_troglodytes        PHCTVPDDVLASDITSTSMSMPEHVLTSES-MHVCDIEHVEHMVHDS---VVEAEIITDP [180]
Q52V16.1|G_gorilla                  PHCTVPDDVLASDITSTSTSMPEHVLTSES-MHVCDIGHVEHMVHDS---VVEAEIITDP [180]
XP_014984082.1|M_mulatta            PHCTVPDDVLASDITSASMSMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIITDP [180]
XP_033067617.1|T_francoisi          PHCTVPDDVLASDITSASMSMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIITDP [180]
XP_031516968.1|P_anubis             PHCTVPDDVLASDITSASISMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIITDP [180]
XP_008017167.1|C_sabaeus            AHCTVPDDVLASDITSASMSMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIVTDP [180]
XP_030782172.1|R_roxellana          PHCTVPDDVLASDITSASMSMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIITDP [180]
XP_032612406.1|H_moloch             PHCTVPDDVLASDITSTSMSMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIITDP [180]
XP_035145821.1|C_jacchus            SHCTVPDDVLASDITSSSVSMPEHVLTSES-MHVCDIGHVEHVVRDN---VVEAEIITDP [180]
P10925.3|M_musculus                 AQFLIP-DILTSGITSTSLTMPEHVLMSEA-IHVSDVGHFEQVIHDS---LVETEVITDP [180]
P20662.2|M_musculus                 AQFLIP-DILTSSITSTSLTMPEHVLMSEA-IHVSNVGHFEQVIHDS---LVEREIITDP [180]
XP_008771898.1|R_norvegicus         AQFPIP-DILASSITSTSLTMPEHILMSEA-IHVSDVGHIEQVIHDS---LVETEVITDP [180]
XP_015343506.1|M_marmota            AHCTVPDDVLPSDITSTSMSMPEHVLTSES-IHMSNVGHVEHVVHDS---EVEAEIVTDP [180]
Q95LI3.1|B_taurus                   AHCTVPDDVLASDITSASMSMPEHVLTSES-VHVSDVGHVEHIVHGS---VVEAEIVTDP [180]
XP_010855418.1|B_bison              AHCTVPDDVLASDITSASMSMPEHVLTSES-VHVSDVGHVEHIVHGS---VVEAEIVTDP [180]
XP_017900383.1|C_hircus             AHCTVPDDVLASDITSASMSMPEHVLTSES-VHVSDVGHVEHIVHGS---VVEAEIVTDP [180]
AMY96563.1|C_elaphus                AHCTVPDDVLASDVTSASMCMPEHVLTSES-VHVSDVGHVEHIVHDS---VVEAEIVTDP [180]
XP_020759307.1|O_virginianus        AHCTVPDDVLASDVTSASMCMPEHVLTSES-VHVSDVGHVEHIVHDS---VVEAEIVTDP [180]
F1SPY3|S_scrofa                     AHCTVPDDVLASDITSASISMPEQVLTSES-IHVS--EHIEH-IHNS---VVEAEIVTDP [180]
XP_024612082.1|N_asiaeorientalis    AHCTVPDDVLASDITSASMSMPEHVLTSES-IHVSDIGHVEH-VHDS---VVEAEIITDP [180]
AKI82174.1|C_lupus                  AHCTVPDDVLASDITSASMSMPEHVLTSDS-IHVSDVGHVEHVVHDS---VVAAEIITDP [180]
XP_032187800.1|M_erminea            AHCTVPDDVLASDITSASMSVPEHVLTSDS-IHVSDIGHVEHMVHDS---VVEAEIITDP [180]
JAC06687.1|L_africana               THCTVPNDVLASDVTSASMSMPEHVLTHEP-IRVPDVGNVEHVVHDN---VVEAEIVTDT [180]
No_accession|E_caballus             AHCTVPDDVLASDITSASMSMPEHVLTSES-IHVSDVGHVEHIVHDS---VVEAEIVTDP [180]
XP_028935710.1|O_anatinus           AHCAVPEDVLAPDVPAAVAAVPEHVLAGEPVHIPPAAGHVGHVEHVVHDGVVDAEMVADP [180]
XP_016288863.1|M_domestica          AHCTVPDDVLASDITTATMSIPEHVLTSDS-MHVPDVGHVEHVVHDN---VVEAEIVTDP [180]
XP_015127980.1|G_gallus             AHCTVPDDVLASDITAEAMSIPEHVLTSES-MHVPEVGHVEHVVHDN---VEEADIVTDT [180]
Q01611.1|X_laevis                   TSCDVPDNVLDPELVDGELTIPDPETG-----MHSVSGHV----------VIGEEITDDA [180]

                                           *  *  *   *   *  *                      **** ***   *
NP_001356631.1|H_sapiens            LTSDIVSEEVLVADCAPEAVIDASGISVD--------QQDNDKASCEDYLMISLDDAGKI [240]
XP_009443992.1|P_troglodytes        LTSDIVSEEVLVADCAPEAIIDASGISVD--------QQDNDKASCEDYLMISLDDAGKI [240]
Q52V16.1|G_gorilla                  LTSDIVSEEVLVADCAPEAIIDASGISVD--------QQDNDKASCEDYLMISLDDAGKI [240]
XP_014984082.1|M_mulatta            LTSDVVSEEVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_033067617.1|T_francoisi          LTSDIVS?EVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_031516968.1|P_anubis             LTSDVVSEEVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_008017167.1|C_sabaeus            LTTDVVSEEVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_030782172.1|R_roxellana          LTSDIVSEEVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_032612406.1|H_moloch             LTSDIVSEEVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_035145821.1|C_jacchus            LTSDVVSEEVLIADCAPETITDAG-ISVD--------QRDDDKGNCEDYLMISLDDAGKI [240]
P10925.3|M_musculus                 ITADTSD--ILVADCVSEAVLDSSGMPLE--------QQDNDKINCEDYLMMSLDEPSKA [240]
P20662.2|M_musculus                 LTADISD--ILVADWASEAVLDSSGMPLE--------QQDDARINCEDYLMMSLDEPSKT [240]
XP_008771898.1|R_norvegicus         LTADISE--ILVTDCASEAVLDSSGMPLE--------QQDDTKVNRDDYLMISLDDAGKT [240]
XP_015343506.1|M_marmota            LTTNLVS-EVLVADCASEAVIDANGIPVD--------HQDDDKSNCEDYLMISLDDAGKI [240]
Q95LI3.1|B_taurus                   LTADVVSEEVLVADCASEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDDGKM [240]
XP_010855418.1|B_bison              LTADVVSEEVLVADCASEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDDGKM [240]
XP_017900383.1|C_hircus             LTDDVVSEEVLVADCASEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDDGKI [240]
AMY96563.1|C_elaphus                LTANIVSEDVLVADCASEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDDGKM [240]
XP_020759307.1|O_virginianus        LTTNIVSEDVLVADCASEAVIDANGIPVD--------QQNDDKGNCEDYLMISLDDDGKM [240]
F1SPY3|S_scrofa                     LTADVVSEEVLVADCASEAVIDANGIPVD--------QQDGDKSSCEDYLMISLDDAGKI [240]
XP_024612082.1|N_asiaeorientalis    LTTDVVSEEVLVADCASEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDAGKI [240]
AKI82174.1|C_lupus                  LTTDVISEEVLVADCASEAVIDASGIPVE--------QQDDDKNNCEDYLMISLDDAGKI [240]
XP_032187800.1|M_erminea            LTADVVSEEVLVADCASEAVIDANGIPVD--------QQDDDKSNCEDYLMISLDDAGKI [240]
JAC06687.1|L_africana               LTTDIVSEEVLVADCTSEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDARKL [240]
No_accession|E_caballus             LTTDVVSEEVLVTDCASEAVIDANGIPVE--------QQ-DDKSNCEDYLMISLDDAGKI [240]
XP_028935710.1|O_anatinus           LAAGVVSEEVLVADCASEAVIDANGIPVERRDDDEDDEDDDDKGNCEDYLMISLDDAGKV [240]
XP_016288863.1|M_domestica          LTTDVVSEEVLVADCASEAVIDANGIPVE--------QQDDDKSNCEDYLMISLDDAGKI [240]
XP_015127980.1|G_gallus             LGTDVVSEEVLVADCASEAVIDANGIPVE---------HQDEKGNCEDYLMISLDDAGKI [240]
Q01611.1|X_laevis                   LEEDMISEEVLVADCVSEAVIDANGIPVH--------ENDSEEVNCDDYLMISLDDAEKI [240]
```

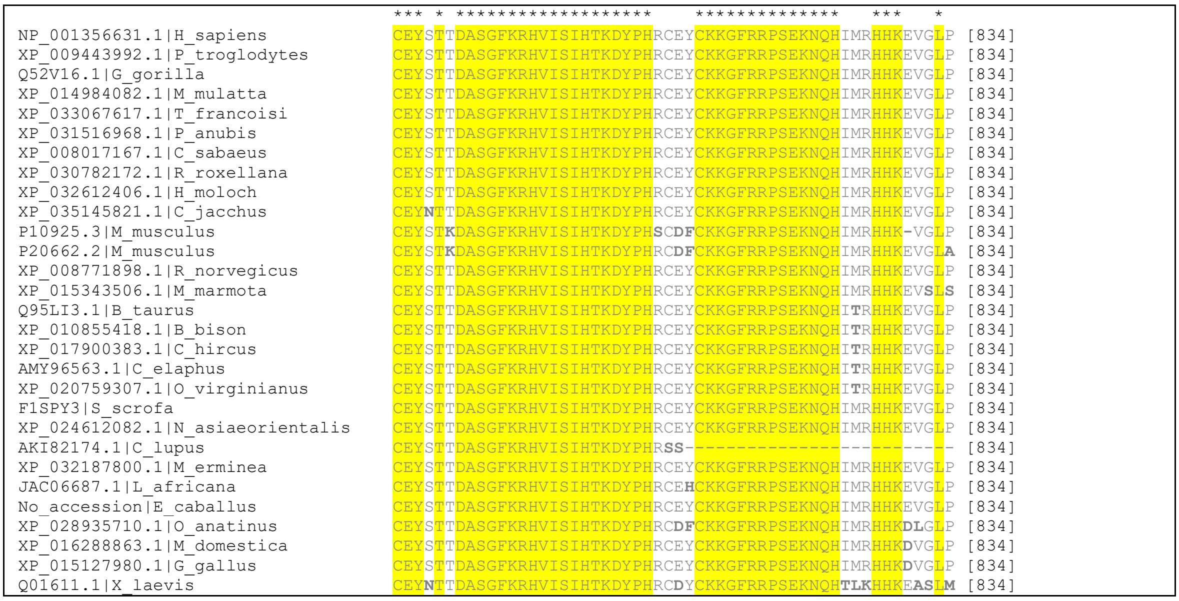


```
                                                       *     ****** *  **      * *   **    *
NP_001356631.1|H_sapiens            EHDGSTGVTIDAESEMDPCKVDSTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_009443992.1|P_troglodytes        EHDGSTGVTIDAESEMDPCKVDSTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
Q52V16.1|G_gorilla                  EHDGSTGVTIDAESEMDPCKVDSTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_014984082.1|M_mulatta            EHDGSTGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_033067617.1|T_francoisi          EHDGSTGVTIDGESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_031516968.1|P_anubis             EHDGSTGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_008017167.1|C_sabaeus            EHDGSTGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_030782172.1|R_roxellana          EHDGSTGVTIDGESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_032612406.1|H_moloch             EHDGSTGVTVDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_035145821.1|C_jacchus            EHDGSSGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVENESENDHGV [300]
P10925.3|M_musculus                 DLEGSSEVTMNAESGTDSSKLDEASPEVIKVCILKADSEVDELGETIHAVESETKNGNEA [300]
P20662.2|M_musculus                 DHEGSSEVTMNAESETDSSKLDEASPEVIKVCILKADSEVDDVGETIQAVESETDNGNEA [300]
XP_008771898.1|R_norvegicus         ENEGSSEVTTNAESESDPYKLNETSPEVIKVYIFKADPEEDDVGETVDIVESKTDNGNEA [300]
XP_015343506.1|M_marmota            EHNGSTAVNTSAESDIDSCKVEGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
Q95LI3.1|B_taurus                   EHDCSSGMTMDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_010855418.1|B_bison              EQDCSAGMTIDRESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_017900383.1|C_hircus             EQDCSAGMTIDRESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
AMY96563.1|C_elaphus                EQDCSAGMTIDAESEIDPCKVDGTCPEVIKVYIFRADPGEDDLGGTVDIVESEPENDHGV [300]
XP_020759307.1|O_virginianus        EQDCSAGVTIDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
F1SPY3|S_scrofa                     EHDGSSEMTMDAESEINPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_024612082.1|N_asiaeorientalis    EHDGSSGMTMDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDRGV [300]
AKI82174.1|C_lupus                  EHGGSSGMTIDTESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_032187800.1|M_erminea            EHGGSSGMTMNTESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
JAC06687.1|L_africana               GHDGTSGITMDTESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
No_accession|E_caballus             EQDGSSGMTMDTELEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHEV [300]
XP_028935710.1|O_anatinus           DHDGSSEMTMDAEPEIDPCKVDGGCPEVIKVYIFKADPGEDDLGGTVDIVESEPENEHGV [300]
XP_016288863.1|M_domestica          EHDGSSEITMDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_015127980.1|G_gallus             EHEGSAEITMEAESESGSCKVDGICPEVIKVYIFKADPGEDDLGGTVDVVESEPENDHAV [300]
Q01611.1|X_laevis                   DEDGAEEITMGSVVEGDSSKLDGSCPEVIKVYIFKADPGEEDLGGTVDIVESESENDHGD [300]

                                               ** *  **   ***    ** **** ****  **** ***    *
NP_001356631.1|H_sapiens            ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa- [360]
XP_009443992.1|P_troglodytes        ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa- [360]
Q52V16.1|G_gorilla                  ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa- [360]
XP_014984082.1|M_mulatta            ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa- [360]
XP_033067617.1|T_francoisi          ELLDQNSSIRVPREKMVYMTVSDSQQED----VAEIADEVYMEVIVGEEDaaVaaaaaa- [360]
XP_031516968.1|P_anubis             ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa- [360]
XP_008017167.1|C_sabaeus            ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa- [360]
XP_030782172.1|R_roxellana          ELLDQNSSIRVPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaa- [360]
XP_032612406.1|H_moloch             ELLDQNSSIRVPREKMVYMTVNDSQREDEDLNVAEIADEVYMEVIVGEEDaaVaaaaa-- [360]
XP_035145821.1|C_jacchus            ELLEQSSSVRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaPATVaaaa- [360]
P10925.3|M_musculus                 EVTDQSTSIRVPRV-NIYMSASDSQKEEED-----------TEVIVGDEDaGGTaaDTP- [360]
P20662.2|M_musculus                 EVTDQRTSIHVPRV-NIYMLASDSQKEEED-----------TKVIVGDEDaGGTaaDTP- [360]
XP_008771898.1|R_norvegicus         EVIDQSSSIYVPRD-NVYMPVSDSQKEEED-----------TKVIVGDEDaGDTaaDTS- [360]
XP_015343506.1|M_marmota            ELLDQNSTIRVPREKMVYMTVNDSQQEDEDLNVAEITDEVYMEVIVGEEDaaVTaaaaa- [360]
Q95LI3.1|B_taurus                   ELLDQNNSIRMPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaTT [360]
XP_010855418.1|B_bison              ELLDQNNSIRMPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaTT [360]
XP_017900383.1|C_hircus             ELLDQSNSIRMPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaTT [360]
AMY96563.1|C_elaphus                EILDQNNSIRVPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaaT [360]
XP_020759307.1|O_virginianus        EILDQNNSIRVPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEED-aVaaaaaaT [360]
F1SPY3|S_scrofa                     ELLDQNSSMRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaaa [360]
XP_024612082.1|N_asiaeorientalis    ELLDQNSSIRVPREKMVYMTVNDSQQ-DEDLNVAEIADEVYMEVIVGEEDaaVaaaaaaa [360]
AKI82174.1|C_lupus                  ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaaa [360]
XP_032187800.1|M_erminea            ELLDQNSSIRVPREKMVYMTVNDSQQEDDDLNVAEIADEVYMEVIVGEEDaaVaaaaaaa [360]
JAC06687.1|L_africana               ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaaa [360]
No_accession|E_caballus             ELLDQNNSIRVPRDKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVaaaaaaa [360]
XP_028935710.1|O_anatinus           GLLDQSSSIRVPREKMVYMTVNDSQQEDEDLSVAEIADEVYMEVIVGEEDaaVa------ [360]
XP_016288863.1|M_domestica          GLLDQSSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDaaVa------ [360]
XP_015127980.1|G_gallus             GLLDQNSSIRIPREKMVYMTVNDSQHEDEDLNVAEIADEVYMEVIVGEEDaaVa------ [360]
Q01611.1|X_laevis                   GFLDSHNGGRLPREKMVYMTVNDSQNDD-DLDVAEIADEVYMEVIVGEEDaaVa------ [360]

                                     **      *    * * ** *** ** * **     *** *   ** ** *  * * **
NP_001356631.1|H_sapiens            VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_009443992.1|P_troglodytes        VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
Q52V16.1|G_gorilla                  VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_014984082.1|M_mulatta            VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESTGLGRLAKQKPKK [420]
XP_033067617.1|T_francoisi          VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESAGLGRLAKQKPKK [420]
XP_031516968.1|P_anubis             VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESTGLGRLAKQKPKK [420]
XP_008017167.1|C_sabaeus            VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESTGLGRLAKQKPKK [420]
XP_030782172.1|R_roxellana          VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESAGLGRLAKQKPKK [420]
XP_032612406.1|H_moloch             VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESAGLGRLAKQKPKK [420]
XP_035145821.1|C_jacchus            MHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGSASAVLHVDESVGLSRLTKQKPKK [420]
P10925.3|M_musculus                 EHEQQMDVSEIKAAFLPIAWTAAYDNNSDEIEDQNVTASALLNQDESGGLDRVPKQKSKK [420]
P20662.2|M_musculus                 EHEQQMDVSEIKAAFLPIAWTAAYDNNSDEIEVQNATASAMLHHDESGGLDRVPKQKSKK [420]
XP_008771898.1|R_norvegicus         EHEQQMDDSEIKAAFLPIAWAAAYDNNSDEIEEQNVTASAVLHQNESGGLDRVHKQKAKK [420]
XP_015343506.1|M_marmota            VHEQQIDDSEM-KAFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLSRLAKQKPKK [420]
Q95LI3.1|B_taurus                   VHEQEMDDSEI-KTFMPIAWAAAYGNNSDGIENRSGTASALLHIDESAGLGRLTKHKPKK [420]
XP_010855418.1|B_bison              VHEQEMDDSEI-KTFMPIAWAAAYGNNSDGIENRSGTASALLHIDESAGLGRLTKHKPKK [420]
XP_017900383.1|C_hircus             VHEQEMDDSEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
AMY96563.1|C_elaphus                VHEQEMDDNEM-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_020759307.1|O_virginianus        VHEQEMDDSEM-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
F1SPY3|S_scrofa                     VHEQQMDDSEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_024612082.1|N_asiaeorientalis    VHEQQMDNSEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
AKI82174.1|C_lupus                  VHEQQMDDNEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_032187800.1|M_erminea            VHEQQMDDNEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLSRLAKQKPKK [420]
JAC06687.1|L_africana               VHEQQMDDSEI-KTFVPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRVTKQKPKK [420]
No_accession|E_caballus             VHEQQMDDSEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_028935710.1|O_anatinus           -HEQQMDDTEI-KTFMPIAWAAAYGNNTDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_016288863.1|M_domestica          -HEQQIDDTEI-KTFMPIAWAAAYGNNTDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_015127980.1|G_gallus             -HEQQIDDNEI-KTFMPIAWAAAYGNNNDGIESRNGTASALLHIDESAGLGRLAKQKPKK [420]
Q01611.1|X_laevis                   -HEHQLEDAELSKTFMPVAWAAAYGNNTDGIEHRNGTASALLHIDESDGLDRLTKQKLKK [420]
```



```
                                      *    *** **   ***  * ****  ******   ***** *****     *** **
NP_001356631.1|H_sapiens            KRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLA-KKKYHCT [480]
XP_009443992.1|P_troglodytes        KRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLA-KKKYHCT [480]
Q52V16.1|G_gorilla                  KRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLA-KKKYHCT [480]
XP_014984082.1|M_mulatta            KRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLA-KKKYHCT [480]
XP_033067617.1|T_francoisi          KRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLA-KKKYHCT [480]
XP_031516968.1|P_anubis             KRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLA-KKKYHCT [480]
XP_008017167.1|C_sabaeus            KRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLA-KKKYHCT [480]
XP_030782172.1|R_roxellana          KRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLA-KKKYHCT [480]
XP_032612406.1|H_moloch             KRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLA-KKKYHCT [480]
XP_035145821.1|C_jacchus            KRRSDARQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLA-KKKYHCT [480]
P10925.3|M_musculus                 KKRPESKQYQSAIFVAPDGQTLRVYPCMFCGKKFKTKRFLKRHTKNHPEYLA-NKKYHCT [480]
P20662.2|M_musculus                 KKRPESKQYQSAIFVAPDGQTLRVYPCMFCGKKFKTKRFLKRHIKNHPEYLA-NKKYHCT [480]
XP_008771898.1|R_norvegicus         KKRPESKQYQTAIIVAPDGQTLIVYPCMFCGKKFKTKSFLKRHIKNHPEYLA-KKKYHCT [480]
XP_015343506.1|M_marmota            RRRPDSKQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHIA-KKKYCCT [480]
Q95LI3.1|B_taurus                   RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLT-KKKYRCT [480]
XP_010855418.1|B_bison              RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLT-KKKYRCT [480]
XP_017900383.1|C_hircus             RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLT-KKKYRCT [480]
AMY96563.1|C_elaphus                RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLT-KKKYRCT [480]
XP_020759307.1|O_virginianus        RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLT-KKKYRCT [480]
F1SPY3|S_scrofa                     RRRPDSRQYQTAIIIGPDGHPLTVYPCLICGKKFKSRGFLKRHMKNHPEHLT-KKKYRCT [480]
XP_024612082.1|N_asiaeorientalis    RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLT-KKKYHCT [480]
AKI82174.1|C_lupus                  RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLS-KKKYRCT [480]
XP_032187800.1|M_erminea            RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLT-KKKYRCT [480]
JAC06687.1|L_africana               RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLT-KKKYRCT [480]
No_accession|E_caballus             RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLT-KKKYHCT [480]
XP_028935710.1|O_anatinus           RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLS-KKKYRCT [480]
XP_016288863.1|M_domestica          RRRPDSRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLT-KKKYRCT [480]
XP_015127980.1|G_gallus             KRRPESRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLLTKKKYRCT [480]
Q01611.1|X_laevis                   KRRGENRQYQTAIIIGPDGHPLTVYPCMICGKKFKSRGFLKRHMKNHPEHLV-RKKYRCT [480]

                                     * * **** * *** ******    *         *  * *   *   *         *
NP_001356631.1|H_sapiens            DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
XP_009443992.1|P_troglodytes        DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
Q52V16.1|G_gorilla                  DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
XP_014984082.1|M_mulatta            DCDYTTNKKISLHNHLESHKLTSKAEK------AIQCDECGKHFSHAGALFTHKMVHKEK [540]
XP_033067617.1|T_francoisi          DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
XP_031516968.1|P_anubis             DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
XP_008017167.1|C_sabaeus            DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
XP_030782172.1|R_roxellana          DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
XP_032612406.1|H_moloch             DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
XP_035145821.1|C_jacchus            DCDYTTNKKISLHNHLESHKLTSKAEK------TIECVECGKHFSHAGALFTHKMVHKEK [540]
P10925.3|M_musculus                 ECDYSTNKKISLHNHMESHKLTIKTEK------TTECDDCRKNLSHAGTLCTHKTMHTEK [540]
P20662.2|M_musculus                 ECDYSTNKKISLHNHMESHKLTIKTEK------TTECDDCRKNLSHAG------TMHTEK [540]
XP_008771898.1|R_norvegicus         DCDYSTNKKISLHNHMESHKLTIKTEK------TTECDDCGKHLSHAG---TLCTHKKEK [540]
XP_015343506.1|M_marmota            DCDYTTNKKISLHNHLESHKLTSKVEK------VIECDECGKHFSHTGALFTHKMVHKEK [540]
Q95LI3.1|B_taurus                   DCDYTTNKKISLHNHLESHKLTSKSEK------AIECDDCGKHFSHAGALFTHKMVHKEK [540]
XP_010855418.1|B_bison              DCDYTTNKKISLHNHLESHKLTSKSEK------AIECDDCGKHFSHAGALFTHKMVHKEK [540]
XP_017900383.1|C_hircus             DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
AMY96563.1|C_elaphus                DCDYTTNKKMSLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
XP_020759307.1|O_virginianus        DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
F1SPY3|S_scrofa                     DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
XP_024612082.1|N_asiaeorientalis    ACDYTTNKKISFHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
AKI82174.1|C_lupus                  DCDYTTNKKISLHNHLESHKLTSKAEK------SIECEECGKHFSHAGALFTHKMVHKEK [540]
XP_032187800.1|M_erminea            DCEYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
JAC06687.1|L_africana               DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHTGALFTHKMVHKEK [540]
No_accession|E_caballus             DCDYTTNKKISLHNHLESHKLTSKAEK------AIECDECGKHFSHAGALFTHKMVHKEK [540]
XP_028935710.1|O_anatinus           DCDYTTNKKVSLHNHLESHKLTGKAEKAAAPGAGAECDECGKHFSHAGALFTHKTVHKEK [540]
XP_016288863.1|M_domestica          DCDYTTNKKISLHNHLESHKLTNKTEKAIEC------DECGKHFSHAGALFTHKMVHKEK [540]
XP_015127980.1|G_gallus             DCDYTTNKKISLHNHLESHKLTNKTEKLIER------DECGKSFSHAGALFAHKMVHRDK [540]
Q01611.1|X_laevis                   DCDYTTNKKVSLHNHLESHKLTATVIKTEKD---LECEECGKIFLHANALFAHKLTHNEK [540]


                                            **** * ****  *  *** ** *  ** * ********** * ** * ***
NP_001356631.1|H_sapiens            -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELRKHMRIHTG [600]
XP_009443992.1|P_troglodytes        -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
Q52V16.1|G_gorilla                  -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_014984082.1|M_mulatta            -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_033067617.1|T_francoisi          -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_031516968.1|P_anubis             -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_008017167.1|C_sabaeus            -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_030782172.1|R_roxellana          -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_032612406.1|H_moloch             -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_035145821.1|C_jacchus            -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
P10925.3|M_musculus                 G-VNKTCKCKFCDYETAEQTLLNHHLLVVHRKKFPHICGECGKGFRHPSALKKHIRVHTG [600]
P20662.2|M_musculus                 G-VNKTCKCKFCDYETAEQTLLNHHLLVVHRKKFPHICGECGKGFRHPSALKKHIRVHTG [600]
XP_008771898.1|R_norvegicus         EKVSKTYKCKFCDYETAEQTSLNHHLLAVHSKKYPHVCVECGKGFRHPSELKKHIRVHTG [600]
XP_015343506.1|M_marmota            -GNNKMHKCKFCEYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRIHTG [600]
Q95LI3.1|B_taurus                   -GASKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_010855418.1|B_bison              -GASKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_017900383.1|C_hircus             -GASKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELRKHMRIHTG [600]
AMY96563.1|C_elaphus                -GANKMHKCKFCEYGTAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_020759307.1|O_virginianus        -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
F1SPY3|S_scrofa                     -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_024612082.1|N_asiaeorientalis    -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
AKI82174.1|C_lupus                  -GTNKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_032187800.1|M_erminea            -GANKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
JAC06687.1|L_africana               -GSSKMHKCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
No_accession|E_caballus             -GANKMHRCKFCEYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_028935710.1|O_anatinus           GAGGRTHKCKFCDYETAEQGLLNRHLLAVHSKNFPHVCVECGKGFRHPSELKKHMRIHTG [600]
XP_016288863.1|M_domestica          -GANKMHKCKFCDYETAEQGLLNRHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
XP_015127980.1|G_gallus             -GVNKMHKCKFCDYETAEQGLLSHHLLAVHSKNFPHICVECGKGFRHPSELKKHMRIHTG [600]
Q01611.1|X_laevis                   -AGNKMHICKFCDYETAEQGLLNRHLLAVHSKSFPHICVECGKGFRHPSELKKHMRTHTG [600]
```

```
                                    **** ** * * ************ ** *     *  *  ** * *  * *   *  *
NP_001356631.1|H_sapiens            EKPYQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ESK [660]
XP_009443992.1|P_troglodytes        EKPYQCQYCEYRSADSSNLKTHIKTKHSKEMPLKCDICLLTFSDTKEVQQHTLVHQ-ESK [660]
Q52V16.1|G_gorilla                  EKPYQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ESK [660]
XP_014984082.1|M_mulatta            EKPYQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ENK [660]
XP_033067617.1|T_francoisi          EKPYQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ENK [660]
XP_031516968.1|P_anubis             EKPYQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ESK [660]
XP_008017167.1|C_sabaeus            EKPYQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHALIHQ-ESK [660]
XP_030782172.1|R_roxellana          EKPYQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQHHTLVHQ-ENR [660]
XP_032612406.1|H_moloch             EKPYQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ESK [660]
XP_035145821.1|C_jacchus            EKPYQCQYCEYRSADSSNLKTHIKTKHSKEMPFKCDICLLTFSDTKEVQQHTLVHQ-ESR [660]
P10925.3|M_musculus                 EKPYECQYCEYKSADSSNLKTHIKSKHSKEIPLKCGICLLTFSDNKEAQQHAVLHQ-ESR [660]
P20662.2|M_musculus                 EKPYECQYCEYKSADSSNLKTHIKSKHSKEIPLKCGICLLTFSDTKEAQQHAVLHQ-ESR [660]
XP_008771898.1|R_norvegicus         EKPYQCQYCEYKSADSSNLKTHIKTKHSKDIPLKCGICLMTFSDTKEAQQHALIHQ-ENR [660]
XP_015343506.1|M_marmota            EKPYQCQYCEYRSADSSNLKTHVKTKHTKEMPFKCDICLLTFSDTKEVQQHALIHQ-ESK [660]
Q95LI3.1|B_taurus                   EKPYQCQYCEYRSADSSNLKTHVKTKHSKEMSFKCDICLLTFSDTKEVQQHALIHQ-ESK [660]
XP_010855418.1|B_bison              EKPYQCQYCEYRSADSSNLKTHVKTKHSKEMSFKCDICLLTFSDTKEVQQHALIHQ-ESK [660]
XP_017900383.1|C_hircus             EKPYQCQYCEYRSADSSNLKTHVKTKHSKEMSFKCDICLLTFSDTKEVQQHALIHQ-ESK [660]
AMY96563.1|C_elaphus                EKPYRCQYCEYRSADSSNLKTHVKTKHSKEMSFKCDICLLTFSDTKEVQQHALIHQ-ESK [660]
XP_020759307.1|O_virginianus        EKPYQCQYCEYRSADSSNLKTHVKTKHSKEMSFNCDICLLTFSDTKEVQQHALIHQ-ESK [660]
F1SPY3|S_scrofa                     EKPYQCQYCEYRSADSSNLKTHVKTKHSKEMPFKCDICLLTFSDTKDVQQHALIHQ-ESK [660]
XP_024612082.1|N_asiaeorientalis    EKPYQCQYCEYRSADSSNLKTHVKTKHSKEMPFKCDSCLLTFSDTKEVQQHALIHQ-ESK [660]
AKI82174.1|C_lupus                  EKPYQCQYCEYRSADSSNLKTHVKTKHSKEMPFKCDICLLTFSDTKEVQQHAVIHQ-ESK [660]
XP_032187800.1|M_erminea            EKPYQCQYCEYRSADSSNLKTHVKTKHSKEMPFKCDICLLTFSDTKEVQQHALIHQ-ESK [660]
JAC06687.1|L_africana               EKPYQCQYCEYRSADSSNLKTHVKTKHSKEMPYRCDICLLTFSDTKEVQQHAVIHQ-ESK [660]
No_accession|E_caballus             EKPYHCQYCEYRSADSSNLKTHVKTKHSKEMPFKCDICLLTFSDTKEVQQHGLIHQ-ESK [660]
XP_028935710.1|O_anatinus           EKPYQCQFCPYRSADSSNLKTHVKTKHSKETPFRCEACPLTFADPKELQQHALLHHQESR [660]
XP_016288863.1|M_domestica          EKPYQCQYCEYRSADSSNLKTHVKTKHSKEMPFKCEICLLTFSDTKEVQQHALIHQ-ESK [660]
XP_015127980.1|G_gallus             EKPYQCQYCEYRSADSSNLKTHVKTKHSKETSSKCDICFQTFSDTKELQQHTLMHQ-ESK [660]
Q01611.1|X_laevis                   EKPYLCQYCDYRSADSSNLKTHVKTKHSKEMPFKCDICLQTFTDSKDLQEHAILHQ-ESK [660]

                                     ***  * ************ ****** *****  * ************ * *   * **
NP_001356631.1|H_sapiens            THQCLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQ [720]
XP_009443992.1|P_troglodytes        THQCLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQ [720]
Q52V16.1|G_gorilla                  THQCLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQ [720]
XP_014984082.1|M_mulatta            THQCLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQ [720]
XP_033067617.1|T_francoisi          THQCLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQ [720]
XP_031516968.1|P_anubis             THQCLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQ [720]
XP_008017167.1|C_sabaeus            THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
XP_030782172.1|R_roxellana          THQCLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQ [720]
XP_032612406.1|H_moloch             THQCLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCEKGFHRPSELKKHVAVHKGKKMHQ [720]
XP_035145821.1|C_jacchus            THQCLHCDHKSSNSSDLKRHVISVHTKDYPHKCEMCDKGFHRPSELKKHVAVHKGKKMHQ [720]
P10925.3|M_musculus                 THQCSHCNHKSSNSSDLKRHIISVHTKAYPHKCDMCSKGFHRPSELKKHVATHKSKKMHQ [720]
P20662.2|M_musculus                 THQCSHCNHKSSNSSDLKRHIISVHTKAYPHKCDMCSKGFHRPSELKKHVATHKSKKMHQ [720]
XP_008771898.1|R_norvegicus         THQCSYCNHKSSNSSDLKRHIISVHTKDYPHKCDMCSKGFHRPSELKKHVATHKSKKMHQ [720]
XP_015343506.1|M_marmota            THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
Q95LI3.1|B_taurus                   THQCVHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
XP_010855418.1|B_bison              THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
XP_017900383.1|C_hircus             THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
AMY96563.1|C_elaphus                THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
XP_020759307.1|O_virginianus        THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
F1SPY3|S_scrofa                     THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCEMCEKGFHRPSELKKHVAAHKGKKMHQ [720]
XP_024612082.1|N_asiaeorientalis    THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
AKI82174.1|C_lupus                  THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
XP_032187800.1|M_erminea            THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
JAC06687.1|L_africana               THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
No_accession|E_caballus             THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCEMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
XP_028935710.1|O_anatinus           AHQCLHCDHKSSNSSDLKRHVISVHTKDYPHKCDTCDKGFHRPSELKKHAAAHRGRKLHQ [720]
XP_016288863.1|M_domestica          THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKMHQ [720]
XP_015127980.1|G_gallus             THQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCDMCDKGFHRPSELKKHVAAHKGKKLHQ [720]
Q01611.1|X_laevis                   NHQCLHCDHKSSNSSDLKRHIISVHTKDYPHKCEVCEKGFHRPSELKKHEAAHKGKKMHQ [720]

                                    **** *   *** ** **** * *  *  **** * * **  *  ******  *******
NP_001356631.1|H_sapiens            CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_009443992.1|P_troglodytes        CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
Q52V16.1|G_gorilla                  CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_014984082.1|M_mulatta            CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_033067617.1|T_francoisi          CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_031516968.1|P_anubis             CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_008017167.1|C_sabaeus            CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_030782172.1|R_roxellana          CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_032612406.1|H_moloch             CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_035145821.1|C_jacchus            CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFNQQNELKKHMKTHSGRKVYQCEY [780]
P10925.3|M_musculus                 CRHCDFNSPDPFLLSHHILSAHTKNVPFKCKRCKKEFQQQCELQTHMKTHSSRKVYQCEY [780]
P20662.2|M_musculus                 CRHCDFKSPDPFLLSHHILSAHTKNVPFKCKRCKKEFQQQCELQTHMKTHSSRKVYQCEY [780]
XP_008771898.1|R_norvegicus         CRHCDFKSPDPFLLSRHILSVHTKNVPFKCKRCKKGFRQQCELQKHMKTHSGRKVYQCEY [780]
XP_015343506.1|M_marmota            CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
Q95LI3.1|B_taurus                   CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCKKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_010855418.1|B_bison              CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCKKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_017900383.1|C_hircus             CRHCDFKIADPFVLSRHILSIHTKDLPFRCKRCKKGFRQQNELKKHMKTHSGRKVYQCEY [780]
AMY96563.1|C_elaphus                CRHCDFRIADPFVLSRHILSVHAKDLPFRCKRCKKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_020759307.1|O_virginianus        CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCKKGFRQQNDLKKHMKTHSGRKVYQCEY [780]
F1SPY3|S_scrofa                     CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_024612082.1|N_asiaeorientalis    CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
AKI82174.1|C_lupus                  CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_032187800.1|M_erminea            CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
JAC06687.1|L_africana               CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
No_accession|E_caballus             CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQTELKKHMKTHSGRKVYQCEY [780]
XP_028935710.1|O_anatinus           CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQGELKKHMKTHSGRKVYQCEY [780]
XP_016288863.1|M_domestica          CRHCDFKIADPFVLSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
XP_015127980.1|G_gallus             CRHCDFKIADPFILSRHILSVHTKDLPFRCKRCRKGFRQQNELKKHMKTHSGRKVYQCEY [780]
Q01611.1|X_laevis                   CRHCEFHIADPFVLSRHILSVHTKELPYRCKRCKKGFRQQIELKKHMKTHSGKKVYQCEY [780]
```



```
                                    *** * *******************    **************   ***   *
NP_001356631.1|H_sapiens            CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
XP_009443992.1|P_troglodytes        CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
Q52V16.1|G_gorilla                  CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
XP_014984082.1|M_mulatta            CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
XP_033067617.1|T_francoisi          CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
XP_031516968.1|P_anubis             CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
XP_008017167.1|C_sabaeus            CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
XP_030782172.1|R_roxellana          CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
XP_032612406.1|H_moloch             CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
XP_035145821.1|C_jacchus            CEYNTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
P10925.3|M_musculus                 CEYSTKDASGFKRHVISIHTKDYPHSCDFCKKGFRRPSEKNQHIMRHHK-VGLP [834]
P20662.2|M_musculus                 CEYSTKDASGFKRHVISIHTKDYPHRCDFCKKGFRRPSEKNQHIMRHHKEVGLA [834]
XP_008771898.1|R_norvegicus         CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
XP_015343506.1|M_marmota            CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVSLS [834]
Q95LI3.1|B_taurus                   CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHITRHHKEVGLP [834]
XP_010855418.1|B_bison              CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHITRHHKEVGLP [834]
XP_017900383.1|C_hircus             CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHITRHHKEVGLP [834]
AMY96563.1|C_elaphus                CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHITRHHKEVGLP [834]
XP_020759307.1|O_virginianus        CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHITRHHKEVGLP [834]
F1SPY3|S_scrofa                     CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
XP_024612082.1|N_asiaeorientalis    CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
AKI82174.1|C_lupus                  CEYSTTDASGFKRHVISIHTKDYPHRSS-------------------------- [834]
XP_032187800.1|M_erminea            CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
JAC06687.1|L_africana               CEYSTTDASGFKRHVISIHTKDYPHRCEHCKKGFRRPSEKNQHIMRHHKEVGLP [834]
No_accession|E_caballus             CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKEVGLP [834]
XP_028935710.1|O_anatinus           CEYSTTDASGFKRHVISIHTKDYPHRCDFCKKGFRRPSEKNQHIMRHHKDLGLP [834]
XP_016288863.1|M_domestica          CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKDVGLP [834]
XP_015127980.1|G_gallus             CEYSTTDASGFKRHVISIHTKDYPHRCEYCKKGFRRPSEKNQHIMRHHKDVGLP [834]
Q01611.1|X_laevis                   CEYNTTDASGFKRHVISIHTKDYPHRCDYCKKGFRRPSEKNQHTLKHHKEASLM [834]
```

Table 11. Multiple CLUSTALW sequence alignment of land vertebrate ZFY primary sequence. The alignment shows conserved amino acids indicated by the YELLOW highlight and the asterisk (*). Each exon has exon colour coding with RED representing Exon 1, GREEN representing Exon 2, CYAN representing Exon 3, MAGENTA representing Exon 4, BROWN representing Exon 5, PURPLE representing Exon 6 and GREY representing Exon 7. The BLACK represents amino acids across a splice junction. `lowercase a` represents alanine residues in polyalanine motif and **`BOLD UPPERCASE`** residues represent amino acids that are not conserved throughout the whole aligned species.

Exons	Percentage Identity (%)	Number of conserved sites	Number of Parsimony informative sites
Exon 1	14.3	3	11
Exon 2	22.3	47	93
Exon 3	24.5	12	27
Exon 4	29.8	14	22
Exon 5	50.0	26	20
Exon 6	47.8	22	18
Exon 7	61.4	247	98

Table 12. Table demonstrating the percentage identity, conserved sites, and parsimony informative sites of land vertebrates and Cetartiodactylaa. The table shows the percentage identity, calculated using the conserved sites and the total number of sites in aligned sequence (by exon).

Phylogenetic analysis of the land vertebrates as illustrated by **Figure 9** indicated that ZFY was highly conserved within the primates, and they were closely related. **Figure 9** shows similarly to **Figure 8** that *M. musculus* and *R. norvegicus* were phylogenetically discordant as they were not in the expected region of the tree. The outgroup species was the *X. laevis* as this is the most distantly related to the warm-blooded land vertebrates (and *N. asiaeorientalis*) and as **Figure 9** indicated, the branch was long showing it was distantly related to the other land vertebrates. However, as genes display codon redundancy as some amino acids have multiple codons due to there being 64 codons and only 20 amino acids, the alignment and phylogenetic tree were done on a nucleotide level to further observe the conservation of the nucleotide sequence (cDNA).

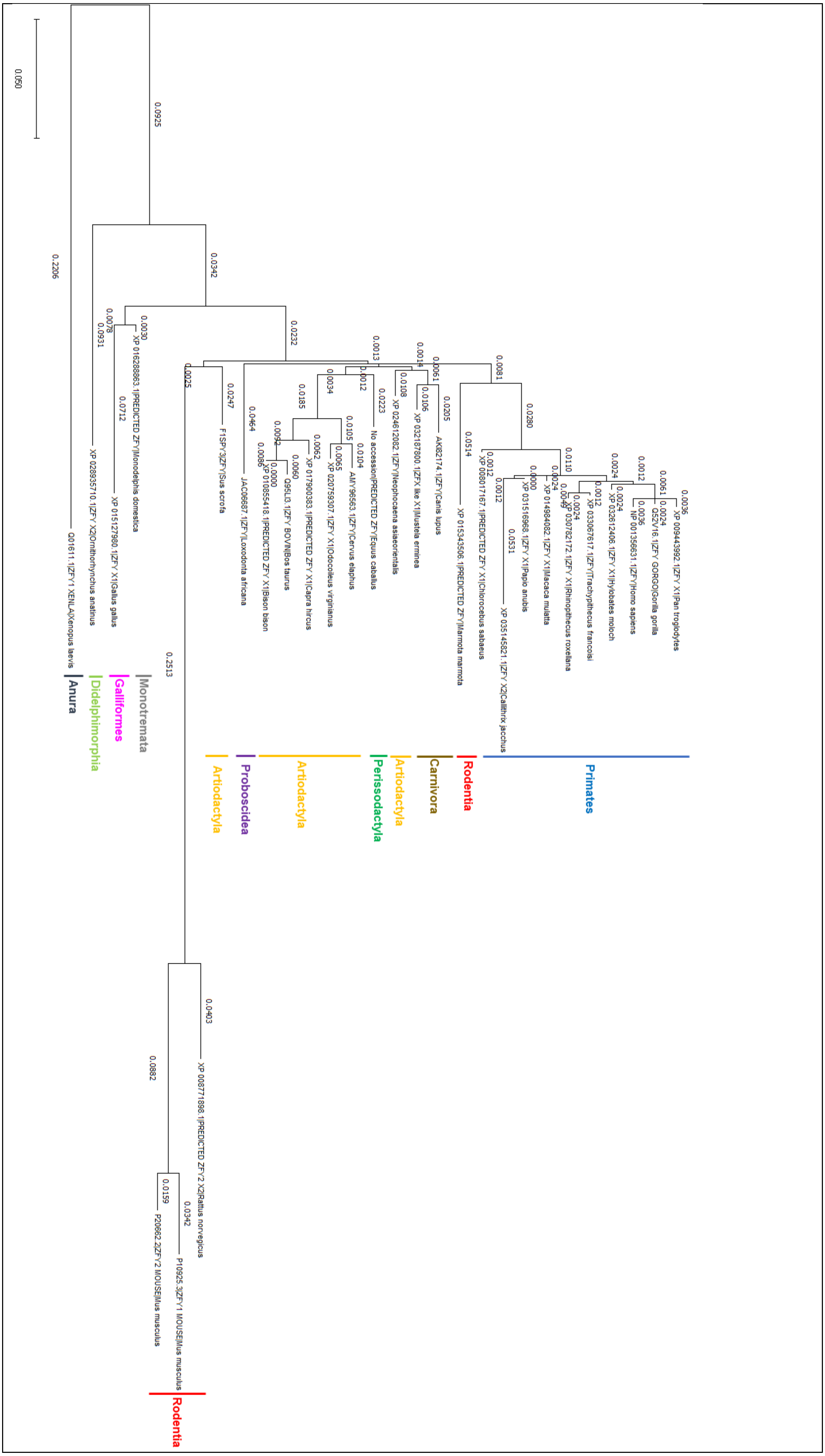

0.050
XP 009443992.1|ZFY X1|Pan troglodytes
Q52V16.1|ZFY GORGO|Gorilla gorilla
NP 001356631.1|ZFY|Homo sapiens
XP 032612406.1|ZFY X1|Hylobates moloch
XP 033067617.1|ZFY|Trachypithecus francoisi
XP 030782172.1|ZFY X1|Rhinopithecus roxellana
XP 014984082.1|ZFY X1|Macaca mulatta
XP 031516968.1|ZFY X1|Papio anubis
XP 035145821.1|ZFY X2|Callithrix jacchus
XP 008017167.1|PREDICTED ZFY X1|Chlorocebus sabaeus
XP 015343506.1|PREDICTED ZFY|Marmota marmota
AKI82174.1|ZFY|Canis lupus
XP 032187800.1|ZFX like X1|Mustela erminea
XP 024612082.1|ZFY|Neophocaena asiaeorientalis
No accession|PREDICTED ZFY|Equus caballus
AMY96563.1|ZFY|Cervus elaphus
XP 020759307.1|ZFY X1|Odocoileus virginianus
XP 017900383.1|PREDICTED ZFY X1|Capra hircus
Q95LI3.1|ZFY BOVIN|Bos taurus
XP 010855418.1|PREDICTED ZFY X1|Bison bison
JAC06687.1|ZFY|Loxodonta africana
F1SPY3|ZFY|Sus scrofa
XP 016288863.1|PREDICTED ZFY|Monodelphis domestica
XP 015127980.1|ZFY X1|Gallus gallus
XP 028935710.1|ZFY X2|Ornithorhynchus anatinus
Q01611.1|ZFY1 XENLA|Xenopus laevis
XP 008771898.1|PREDICTED ZFY2 X2|Rattus norvegicus
P10925.3|ZFY1 MOUSE|Mus musculus
P20662.2|ZFY2 MOUSE|Mus musculus
Primates
Rodentia
Carnivora
Artiodactyla
Perissodactyla
Artiodactyla
Proboscidea
Artiodactyla
Monotremata
Galliformes
Didelphimorphia
Anura
Rodentia
0.0036
0.0061
0.0024
0.0036
0.0012
0.0024
0.0024
0.0012
0.0024
0.0049
0.0110
0.0024
0.0000
0.0012
0.0531
0.0012
0.0280
0.0514
0.0081
0.0205
0.0061
0.0106
0.0014
0.0108
0.0013
0.0223
0.0012
0.0104
0.0105
0.0065
0.0034
0.0062
0.0185
0.0060
0.0092
0.0000
0.0086
0.0464
0.0247
0.0025
0.0232
0.0342
0.0925
0.0030
0.0078
0.0712
0.0931
0.2206
0.2513
0.0403
0.0882
0.0342
0.0159

Figure 9. Maximum likelihood (Bootstrap) tree of land vertebrate ZFY protein sequences shown by Table 10. The tree was inferred by Maximum Likelihood method and JTT matrix-based model. The tree represents the coding exons of different ZFY species. Species were grouped taxonomically and each order is indicated by the coloured text and labelling. The scale as indicated represents 0.050 amino acid substitutions per site. The species used as an outgroup was X. laevis.

The nucleotide alignment showed that 1065 nucleotides of the aligned sequences were conserved (2568 aligned sites in total). Though most nucleotides of the different land vertebrate species were not conserved, most of the change in amino acids resulted in the same codon because of codon degeneracy as the translated nucleotide sequence resulted in also 375 conserved amino acids in total. This nucleotide alignment was used to generate a phylogenetic tree shown by **Figure 10**, which showed the rodents were all grouped together. Though, the branch length in the phylogenetic tree consistently remains longer than the other placental mammals as previously observed also in **Figure 8** and **Figure 9**.

Figure 10. Maximum likelihood (Bootstrap) tree of multiple ZFY nucleotide sequence alignment. The tree was inferred by Maximum Likelihood method and Kimura 2-parameter model. The tree was generated using cDNA sequences obtained from NCBI GenBank, Uniprot, Ensembl, EMBL-EBI and published papers. The sequences were labelled by order and the outgroup species used was the *X. laevis*. The scale represents 0.10 nucleotide substitutions per site. The accession numbers correspond to the nucleotide accession numbers provided by **Table 1**.

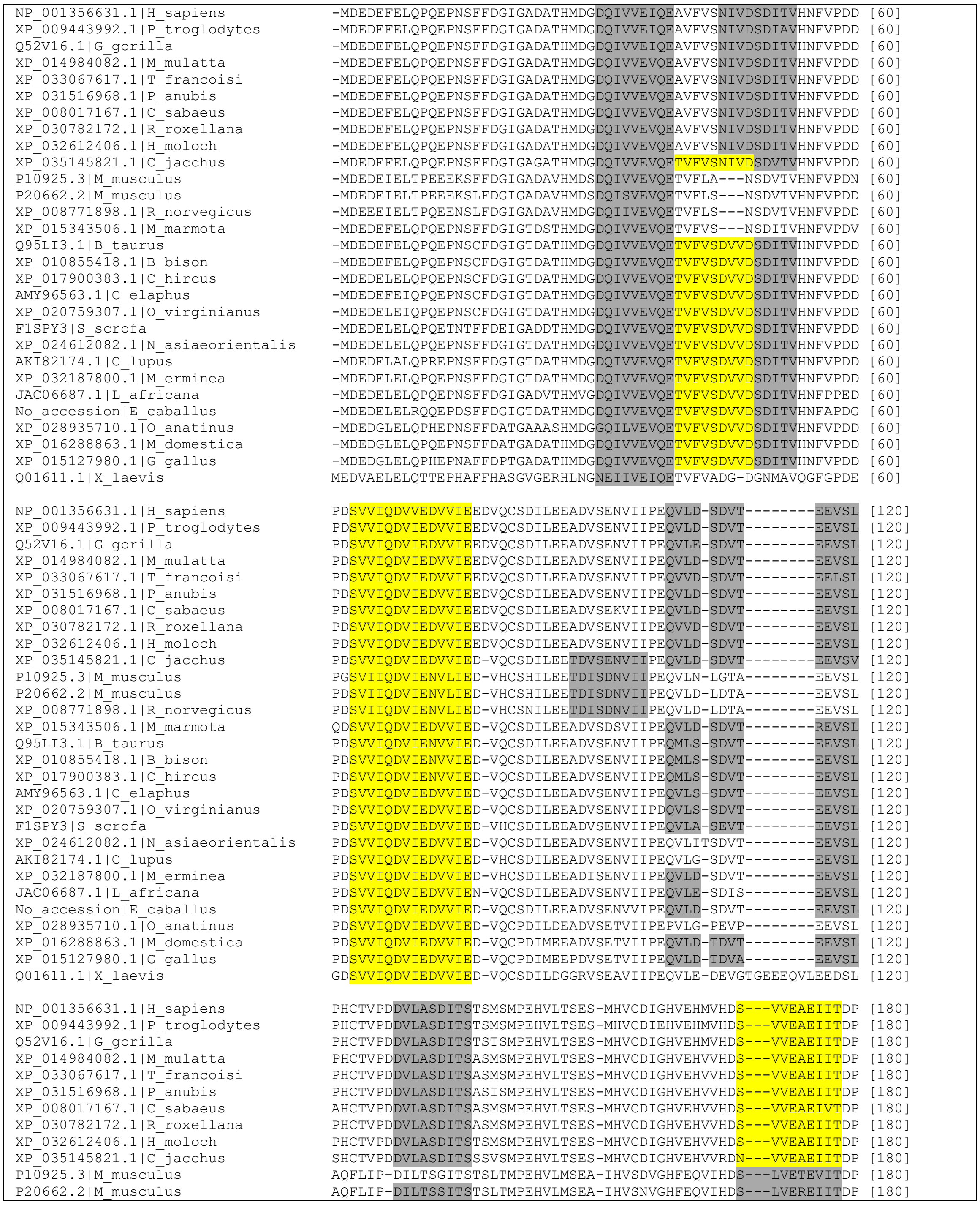
3.4 9aa TAD and DNA binding site prediction

```
                 ********************
hZFY_long     MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEIQEAVFVSNIVDSDITVHNFVPDDP  [60]
hZFY_short    MDEDEFELQPQEPNSFFDGI----------------------------------------  [60]

hZFY_long     DSVVIQDVVEDVVIEEDVQCSDILEEADVSENVIIPEQVLDSDVTEEVSLPHCTVPDDVL  [120]
hZFY_short    ------------------------------------------------------------  [120]

hZFY_long     ASDITSTSMSMPEHVLTSESMHVCDIGHVEHMVHDSVVEAEIITDPLTSDIVSEEVLVAD  [180]
hZFY_short    ------------------------------------------------------------  [180]
                                              ****************************
hZFY_long     CAPEAVIDASGISVDQQDNDKASCEDYLMISLDDAGKIEHDGSTGVTIDAESEMDPCKVD  [240]
hZFY_short    -------------------------------VDDAGKIEHDGSTGVTIDAESEMDPCKVD  [240]

              ************************************************************
hZFY_long     STCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGVELLDQNSSIRVPREKMVYMTVN  [300]
hZFY_short    STCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGVELLDQNSSIRVPREKMVYMTVN  [300]

              ************************************************************
hZFY_long     DSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAAVHEQQIDEDEMKTFVPIAWAAAY  [360]
hZFY_short    DSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAAVHEQQIDEDEMKTFVPIAWAAAY  [360]

              ************************************************************
hZFY_long     GNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKKKRRPDSRQYQTAIIIGPDGHPLTV  [420]
hZFY_short    GNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKKKRRPDSRQYQTAIIIGPDGHPLTV  [420]

              ************************************************************
hZFY_long     YPCMICGKKFKSRGFLKRHMKNHPEHLAKKKYHCTDCDYTTNKKISLHNHLESHKLTSKA  [480]
hZFY_short    YPCMICGKKFKSRGFLKRHMKNHPEHLAKKKYHCTDCDYTTNKKISLHNHLESHKLTSKA  [480]

              ************************************************************
hZFY_long     EKAIECDECGKHFSHAGALFTHKMVHKEKGANKMHKCKFCEYETAEQGLLNRHLLAVHSK  [540]
hZFY_short    EKAIECDECGKHFSHAGALFTHKMVHKEKGANKMHKCKFCEYETAEQGLLNRHLLAVHSK  [540]

              ************************************************************
hZFY_long     NFPHICVECGKGFRHPSELRKHMRIHTGEKPYQCQYCEYRSADSSNLKTHIKTKHSKEMP  [600]
hZFY_short    NFPHICVECGKGFRHPSELRKHMRIHTGEKPYQCQYCEYRSADSSNLKTHIKTKHSKEMP  [600]

                 ************************************************************
hZFY_long     FKCDICLLTFSDTKEVQQHTLVHQESKTHQCLHCDHKSSNSSDLKRHVISVHTKDYPHKC  [660]
hZFY_short    FKCDICLLTFSDTKEVQQHTLVHQESKTHQCLHCDHKSSNSSDLKRHVISVHTKDYPHKC  [660]

              ************************************************************
hZFY_long     EMCEKGFHRPSELKKHVAVHKGKKMHQCRHCDFKIADPFVLSRHILSVHTKDLPFRCKRC  [720]
hZFY_short    EMCEKGFHRPSELKKHVAVHKGKKMHQCRHCDFKIADPFVLSRHILSVHTKDLPFRCKRC  [720]

              ************************************************************
hZFY_long     RKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEYCKKG  [780]
hZFY_short    RKGFRQQNELKKHMKTHSGRKVYQCEYCEYSTTDASGFKRHVISIHTKDYPHRCEYCKKG  [780]

              *********************
hZFY_long     FRRPSEKNQHIMRHHKEVGLP                                         [801]
hZFY_short    FRRPSEKNQHIMRHHKEVGLP                                         [801]
```

Table 13. Human ZFY (hZFY) spliced variants alignment for 9aa TAD prediction. The sequences were individually used for 9aa TAD prediction using the 9aa TAD prediction tool. As indicated by the names, the long (full length ZFY) and short (lacks half the acidic domain) isoforms of the proteins were used and then aligned using the CLUSTALW tool on the MEGA-X programme. The 9aa TAD motifs are identified by highlights, with the 9aa TAD predictions that displayed 100% match highlighted in yellow and the grey highlight signifying the match of the 9aa TAD prediction being ≥ 67%. RED represents the acidic portion and BLACK represents the remaining zinc finger domain portion of the protein. ***** represents amino acids that are conserved, and **BOLD** represents amino acids not conserved.

Table 13 showed that hZFY-long had three transactivation domains predictions within the acidic domain with perfect matches (100%), whereas the hZFY-short only had one motif with a perfect match, which it shared with hZFY-long between sites 316-324 (`DEVYMEVIV`) of the alignment. However, the other two motifs

(`SVVIQDVVEDVVIE` between sites 62-75 and `SVVEAEIIT` between sites 156-164) that were perfect matches were located between the acidic domain of hZFY-long with 191 additional amino acids. The remaining 9aa TADs predicted were not a perfect score and were therefore not considered in our analysis as we could not be confident due to the criteria for 9aa TAD not being 100%. The percentage identity between the two hZFY protein variant sequences was 76%, owing a reduction in percentage identity largely to the missing half of the acidic activating domain.



```
NP_001356631.1|H_sapiens          -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEIQEAVFVSNIVDSDITVHNFVPDD [60]
XP_009443992.1|P_troglodytes      -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEIQEAVFVSNIVDSDIAVHNFVPDD [60]
Q52V16.1|G_gorilla                -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEIQEAVFVSNIVDSDITVHNFVPDD [60]
XP_014984082.1|M_mulatta          -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_033067617.1|T_francoisi        -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_031516968.1|P_anubis           -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_008017167.1|C_sabaeus          -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_030782172.1|R_roxellana        -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_032612406.1|H_moloch           -MDEDEFELQPQEPNSFFDGIGADATHMDGDQIVVEVQEAVFVSNIVDSDITVHNFVPDD [60]
XP_035145821.1|C_jacchus          -MDEDEFELQPQEPNSFFDGIGAGATHMDGDQIVVEVQETVFVSNIVDSDVTVHNFVPDD [60]
P10925.3|M_musculus               -MDEDEIELTPEEEKSFFDGIGADAVHMDSDQIVVEVQETVFLA---NSDVTVHNFVPDN [60]
P20662.2|M_musculus               -MDEDEIELTPEEEKSLFDGIGADAVHMDSDQISVEVQETVFLS---NSDVTVHNFVPDD [60]
XP_008771898.1|R_norvegicus       -MDEEEIELTPQEENSLFDGIGADAVHMDGDQIIVEVQETVFLS---NSDVTVHNFVPDD [60]
XP_015343506.1|M_marmota          -MDEDEFELQPQEPNSFFDGIGTDSTHMDGDQIVVEVQETVFVS---NSDITVHNFVPDV [60]
Q95LI3.1|B_taurus                 -MDEDEFELQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_010855418.1|B_bison            -MDEDEFELQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_017900383.1|C_hircus           -MDEDELELQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
AMY96563.1|C_elaphus              -MDEDEFEIQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_020759307.1|O_virginianus      -MDEDELEIQPQEPNSCFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
F1SPY3|S_scrofa                   -MDEDELELQPQETNTFFDEIGADDTHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_024612082.1|N_asiaeorientalis  -MDEDELELQPQEPNSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
AKI82174.1|C_lupus                -MDEDELALQPREPNSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_032187800.1|M_erminea          -MDEDELELQPQEPNSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
JAC06687.1|L_africana             -MDEDELELQPQEPNSFFDGIGADVTHMVGDQIVVEVQETVFVSDVVDSDITVHNFPPED [60]
No_accession|E_caballus           -MDEDELELRQQEPDSFFDGIGTDATHMDGDQIVVEVQETVFVSDVVDSDITVHNFAPDG [60]
XP_028935710.1|O_anatinus         -MDEDGLELQPHEPNSFFDATGAAASHMDGGQILVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_016288863.1|M_domestica        -MDEDGLELQPQEPNSFFDATGADATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
XP_015127980.1|G_gallus           -MDEDGLELQPHEPNAFFDPTGADATHMDGDQIVVEVQETVFVSDVVDSDITVHNFVPDD [60]
Q01611.1|X_laevis                 MEDVAELELQTTEPHAFFHASGVGERHLNGNEIIVEIQETVFVADG-DGNMAVQGFGPDE [60]

NP_001356631.1|H_sapiens          PDSVVIQDVVEDVVIEEDVQCSDILEEADVSENVIIPEQVLD-SDVT--------EEVSL [120]
XP_009443992.1|P_troglodytes      PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD-SDVT--------EEVSL [120]
Q52V16.1|G_gorilla                PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLE-SDVT--------EEVSL [120]
XP_014984082.1|M_mulatta          PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD-SDVT--------EEVSL [120]
XP_033067617.1|T_francoisi        PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVVD-SDVT--------EELSL [120]
XP_031516968.1|P_anubis           PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD-SDVT--------EEVSL [120]
XP_008017167.1|C_sabaeus          PDSVVIQDVIEDVVIEEDVQCSDILEEADVSEKVIIPEQVLD-SDVT--------EEVSL [120]
XP_030782172.1|R_roxellana        PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVVD-SDVT--------EEVSL [120]
XP_032612406.1|H_moloch           PDSVVIQDVIEDVVIEEDVQCSDILEEADVSENVIIPEQVLD-SDVT--------EEVSL [120]
XP_035145821.1|C_jacchus          PDSVVIQDVIEDVVIED-VQCSDILEETDVSENVIIPEQVLD-SDVT--------EEVSV [120]
P10925.3|M_musculus               PGSVIIQDVIENVLIED-VHCSHILEETDISDNVIIPEQVLN-LGTA--------EEVSL [120]
P20662.2|M_musculus               PDSVIIQDVIENVLIED-VHCSHILEETDISDNVIIPEQVLD-LDTA--------EEVSL [120]
XP_008771898.1|R_norvegicus       PDSVIIQDVIENVLIED-VHCSNILEETDISDNVIIPEQVLD-LDTA--------EEVSL [120]
XP_015343506.1|M_marmota          QDSVVIQDVIEDVVIED-VQCSDILEEADVSDSVIIPEQVLD-SDVT--------REVSL [120]
Q95LI3.1|B_taurus                 PDSVVIQDVIENVVIED-VQCSDILEEADVSENVIIPEQMLS-SDVT--------EEVSL [120]
XP_010855418.1|B_bison            PDSVVIQDVIENVVIED-VQCSDILEEADVSENVIIPEQMLS-SDVT--------EEVSL [120]
XP_017900383.1|C_hircus           PDSVVIQDVIENVVIED-VQCSDILEEADVSENVIIPEQMLS-SDVT--------EEVSL [120]
AMY96563.1|C_elaphus              PDSVVIQDVIEDVVIED-VQCSDILEEADVSENVIIPEQVLS-SDVT--------EEVSL [120]
XP_020759307.1|O_virginianus      PDSVVIQDVIEDVVIED-VQCSDILEEADVSENVIIPDQVLS-SDVT--------EEVSL [120]
F1SPY3|S_scrofa                   PDSVVIQDVIEDVVIED-VHCSDILEEADVSENVIIPEQVLA-SEVT--------EEVSL [120]
XP_024612082.1|N_asiaeorientalis  PDSVVIQDVIEDVVIED-VQCSDILEEADVSENVIIPEQVLITSDVT--------EEVSL [120]
AKI82174.1|C_lupus                PDSVVIQDVIEDVVIED-VHCSDILEEADVSENVIIPEQVLG-SDVT--------EEVSL [120]
XP_032187800.1|M_erminea          PDSVVIQDVIEDVVIED-VHCSDILEEADISENVIIPEQVLD-SDVT--------EEVSL [120]
JAC06687.1|L_africana             PDSVVIQDVIEDVVIEN-VQCSDILEEADVSENVIIPEQVLE-SDIS--------EEVSL [120]
No_accession|E_caballus           PDSVVIQDVIEDVVIED-VQCSDILEEADVSENVVIPEQVLD-SDVT--------EEVSL [120]
XP_028935710.1|O_anatinus         PDSVVIQDVIEDVVIED-VQCPDILDEADVSETVIIPEPVLG-PEVP--------EEVSL [120]
XP_016288863.1|M_domestica        PDSVVIQDVIEDVVIED-VQCPDIMEEADVSETVIIPEQVLD-TDVT--------EEVSL [120]
XP_015127980.1|G_gallus           PDSVVIQDVIEDVVIED-VQCPDIMEEPDVSETVIIPEQVLD-TDVA--------EEVSL [120]
Q01611.1|X_laevis                 GDSVVIQDVIEDVVIED-VQCSDILDGGRVSEAVIIPEQVLE-DEVGTGEEEQVLEEDSL [120]

NP_001356631.1|H_sapiens          PHCTVPDDVLASDITSTSMSMPEHVLTSES-MHVCDIGHVEHMVHDS---VVEAEIITDP [180]
XP_009443992.1|P_troglodytes      PHCTVPDDVLASDITSTSMSMPEHVLTSES-MHVCDIEHVEHMVHDS---VVEAEIITDP [180]
Q52V16.1|G_gorilla                PHCTVPDDVLASDITSTSTSMPEHVLTSES-MHVCDIGHVEHMVHDS---VVEAEIITDP [180]
XP_014984082.1|M_mulatta          PHCTVPDDVLASDITSASMSMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIITDP [180]
XP_033067617.1|T_francoisi        PHCTVPDDVLASDITSASMSMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIITDP [180]
XP_031516968.1|P_anubis           PHCTVPDDVLASDITSASISMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIITDP [180]
XP_008017167.1|C_sabaeus          AHCTVPDDVLASDITSASMSMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIVTDP [180]
XP_030782172.1|R_roxellana        PHCTVPDDVLASDITSASMSMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIITDP [180]
XP_032612406.1|H_moloch           PHCTVPDDVLASDITSTSMSMPEHVLTSES-MHVCDIGHVEHVVHDS---VVEAEIITDP [180]
XP_035145821.1|C_jacchus          SHCTVPDDVLASDITSSSVSMPEHVLTSES-MHVCDIGHVEHVVRDN---VVEAEIITDP [180]
P10925.3|M_musculus               AQFLIP-DILTSGITSTSLTMPEHVLMSEA-IHVSDVGHFEQVIHDS---LVETEVITDP [180]
P20662.2|M_musculus               AQFLIP-DILTSSITSTSLTMPEHVLMSEA-IHVSNVGHFEQVIHDS---LVEREIITDP [180]
```

```
XP_008771898.1|R_norvegicus         AQFPIP-DILASSITSTSLTMPEHILMSEA-IHVSDVGHIEQVIHDS---LVETEVITDP [180]
XP_015343506.1|M_marmota            AHCTVPDDVLPSDITSTSMSMPEHVLTSES-IHMSNVGHVEHVVHDS---EVEAEIVTDP [180]
Q95LI3.1|B_taurus                   AHCTVPDDVLASDITSASMSMPEHVLTSES-VHVSDVGHVEHIVHGS---VVEAEIVTDP [180]
XP_010855418.1|B_bison              AHCTVPDDVLASDITSASMSMPEHVLTSES-VHVSDVGHVEHIVHGS---VVEAEIVTDP [180]
XP_017900383.1|C_hircus             AHCTVPDDVLASDITSASMSMPEHVLTSES-VHVSDVGHVEHIVHGS---VVEAEIVTDP [180]
AMY96563.1|C_elaphus                AHCTVPDDVLASDVTSASMCMPEHVLTSES-VHVSDVGHVEHIVHDS---VVEAEIVTDP [180]
XP_020759307.1|O_virginianus        AHCTVPDDVLASDVTSASMCMPEHVLTSES-VHVSDVGHVEHIVHDS---VVEAEIVTDP [180]
F1SPY3|S_scrofa                     AHCTVPDDVLASDITSASISMPEQVLTSES-IHVS--EHIEH-IHNS---VVEAEIVTDP [180]
XP_024612082.1|N_asiaeorientalis    AHCTVPDDVLASDITSASMSMPEHVLTSES-IHVSDIGHVEH-VHDS---VVEAEIITDP [180]
AKI82174.1|C_lupus                  AHCTVPDDVLASDITSASMSMPEHVLTSDS-IHVSDVGHVEHVVHDS---VVAAEIITDP [180]
XP_032187800.1|M_erminea            AHCTVPDDVLASDITSASMSVPEHVLTSDS-IHVSDIGHVEHMVHDS---VVEAEIITDP [180]
JAC06687.1|L_africana               THCTVPNDVLASDVTSASMSMPEHVLTHEP-IRVPDVGNVEHVVHDN---VVEAEIVTDT [180]
No_accession|E_caballus             AHCTVPDDVLASDITSASMSMPEHVLTSES-IHVSDVGHVEHIVHDS---VVEAEIVTDP [180]
XP_028935710.1|O_anatinus           AHCAVPEDVLAPDVPAAVAAVPEHVLAGEPVHIPPAAGHVGHVEHVVHDGVVDAEMVADP [180]
XP_016288863.1|M_domestica          AHCTVPDDVLASDITTATMSIPEHVLTSDS-MHVPDVGHVEHVVHDN---VVEAEIVTDP [180]
XP_015127980.1|G_gallus             AHCTVPDDVLASDITAEAMSIPEHVLTSES-MHVPEVGHVEHVVHDN---VEEADIVTDT [180]
Q01611.1|X_laevis                   TSCDVPDNVLDPELVDGELTIPDPETG-----MHSVSGHV----------VIGEEITDDA [180]

NP_001356631.1|H_sapiens            LTSDIVSEEVLVADCAPEAVIDASGISVD--------QQDNDKASCEDYLMISLDDAGKI [240]
XP_009443992.1|P_troglodytes        LTSDIVSEEVLVADCAPEAIIDASGISVD--------QQDNDKASCEDYLMISLDDAGKI [240]
Q52V16.1|G_gorilla                  LTSDIVSEEVLVADCAPEAIIDASGISVD--------QQDNDKASCEDYLMISLDDAGKI [240]
XP_014984082.1|M_mulatta            LTSDVVSEEVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_033067617.1|T_francoisi          LTSDIVS?EVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_031516968.1|P_anubis             LTSDVVSEEVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_008017167.1|C_sabaeus            LTTDVVSEEVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_030782172.1|R_roxellana          LTSDIVSEEVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_032612406.1|H_moloch             LTSDIVSEEVLVADCAPEAIIDASGISVD--------QQDNDKANCEDYLMISLDDAGKI [240]
XP_035145821.1|C_jacchus            LTSDVVSEEVLIADCAPETITDAG-ISVD--------QRDDDKGNCEDYLMISLDDAGKI [240]
P10925.3|M_musculus                 ITADTSD--ILVADCVSEAVLDSSGMPLE--------QQDNDKINCEDYLMMSLDEPSKA [240]
P20662.2|M_musculus                 LTADISD--ILVADWASEAVLDSSGMPLE--------QQDDARINCEDYLMMSLDEPSKT [240]
XP_008771898.1|R_norvegicus         LTADISE--ILVTDCASEAVLDSSGMPLE--------QQDDTKVNRDDYLMISLDDAGKT [240]
XP_015343506.1|M_marmota            LTTNLVS-EVLVADCASEAVIDANGIPVD--------HQDDDKSNCEDYLMISLDDAGKI [240]
Q95LI3.1|B_taurus                   LTADVVSEEVLVADCASEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDDGKM [240]
XP_010855418.1|B_bison              LTADVVSEEVLVADCASEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDDGKM [240]
XP_017900383.1|C_hircus             LTDDVVSEEVLVADCASEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDDGKI [240]
AMY96563.1|C_elaphus                LTANIVSEDVLVADCASEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDDGKM [240]
XP_020759307.1|O_virginianus        LTTNIVSEDVLVADCASEAVIDANGIPVD--------QQNDDKGNCEDYLMISLDDDGKM [240]
F1SPY3|S_scrofa                     LTADVVSEEVLVADCASEAVIDANGIPVD--------QQDGDKSSCEDYLMISLDDAGKI [240]
XP_024612082.1|N_asiaeorientalis    LTTDVVSEEVLVADCASEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDAGKI [240]
AKI82174.1|C_lupus                  LTTDVISEEVLVADCASEAVIDASGIPVE--------QQDDDKNNCEDYLMISLDDAGKI [240]
XP_032187800.1|M_erminea            LTADVVSEEVLVADCASEAVIDANGIPVD--------QQDDDKSNCEDYLMISLDDAGKI [240]
JAC06687.1|L_africana               LTTDIVSEEVLVADCTSEAVIDANGIPVD--------QQDDDKGNCEDYLMISLDDARKL [240]
No_accession|E_caballus             LTTDVVSEEVLVTDCASEAVIDANGIPVE--------QQ-DDKSNCEDYLMISLDDAGKI [240]
XP_028935710.1|O_anatinus           LAAGVVSEEVLVADCASEAVIDANGIPVERRDDDEDDEDDDDKGNCEDYLMISLDDAGKV [240]
XP_016288863.1|M_domestica          LTTDVVSEEVLVADCASEAVIDANGIPVE--------QQDDDKSNCEDYLMISLDDAGKI [240]
XP_015127980.1|G_gallus             LGTDVVSEEVLVADCASEAVIDANGIPVE---------HQDEKGNCEDYLMISLDDAGKI [240]
Q01611.1|X_laevis                   LEEDMISEEVLVADCVSEAVIDANGIPVH--------ENDSEEVNCDDYLMISLDDAEKI [240]

NP_001356631.1|H_sapiens            EHDGSTGVTIDAESEMDPCKVDSTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_009443992.1|P_troglodytes        EHDGSTGVTIDAESEMDPCKVDSTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
Q52V16.1|G_gorilla                  EHDGSTGVTIDAESEMDPCKVDSTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_014984082.1|M_mulatta            EHDGSTGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_033067617.1|T_francoisi          EHDGSTGVTIDGESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_031516968.1|P_anubis             EHDGSTGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_008017167.1|C_sabaeus            EHDGSTGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_030782172.1|R_roxellana          EHDGSTGVTIDGESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_032612406.1|H_moloch             EHDGSTGVTVDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_035145821.1|C_jacchus            EHDGSSGVTIDAESEMDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVENESENDHGV [300]
P10925.3|M_musculus                 DLEGSSEVTMNAESGTDSSKLDEASPEVIKVCILKADSEVDELGETIHAVESETKNGNEA [300]
P20662.2|M_musculus                 DHEGSSEVTMNAESETDSSKLDEASPEVIKVCILKADSEVDDVGETIQAVESETDNGNEA [300]
XP_008771898.1|R_norvegicus         ENEGSSEVTTNAESESDPYKLNETSPEVIKVYIFKADPEEDDVGETVDIVESKTDNGNEA [300]
XP_015343506.1|M_marmota            EHNGSTAVNTSAESDIDSCKVEGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
Q95LI3.1|B_taurus                   EHDCSSGMTMDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_010855418.1|B_bison              EQDCSAGMTIDRESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_017900383.1|C_hircus             EQDCSAGMTIDRESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
AMY96563.1|C_elaphus                EQDCSAGMTIDAESEIDPCKVDGTCPEVIKVYIFRADPGEDDLGGTVDIVESEPENDHGV [300]
XP_020759307.1|O_virginianus        EQDCSAGVTIDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
F1SPY3|S_scrofa                     EHDGSSEMTMDAESEINPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_024612082.1|N_asiaeorientalis    EHDGSSGMTMDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDRGV [300]
AKI82174.1|C_lupus                  EHGGSSGMTIDTESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_032187800.1|M_erminea            EHGGSSGMTMNTESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
JAC06687.1|L_africana               GHDGTSGITMDTESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
No_accession|E_caballus             EQDGSSGMTMDTELEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHEV [300]
XP_028935710.1|O_anatinus           DHDGSSEMTMDAEPEIDPCKVDGGCPEVIKVYIFKADPGEDDLGGTVDIVESEPENEHGV [300]
XP_016288863.1|M_domestica          EHDGSSEITMDAESEIDPCKVDGTCPEVIKVYIFKADPGEDDLGGTVDIVESEPENDHGV [300]
XP_015127980.1|G_gallus             EHEGSAEITMEAESESGSCKVDGICPEVIKVYIFKADPGEDDLGGTVDVVESEPENDHAV [300]
Q01611.1|X_laevis                   DEDGAEEITMGSVVEGDSSKLDGSCPEVIKVYIFKADPGEEDLGGTVDIVESESENDHGD [300]

NP_001356631.1|H_sapiens            ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAA- [360]
XP_009443992.1|P_troglodytes        ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAA- [360]
Q52V16.1|G_gorilla                  ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAA- [360]
XP_014984082.1|M_mulatta            ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAA- [360]
XP_033067617.1|T_francoisi          ELLDQNSSIRVPREKMVYMTVSDSQQED----VAEIADEVYMEVIVGEEDAAVAAAAAA- [360]
XP_031516968.1|P_anubis             ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAA- [360]
XP_008017167.1|C_sabaeus            ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAA- [360]
XP_030782172.1|R_roxellana          ELLDQNSSIRVPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAA- [360]
XP_032612406.1|H_moloch             ELLDQNSSIRVPREKMVYMTVNDSQREDEDLNVAEIADEVYMEVIVGEEDAAVAAAAA-- [360]
XP_035145821.1|C_jacchus            ELLEQSSSVRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAPATVAAAA- [360]
P10925.3|M_musculus                 EVTDQSTSIRVPRV-NIYMSASDSQKEEED-----------TEVIVGDEDAGGTAADTP- [360]
P20662.2|M_musculus                 EVTDQRTSIHVPRV-NIYMLASDSQKEEED-----------TKVIVGDEDAGGTAADTP- [360]
XP_008771898.1|R_norvegicus         EVIDQSSSIYVPRD-NVYMPVSDSQKEEED-----------TKVIVGDEDAGDTAADTS- [360]
XP_015343506.1|M_marmota            ELLDQNSTIRVPREKMVYMTVNDSQQEDEDLNVAEITDEVYMEVIVGEEDAAVTAAAAA- [360]
Q95LI3.1|B_taurus                   ELLDQNNSIRMPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAATT [360]
XP_010855418.1|B_bison              ELLDQNNSIRMPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAATT [360]
XP_017900383.1|C_hircus             ELLDQSNSIRMPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAATT [360]
AMY96563.1|C_elaphus                EILDQNNSIRVPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAAT [360]
XP_020759307.1|O_virginianus        EILDQNNSIRVPREKMVYMTVSDSQQEDEDLNVAEIADEVYMEVIVGEED-AVAAAAAAT [360]
F1SPY3|S_scrofa                     ELLDQNSSMRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAAA [360]
XP_024612082.1|N_asiaeorientalis    ELLDQNSSIRVPREKMVYMTVNDSQQ-DEDLNVAEIADEVYMEVIVGEEDAAVAAAAAAA [360]
```

```
AKI82174.1|C_lupus                    ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAAA [360]
XP_032187800.1|M_erminea              ELLDQNSSIRVPREKMVYMTVNDSQQEDDDLNVAEIADEVYMEVIVGEEDAAVAAAAAAA [360]
JAC06687.1|L_africana                 ELLDQNSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAAA [360]
No_accession|E_caballus               ELLDQNNSIRVPRDKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVAAAAAAA [360]
XP_028935710.1|O_anatinus             GLLDQSSSIRVPREKMVYMTVNDSQQEDEDLSVAEIADEVYMEVIVGEEDAAVA------ [360]
XP_016288863.1|M_domestica            GLLDQSSSIRVPREKMVYMTVNDSQQEDEDLNVAEIADEVYMEVIVGEEDAAVA------ [360]
XP_015127980.1|G_gallus               GLLDQNSSIRIPREKMVYMTVNDSQHEDEDLNVAEIADEVYMEVIVGEEDAAVA------ [360]
Q01611.1|X_laevis                     GFLDSHNGGRLPREKMVYMTVNDSQNDD-DLDVAEIADEVYMEVIVGEEDAAVA------ [360]

NP_001356631.1|H_sapiens              VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_009443992.1|P_troglodytes          VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
Q52V16.1|G_gorilla                    VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_014984082.1|M_mulatta              VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESTGLGRLAKQKPKK [420]
XP_033067617.1|T_francoisi            VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESAGLGRLAKQKPKK [420]
XP_031516968.1|P_anubis               VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESTGLGRLAKQKPKK [420]
XP_008017167.1|C_sabaeus              VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESTGLGRLAKQKPKK [420]
XP_030782172.1|R_roxellana            VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESAGLGRLAKQKPKK [420]
XP_032612406.1|H_moloch               VHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGTASALLHVDESAGLGRLAKQKPKK [420]
XP_035145821.1|C_jacchus              MHEQQIDEDEM-KTFVPIAWAAAYGNNSDGIENRNGSASAVLHVDESVGLSRLTKQKPKK [420]
P10925.3|M_musculus                   EHEQQMDVSEIKAAFLPIAWTAAYDNNSDEIEDQNVTASALLNQDESGGLDRVPKQKSKK [420]
P20662.2|M_musculus                   EHEQQMDVSEIKAAFLPIAWTAAYDNNSDEIEVQNATASAMLHHDESGGLDRVPKQKSKK [420]
XP_008771898.1|R_norvegicus           EHEQQMDDSEIKAAFLPIAWAAAYDNNSDEIEEQNVTASAVLHQNESGGLDRVHKQKAKK [420]
XP_015343506.1|M_marmota              VHEQQIDDSEM-KAFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLSRLAKQKPKK [420]
Q95LI3.1|B_taurus                     VHEQEMDDSEI-KTFMPIAWAAAYGNNSDGIENRSGTASALLHIDESAGLGRLTKHKPKK [420]
XP_010855418.1|B_bison                VHEQEMDDSEI-KTFMPIAWAAAYGNNSDGIENRSGTASALLHIDESAGLGRLTKHKPKK [420]
XP_017900383.1|C_hircus               VHEQEMDDSEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
AMY96563.1|C_elaphus                  VHEQEMDDNEM-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_020759307.1|O_virginianus          VHEQEMDDSEM-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
F1SPY3|S_scrofa                       VHEQQMDDSEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_024612082.1|N_asiaeorientalis      VHEQQMDNSEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
AKI82174.1|C_lupus                    VHEQQMDDNEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_032187800.1|M_erminea              VHEQQMDDNEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLSRLAKQKPKK [420]
JAC06687.1|L_africana                 VHEQQMDDSEI-KTFVPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRVTKQKPKK [420]
No_accession|E_caballus               VHEQQMDDSEI-KTFMPIAWAAAYGNNSDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_028935710.1|O_anatinus             -HEQQMDDTEI-KTFMPIAWAAAYGNNTDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_016288863.1|M_domestica            -HEQQIDDTEI-KTFMPIAWAAAYGNNTDGIENRNGTASALLHIDESAGLGRLAKQKPKK [420]
XP_015127980.1|G_gallus               -HEQQIDDNEI-KTFMPIAWAAAYGNNNDGIESRNGTASALLHIDESAGLGRLAKQKPKK [420]
Q01611.1|X_laevis                     -HEHQLEDAELSKTFMPVAWAAAYGNNTDGIEHRNGTASALLHIDESDGLDRLTKQKLKK [420]

NP_001356631.1|H_sapiens              KRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLA-KKKYHct [480]
XP_009443992.1|P_troglodytes          KRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLA-KKKYHct [480]
Q52V16.1|G_gorilla                    KRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLA-KKKYHct [480]
XP_014984082.1|M_mulatta              KRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLA-KKKYHct [480]
XP_033067617.1|T_francoisi            KRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLA-KKKYHct [480]
XP_031516968.1|P_anubis               KRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLA-KKKYHct [480]
XP_008017167.1|C_sabaeus              KRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLA-KKKYHct [480]
XP_030782172.1|R_roxellana            KRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLA-KKKYHct [480]
XP_032612406.1|H_moloch               KRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLA-KKKYHct [480]
XP_035145821.1|C_jacchus              KRRSDARQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLA-KKKYHct [480]
P10925.3|M_musculus                   KKRPESKQYQSAIFVAPDGQTLRVYPcmfcgkkfkTKRFLKRhtknhPEYLA-NKKYHct [480]
P20662.2|M_musculus                   KKRPESKQYQSAIFVAPDGQTLRVYPcmfcgkkfkTKRFLKRhiknhPEYLA-NKKYHct [480]
XP_008771898.1|R_norvegicus           KKRPESKQYQTAIIVAPDGQTLIVYPcmfcgkkfkTKSFLKRhiknhPEYLA-KKKYHct [480]
XP_015343506.1|M_marmota              RRRPDSKQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHIA-KKKYCct [480]
Q95LI3.1|B_taurus                     RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLT-KKKYRct [480]
XP_010855418.1|B_bison                RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLT-KKKYRct [480]
XP_017900383.1|C_hircus               RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLT-KKKYRct [480]
AMY96563.1|C_elaphus                  RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLT-KKKYRct [480]
XP_020759307.1|O_virginianus          RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLT-KKKYRct [480]
F1SPY3|S_scrofa                       RRRPDSRQYQTAIIIGPDGHPLTVYPclicgkkfkSRGFLKRhmknhPEHLT-KKKYRct [480]
XP_024612082.1|N_asiaeorientalis      RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLT-KKKYHct [480]
AKI82174.1|C_lupus                    RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLS-KKKYRct [480]
XP_032187800.1|M_erminea              RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLT-KKKYRct [480]
JAC06687.1|L_africana                 RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLT-KKKYRct [480]
No_accession|E_caballus               RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLT-KKKYHct [480]
XP_028935710.1|O_anatinus             RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLS-KKKYRct [480]
XP_016288863.1|M_domestica            RRRPDSRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLT-KKKYRct [480]
XP_015127980.1|G_gallus               KRRPESRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLLTKKKYRct [480]
Q01611.1|X_laevis                     KRRGENRQYQTAIIIGPDGHPLTVYPcmicgkkfkSRGFLKRhmknhPEHLV-RKKYRct [480]

NP_001356631.1|H_sapiens              dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
XP_009443992.1|P_troglodytes          dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
Q52V16.1|G_gorilla                    dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
XP_014984082.1|M_mulatta              dcdyttnKKISLHNhleshKLTSKAEK------AIQcdecgkhfsHAGALFThkmvhKEK [540]
XP_033067617.1|T_francoisi            dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
XP_031516968.1|P_anubis               dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
XP_008017167.1|C_sabaeus              dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
XP_030782172.1|R_roxellana            dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
XP_032612406.1|H_moloch               dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
XP_035145821.1|C_jacchus              dcdyttnKKISLHNhleshKLTSKAEK------TIEcvecgkhfsHAGALFThkmvhKEK [540]
P10925.3|M_musculus                   ecdystnKKISLHNhmeshKLTIKTEK------TTEcddcrknlshagtlcthktmhTEK [540]
P20662.2|M_musculus                   ecdystnKKISLHNhmeshKLTIKTEK------TTECDDCRKNLSHAG------TMHTEK [540]
XP_008771898.1|R_norvegicus           dcdystnKKISLHNhmeshKLTIKTEK------TTEcddcgkhlshagtlcthk---kek [540]
XP_015343506.1|M_marmota              dcdyttnKKISLHNhleshKLTSKVEK------VIEcdecgkhfsHTGALFThkmvhKEK [540]
Q95LI3.1|B_taurus                     dcdyttnKKISLHNhleshKLTSKSEK------AIEcddcgkhfsHAGALFThkmvhKEK [540]
XP_010855418.1|B_bison                dcdyttnKKISLHNhleshKLTSKSEK------AIEcddcgkhfsHAGALFThkmvhKEK [540]
XP_017900383.1|C_hircus               dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
AMY96563.1|C_elaphus                  dcdyttnKKMSLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
XP_020759307.1|O_virginianus          dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
F1SPY3|S_scrofa                       dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
XP_024612082.1|N_asiaeorientalis      acdyttnkkisfhnhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
AKI82174.1|C_lupus                    dcdyttnKKISLHNhleshKLTSKAEK------SIEceecgkhfsHAGALFThkmvhKEK [540]
XP_032187800.1|M_erminea              dceyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
JAC06687.1|L_africana                 dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHTGALFThkmvhKEK [540]
No_accession|E_caballus               dcdyttnKKISLHNhleshKLTSKAEK------AIEcdecgkhfsHAGALFThkmvhKEK [540]
XP_028935710.1|O_anatinus             dcdyttnKKVSLHNhleshKLTGKAEKAAAPGAGAEcdecgkhfsHAGALFThktvhKEK [540]
XP_016288863.1|M_domestica            dcdyttnKKISLHNhleshKLTNKTEKAIEc------decgkhfsHAGALFThkmvhKEK [540]
XP_015127980.1|G_gallus               dcdyttnKKISLHNhleshKLTNKTEKLIEr------decgksfshagalfahkmvhRDK [540]
Q01611.1|X_laevis                     dcdyttnKKVSLHNhleshKLTATVIKTEKD---LEceecgkiflHANALFAhklthNEK [540]
```

```
NP_001356631.1|H_sapiens            -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELRKhmrihTG [600]
XP_009443992.1|P_troglodytes        -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
Q52V16.1|G_gorilla                  -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_014984082.1|M_mulatta            -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_033067617.1|T_francoisi          -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_031516968.1|P_anubis             -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_008017167.1|C_sabaeus            -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_030782172.1|R_roxellana          -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_032612406.1|H_moloch             -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_035145821.1|C_jacchus            -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
P10925.3|M_musculus                 G-VNKTCKckfcdyetaeqtllnhhllvvhRKKFPHIcgecgkgfrHPSALKKhirvhTG [600]
P20662.2|M_musculus                 G-VNKTCKckfcdyetaeqtllnhhllvvhRKKFPHIcgecgkgfrHPSALKKhirvhTG [600]
XP_008771898.1|R_norvegicus         EKVSKTYKckfcdyetaEQTSLNHhllavhSKKYPHVcvecgkgfrHPSELKKhirvhTG [600]
XP_015343506.1|M_marmota            -GNNKMHKckfceyetaEQGLLNRhllavhSKSFPHIcvecgkgfrHPSELKKhmrihTG [600]
Q95LI3.1|B_taurus                   -GASKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_010855418.1|B_bison              -GASKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_017900383.1|C_hircus             -GASKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELRKhmrihTG [600]
AMY96563.1|C_elaphus                -GANKMHKckfceygtaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_020759307.1|O_virginianus        -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
F1SPY3|S_scrofa                     -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_024612082.1|N_asiaeorientalis    -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
AKI82174.1|C_lupus                  -GTNKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_032187800.1|M_erminea            -GANKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
JAC06687.1|L_africana               -GSSKMHKckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
No_accession|E_caballus             -GANKMHRckfceyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_028935710.1|O_anatinus           GAGGRTHKckfcdyetaEQGLLNRhllavhSKNFPHVcvecgkgfrHPSELKKhmrihTG [600]
XP_016288863.1|M_domestica          -GANKMHKckfcdyetaEQGLLNRhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
XP_015127980.1|G_gallus             -GVNKMHKckfcdyetaeqgllshhllavhSKNFPHIcvecgkgfrHPSELKKhmrihTG [600]
Q01611.1|X_laevis                   -AGNKMHIckfcdyetaEQGLLNRhllavhSKSFPHIcvecgkgfrHPSELKKhmrthTG [600]

NP_001356631.1|H_sapiens            EKPYQcqyceyrsaDSSNLKThiktkhSKEMPFKcdiclltfsDTKEVQQhtlvhQ-ESK [660]
XP_009443992.1|P_troglodytes        EKPYQcqyceyrsaDSSNLKThiktkhSKEMPLKcdiclltfsDTKEVQQhtlvhQ-ESK [660]
Q52V16.1|G_gorilla                  EKPYQcqyceyrsaDSSNLKThiktkhSKEMPFKcdiclltfsDTKEVQQhtlvhQ-ESK [660]
XP_014984082.1|M_mulatta            EKPYQcqyceyrsaDSSNLKThiktkhSKEMPFKcdiclltfsDTKEVQQhtlvhQ-ENK [660]
XP_033067617.1|T_francoisi          EKPYQcqyceyrsaDSSNLKThiktkhSKEMPFKcdiclltfsDTKEVQQhtlvhQ-ENK [660]
XP_031516968.1|P_anubis             EKPYQcqyceyrsaDSSNLKThiktkhSKEMPFKcdiclltfsDTKEVQQhtlvhQ-ESK [660]
XP_008017167.1|C_sabaeus            EKPYQcqyceyrsaDSSNLKThiktkhSKEMPFKcdiclltfsDTKEVQQhalihQ-ESK [660]
XP_030782172.1|R_roxellana          EKPYQcqyceyrsaDSSNLKThiktkhSKEMPFKcdiclltfsDTKEVQHhtlvhQ-ENR [660]
XP_032612406.1|H_moloch             EKPYQcqyceyrsaDSSNLKThiktkhSKEMPFKcdiclltfsDTKEVQQhtlvhQ-ESK [660]
XP_035145821.1|C_jacchus            EKPYQcqyceyrsaDSSNLKThiktkhSKEMPFKcdiclltfsDTKEVQQhtlvhQ-ESR [660]
P10925.3|M_musculus                 EKPYEcqyceyksaDSSNLKThikskhSKEIPLKcgiclltfsdnkeaqqhavlhQ-ESR [660]
P20662.2|M_musculus                 EKPYEcqyceyksaDSSNLKThikskhSKEIPLKcgiclltfsdtkeaqqhavlhQ-ESR [660]
XP_008771898.1|R_norvegicus         EKPYQcqyceyksaDSSNLKThiktkhSKDIPLKcgiclmtfsdtkeaqqhalihQ-ENR [660]
XP_015343506.1|M_marmota            EKPYQcqyceyrsaDSSNLKThvktkhTKEMPFKcdiclltfsDTKEVQQhalihQ-ESK [660]
Q95LI3.1|B_taurus                   EKPYQcqyceyrsaDSSNLKThvktkhSKEMSFKcdiclltfsDTKEVQQhalihQ-ESK [660]
XP_010855418.1|B_bison              EKPYQcqyceyrsaDSSNLKThvktkhSKEMSFKcdiclltfsDTKEVQQhalihQ-ESK [660]
XP_017900383.1|C_hircus             EKPYQcqyceyrsaDSSNLKThvktkhSKEMSFKcdiclltfsDTKEVQQhalihQ-ESK [660]
AMY96563.1|C_elaphus                EKPYRcqyceyrsaDSSNLKThvktkhSKEMSFKcdiclltfsDTKEVQQhalihQ-ESK [660]
XP_020759307.1|O_virginianus        EKPYQcqyceyrsaDSSNLKThvktkhSKEMSFNcdiclltfsDTKEVQQhalihQ-ESK [660]
F1SPY3|S_scrofa                     EKPYQcqyceyrsaDSSNLKThvktkhSKEMPFKcdiclltfsDTKDVQQhalihQ-ESK [660]
XP_024612082.1|N_asiaeorientalis    EKPYQcqyceyrsaDSSNLKThvktkhSKEMPFKcdsclltfsDTKEVQQhalihQ-ESK [660]
AKI82174.1|C_lupus                  EKPYQcqyceyrsaDSSNLKThvktkhSKEMPFKcdiclltfsDTKEVQQhavihQ-ESK [660]
XP_032187800.1|M_erminea            EKPYQcqyceyrsaDSSNLKThvktkhSKEMPFKcdiclltfsDTKEVQQhalihQ-ESK [660]
JAC06687.1|L_africana               EKPYQcqyceyrsaDSSNLKThvktkhSKEMPYRcdiclltfsDTKEVQQhavihQ-ESK [660]
No_accession|E_caballus             EKPYHcqyceyrsaDSSNLKThvktkhSKEMPFKcdiclltfsDTKEVQQhglihQ-ESK [660]
XP_028935710.1|O_anatinus           EKPYQcqfcpyrsaDSSNLKThvktkhSKETPFRceacpltfaDPKELQQhallhHQESR [660]
XP_016288863.1|M_domestica          EKPYQcqyceyrsaDSSNLKThvktkhSKEMPFKceiclltfsDTKEVQQhalihQ-ESK [660]
XP_015127980.1|G_gallus             EKPYQcqyceyrsaDSSNLKThvktkhSKETSSKcdicfqtfsDTKELQQhtlmhQ-ESK [660]
Q01611.1|X_laevis                   EKPYLcqycdyrsaDSSNLKThvktkhSKEMPFKcdiclqtftDSKDLQEhailhQ-ESK [660]

NP_001356631.1|H_sapiens            THQclhcdhkssNSSDLKRhvisvhTKDYPHKcemcekgfhRPSELKKhvavhKGKKMHQ [720]
XP_009443992.1|P_troglodytes        THQclhcdhkssNSSDLKRhvisvhTKDYPHKcemcekgfhRPSELKKhvavhKGKKMHQ [720]
Q52V16.1|G_gorilla                  THQclhcdhkssNSSDLKRhvisvhTKDYPHKcemcekgfhRPSELKKhvavhKGKKMHQ [720]
XP_014984082.1|M_mulatta            THQclhcdhkssNSSDLKRhvisvhTKDYPHKcemcekgfhRPSELKKhvavhKGKKMHQ [720]
XP_033067617.1|T_francoisi          THQclhcdhkssNSSDLKRhvisvhTKDYPHKcemcekgfhRPSELKKhvavhKGKKMHQ [720]
XP_031516968.1|P_anubis             THQclhcdhkssNSSDLKRhvisvhTKDYPHKcemcekgfhRPSELKKhvavhKGKKMHQ [720]
XP_008017167.1|C_sabaeus            THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
XP_030782172.1|R_roxellana          THQclhcdhkssNSSDLKRhvisvhTKDYPHKcemcekgfhRPSELKKhvavhKGKKMHQ [720]
XP_032612406.1|H_moloch             THQclhcdhkssNSSDLKRhvisvhTKDYPHKcemcekgfhRPSELKKhvavhKGKKMHQ [720]
XP_035145821.1|C_jacchus            THQclhcdhkssNSSDLKRhvisvhTKDYPHKcemcdkgfhRPSELKKhvavhKGKKMHQ [720]
P10925.3|M_musculus                 THQcshcnhkssNSSDLKRhiisvhTKAYPHKcdmcskgfhRPSELKKhvathKSKKMHQ [720]
P20662.2|M_musculus                 THQcshcnhkssNSSDLKRhiisvhTKAYPHKcdmcskgfhRPSELKKhvathKSKKMHQ [720]
XP_008771898.1|R_norvegicus         THQcsycnhkssNSSDLKRhiisvhTKDYPHKcdmcskgfhRPSELKKhvathKSKKMHQ [720]
XP_015343506.1|M_marmota            THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
Q95LI3.1|B_taurus                   THQcvhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
XP_010855418.1|B_bison              THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
XP_017900383.1|C_hircus             THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
AMY96563.1|C_elaphus                THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
XP_020759307.1|O_virginianus        THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
F1SPY3|S_scrofa                     THQclhcdhkssNSSDLKRhiisvhTKDYPHKcemcekgfhRPSELKKhvaahKGKKMHQ [720]
XP_024612082.1|N_asiaeorientalis    THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
AKI82174.1|C_lupus                  THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
XP_032187800.1|M_erminea            THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
JAC06687.1|L_africana               THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
No_accession|E_caballus             THQclhcdhkssNSSDLKRhiisvhTKDYPHKcemcdkgfhRPSELKKhvaahKGKKMHQ [720]
XP_028935710.1|O_anatinus           AHQclhcdhkssNSSDLKRhvisvhTKDYPHKcdtcdkgfhRPSELKKhaaahRGRKLHQ [720]
XP_016288863.1|M_domestica          THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKMHQ [720]
XP_015127980.1|G_gallus             THQclhcdhkssNSSDLKRhiisvhTKDYPHKcdmcdkgfhRPSELKKhvaahKGKKLHQ [720]
Q01611.1|X_laevis                   NHQclhcdhkssNSSDLKRhiisvhTKDYPHKcevcekgfhRPSELKKheaahKGKKMHQ [720]


NP_001356631.1|H_sapiens            crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_009443992.1|P_troglodytes        crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
Q52V16.1|G_gorilla                  crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_014984082.1|M_mulatta            crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_033067617.1|T_francoisi          crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_031516968.1|P_anubis             crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_008017167.1|C_sabaeus            crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_030782172.1|R_roxellana          crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
```

```
XP_032612406.1|H_moloch             crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_035145821.1|C_jacchus            crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfnQQNELKKhmkthSGRKVYQcey [780]
P10925.3|M_musculus                 crhcdfnspdpfllshhilsahTKNVPFKckrckkefqQQCELQThmkthSSRKVYQcey [780]
P20662.2|M_musculus                 crhcdfkspdpfllshhilsahTKNVPFKckrckkefqQQCELQThmkthSSRKVYQcey [780]
XP_008771898.1|R_norvegicus         crhcdfkspdpfllsrhilsvhTKNVPFKckrckkgfrQQCELQKhmkthSGRKVYQcey [780]
XP_015343506.1|M_marmota            crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
Q95LI3.1|B_taurus                   crhcdfkiadpfvlsrhilsvhTKDLPFRckrckkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_010855418.1|B_bison              crhcdfkiadpfvlsrhilsvhTKDLPFRckrckkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_017900383.1|C_hircus             crhcdfkiadpfvlsrhilsihTKDLPFRckrckkgfrQQNELKKhmkthSGRKVYQcey [780]
AMY96563.1|C_elaphus                crhcdfriadpfvlsrhilsvhAKDLPFRckrckkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_020759307.1|O_virginianus        crhcdfkiadpfvlsrhilsvhTKDLPFRckrckkgfrQQNDLKKhmkthSGRKVYQcey [780]
F1SPY3|S_scrofa                     crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_024612082.1|N_asiaeorientalis    crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
AKI82174.1|C_lupus                  crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_032187800.1|M_erminea            crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
JAC06687.1|L_africana               crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
No_accession|E_caballus             crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQTELKKhmkthSGRKVYQcey [780]
XP_028935710.1|O_anatinus           crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQGELKKhmkthSGRKVYQcey [780]
XP_016288863.1|M_domestica          crhcdfkiadpfvlsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
XP_015127980.1|G_gallus             crhcdfkiadpfilsrhilsvhTKDLPFRckrcrkgfrQQNELKKhmkthSGRKVYQcey [780]
Q01611.1|X_laevis                   crhcefhiadpfvlsrhilsvhTKELPYRckrckkgfrQQIELKKhmkthSGKKVYQcey [780]

NP_001356631.1|H_sapiens            ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
XP_009443992.1|P_troglodytes        ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
Q52V16.1|G_gorilla                  ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
XP_014984082.1|M_mulatta            ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
XP_033067617.1|T_francoisi          ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
XP_031516968.1|P_anubis             ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
XP_008017167.1|C_sabaeus            ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
XP_030782172.1|R_roxellana          ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
XP_032612406.1|H_moloch             ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
XP_035145821.1|C_jacchus            ceynttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
P10925.3|M_musculus                 ceystkDASGFKRhvisihTKDYPHScdfckkgfrRPSEKNQhimrhHK-VGLP [834]
P20662.2|M_musculus                 ceystkDASGFKRhvisihTKDYPHRcdfckkgfrRPSEKNQhimrhHKEVGLA [834]
XP_008771898.1|R_norvegicus         ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
XP_015343506.1|M_marmota            ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVSLS [834]
Q95LI3.1|B_taurus                   ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhitrhHKEVGLP [834]
XP_010855418.1|B_bison              ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhitrhHKEVGLP [834]
XP_017900383.1|C_hircus             ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhitrhHKEVGLP [834]
AMY96563.1|C_elaphus                ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhitrhHKEVGLP [834]
XP_020759307.1|O_virginianus        ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhitrhHKEVGLP [834]
F1SPY3|S_scrofa                     ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
XP_024612082.1|N_asiaeorientalis    ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
AKI82174.1|C_lupus                  ceysttDASGFKRhvisihTKDYPHRSS-------------------------- [834]
XP_032187800.1|M_erminea            ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
JAC06687.1|L_africana               ceysttDASGFKRhvisihTKDYPHRcehckkgfrRPSEKNQhimrhHKEVGLP [834]
No_accession|E_caballus             ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKEVGLP [834]
XP_028935710.1|O_anatinus           ceysttDASGFKRhvisihTKDYPHRcdfckkgfrRPSEKNQhimrhHKDLGLP [834]
XP_016288863.1|M_domestica          ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKDVGLP [834]
XP_015127980.1|G_gallus             ceysttDASGFKRhvisihTKDYPHRceyckkgfrRPSEKNQhimrhHKDVGLP [834]
Q01611.1|X_laevis                   ceynttDASGFKRhvisihTKDYPHRcdyckkgfrRPSEKNQhtlkhHKEASLM [834]
```

Table 14. Multiple CLUSTALW sequence alignment of ZFY outlining the 9 amino acid transactivation domains and DNA binding sites predictions. **(A)** The 9aa TAD motifs are identified by highlights, with the 9aa TADs that displayed 100% match highlighted in yellow and the grey highlight signifying the match of the 9aa TAD being ≥ 67%. **(B)** The DNA binding sites are indicated by coloured upper/lowercase text; ZF1 = GREEN, ZF2 = RED, ZF3 = MAGENTA, ZF4 = ORANGE, ZF5 = BROWN, ZF6 = CYAN, ZF7 = PURPLE, ZF8 = SAPPHIRE, ZF9 = OLIVE, ZF10 = GREY,ZF11 = LAVENDER, ZF12 = PEACH, ZF13 = DARK RED. The predicted sequence responsible for zinc finger binding is indicated by **`UPPERCASE LETTERS`** within the zinc finger DNA binding site predictions.

Table 14 showed that four regions within the ZFY protein were consistently predicted to have 9 amino acid TADs throughout the majority of the species, and three of the four regions were located exon 2 and the other region located on exon 5. These four regions contained perfect matches throughout most of the species as indicated by the highlight in **Table 14**. Though, *H. sapiens*, majority of the primates, rodents and amphibians do not have the 9aa TAD predicted in other species between sites 40-48 because they do not meet criteria RC4 & RC5 and also RC6 & RC7 as there were four consecutive hydrophobic residues and no threonine or serine present. However, other species such as *C. jacchus* meet every criteria apart from possessing three consecutive hydrophilic residues between sites 41-43, so it was predicted to be a perfect match for a 9aa TAD (t**VFV**sn**IV**d) along with the other species that have perfect match prediction. The next region was between residue sites 63-76 (s**VVI**qd**VV**ed**VVI**e) and what we observed was that this region had three different perfect matches, but the arrangement of the 9aa TAD sequences would be different. For example, in *H. sapiens*, the sequence arrangements for the motifs were SVVIQDVVE, QDVVEDVVI and DVVEDVVIE, but the program conjoined these and they produced a 14 amino acid TAD. This pattern remained consistent in the rest of the species between those specific amino acid sites and the majority of the species contained either a doublet or triplet di-valine cluster.

The conservation of the 9 amino acid TAD was different for each section as the first cluster between residue sites 40-48 had a 78% identity (7 of 9) in species that possessed the 9aa TAD for this particular region, and the second cluster of 9 amino acids TAD had a percentage identity of 71% (10 of 14). Additionally, the third cluster between sites 167-178 (gap between 168-170) had the lowest percentage identity of 44% (4 of 9), and this region is also where *O. anatinus* has 3 more amino acids that the rest of the species. The final cluster of residues between sites 338-346 has a 100% identity and interestingly, in the rodents *M. musculus* and *R. norvegicus* there was a gap of 4 amino acids in the same region which indicated some divergence within this region which explained the lack of a 9aa TAD prediction in this region.

Table 14 also shows the zinc finger predictions, with the majority of the species containing 13 zinc fingers on the seventh exon and only *M. musculus* and *C. lupus* having 11 zinc fingers because of a missing portion of the sequence of zinc finger

3 and also because of an incomplete sequence of the zinc finger 13. The predicted zinc fingers contained the classic C2H2 characteristics. For instance, the H. sapiens zinc finger 1 has the sequence **C**MI**C**GKKFKSRGFLKR**H**MKN**H** and as outlined by the bold letters, the cysteine and histidine residues are present. However, the first residue of predicted zinc finger 3 of *G. gallus* did not have a cysteine and contained an arginine residue instead. It was also observed that the zinc finger 3 was the least conserved zinc finger as it had only a percentage identity of 20%. Furthermore, this zinc finger region was missing 6 amino acids in *M. musculus* Zfy2 and *R. norvegicus* was missing 3 amino acids. Another zinc finger with a low percentage identity was zinc finger 7 as it had a percentage identity of 38%. Nevertheless, the remainder of the zinc fingers (ZF1-ZF2, ZF4-ZF6 and ZF8-ZF13) contained a percentage identity $\geq$ 64% with zinc finger 12 having the highest conservation as it had 91% identity.

Another interesting observation was that Krüppel-type C2H2 zinc fingers usually have the conserved linker sequence TGEKPY and in all this can be seen in all of the species in **Table 9**,**Table 11,** and **Table 14**. In addition, TKDYPH(R/K) (**Table 14**) were also conserved sequences seen frequently as zinc finger linker regions (between sites 686-692 and 800-806) contained this sequence across most of the species, only differing in the last amino acid as one contained arginine (R) and the other lysine (K). The rest of the linker regions between the zinc fingers had no conserved regions or patterns.

	Zinc Finger (ZF) Scores												
Species	**ZF1**	**ZF2**	**ZF3**	**ZF4**	**ZF5**	**ZF6**	**ZF7**	**ZF8**	**ZF9**	**ZF10**	**ZF11**	**ZF12**	**ZF13**
H. sapiens	31.7	20.3	27.5	18.3	35.1	29.5	21.3	19.5	27.6	16.0	30.8	25.5	25.6
P. troglodytes	31.7	20.3	27.5	18.3	35.4	29.5	18.2	19.5	27.6	16.0	30.8	25.5	25.6
G. gorilla	31.7	20.3	27.5	18.3	35.4	29.5	21.3	19.5	27.6	16.0	30.8	25.5	25.6
M. mulatta	31.7	20.3	27.8	18.3	35.4	29.5	21.3	19.5	27.6	16.0	30.8	25.5	25.6
T. francoisi	31.7	20.3	27.5	18.3	35.4	29.5	21.3	19.5	27.6	16.0	30.8	25.5	25.6
P. anubis	31.7	20.3	27.5	18.3	35.4	29.5	21.3	19.5	27.6	16.0	30.8	25.5	25.6
C. sabaeus	31.7	20.3	27.5	18.3	35.4	29.5	21.1	20.8	27.9	16.0	30.8	25.5	25.6
R. roxellana	31.7	20.3	27.5	18.3	35.4	29.5	20.7	19.5	27.6	16.0	30.8	25.5	25.6
H. moloch	31.7	20.3	27.5	18.3	35.4	29.5	21.3	19.5	27.6	16.0	30.8	25.5	25.6
C. jacchus	31.7	20.3	27.5	18.3	35.4	29.5	21.3	19.5	29.3	16.0	30.7	23.8	25.6
M. musculus (ZFY1)	24.7	20.9	15.5	15.8	35.6	27.8	13.8	21.6	27.1	14.7	24.0	25.3	22.9
M. musculus (ZFY2)	27.5	20.9	(-)[§]	15.8	35.6	27.8	12.5	21.6	27.1	14.7	24.0	25.3	23.5
R. norvegicus	30.4	21.7	4.1	19.9	35.4	29.0	15.5	23.3	27.1	17.4	27.8	25.5	25.6
M. marmota	31.7	18.8	26.2	18.3	35.4	28.2	21.1	20.8	27.9	16.0	30.8	25.5	25.6
B. taurus	31.7	23.4	28.2	18.3	35.4	28.2	21.1	21.3	27.9	16.0	31.4	25.5	24.3
B. bison	31.7	23.4	28.2	18.3	35.4	28.2	21.1	20.8	27.9	16.0	31.4	25.5	24.3
C. hircus	31.7	23.4	27.5	18.3	35.1	28.2	21.1	20.8	27.9	15.7	31.4	25.5	24.3
C. elaphus	31.7	24.0	27.5	18.0	35.4	27.7	21.1	20.8	27.9	16.5	31.4	25.5	24.3
O. virginianus	31.7	23.4	27.5	18.3	35.4	28.2	18.8	20.8	27.9	16.0	31.5	25.5	24.3
S .scrofa	32.0	23.4	27.5	18.3	35.4	28.2	21.1	20.8	26.8	16.0	30.8	25.5	25.6
N. asiaeorientalis	31.7	14.2	27.5	18.3	35.4	28.2	18.9	20.8	27.9	16.0	30.8	25.5	25.6
C. lupus	31.7	23.4	28.0	18.3	35.4	28.2	19.7	20.8	27.9	16.0	30.8	25.5	N/A[**]
M. erminea	31.7	21.7	27.5	18.3	35.4	28.2	21.1	20.8	27.9	16.0	30.8	25.5	25.6
L. africana	31.7	23.4	26.2	18.3	35.4	28.2	19.8	20.8	27.9	16.0	30.8	25.5	23.9
E. caballus	31.7	20.3	27.5	18.0	35.4	24.7	19.7	20.8	28.5	16.0	31.2	25.5	25.6
O. anatinus	31.7	24.2	23.5	19.9	36.2	27.1	22.2	19.5	26.3	16.0	32.2	25.5	23.5
M. domestica	31.7	23.4	27.5	19.9	35.4	28.2	21.7	20.8	27.9	16.0	30.8	25.5	25.6
G. gallus	31.7	23.4	13.2	16.6	35.4	28.2	21.2	20.8	27.9	11.7	30.8	25.5	25.6
X. laevis	31.7	24.2	23.0	18.6	35.9	25.7	25.6	20.8	27.0	13.0	31.4	23.8	22.1

Table 15. Zinc finger scores of each individual zinc finger domain. The table illustrates the zinc finger scores of the 13 predicted zinc fingers. As shown, *Mus musculus* has a missing zinc finger 3 which is indicated by the symbol (-), and *Canis lupus* sequence is incomplete at the end which signifies that the database has the incomplete partial sequence and not the complete full sequence, and this is indicated by N/A. The HMMER algorithm was used to detect the zinc fingers and the ZF or HMMER score is ‘most confident’ when > 17.7. Values lower are considered not to be a confident result. We created subcategories for confidence to distinguish significant results from moderately significant results. Therefore: GREEN = Most Confident ($\geq$ 30.0), ORANGE = Fairly Confident (20.0- 29.9) , BLACK = Least Confident (17.8 -19.9) and RED = No Confidence (< 17.7).

Table 15 showed zinc fingers 1, 5-6, 8-9 and 11-13 had a ZF score value above 17.7 in almost all of the land vertebrates. A few anomalies were observed in zinc fingers 2-4 and 7 of some species as their ZF scores were below 17.7. Though zinc finger 10 was predicted in all of the species initially shown by **Table 14**, the

---

[§] *M. musculus* Zfy2 shows that there was no predicted zinc finger 3 due to the truncation within that region of the sequence

[**] *C. lupus* ZFY shows an incomplete sequence, thus the reason zinc finger 13 was not predicted

ZF scores of all the land vertebrates were below 17.7, thus highlighting that though some of the criteria was met by the sequence, it could not confidently be considered a zinc finger. Therefore, we concluded the majority of ZFY sequences had roughly 12 zinc fingers in total (excluding *G. gallus*, *N. asiaeorientalis, R. norvegicus, M. musculus* and *C. lupus*).

3.5 ZFY domain binding

To investigate the possible binding motifs of each zinc finger, we analysed the predicted binding target of each using the B1H screens Nearest Neighbour Prediction tool using the F2F3union data. This tool predicted that a pattern could be established as the majority of the land vertebrates had the same trinucleotide target. For example, as **Table 16** shows zinc fingers 2,6,8, 9, 12 and 13 all had a 100% match for the trinucleotide target per zinc finger which indicates that these trinucleotides are highly likely to be the targets for these specific zinc fingers. However, we noticed instances where *M. musculus* Zfy1 and Zfy2*,* and *R. norvegicus* zinc fingers 1, 3, 4, 5, 7 and 11 had either predicted different trinucleotide targets, no predicted targets or the trinucleotides could be anything as the score for each amino acid was equal. In addition, as indicated by **Table 16**, *R. roxellana, N. asiaeorientalis* and *S. scrofa* had one zinc finger each which contained either a different trinucleotide or no predicted trinucleotide at all, and *G. gallus* and *X. laevis* had two zinc fingers with either different or no predicted trinucleotides. The prediction tool result for zinc finger 10 showed that it had no predicted trinucleotide target for all land vertebrates which corresponded with **Table 15** as this zinc finger had a weak HMMER score, further confirming this zinc finger could not be confirmed as a zinc finger.

	Predicted Trinucleotide Target												
Species	ZF1	ZF2	ZF3	ZF4	ZF5	ZF6	ZF7	ZF8	ZF9	ZF10	ZF11	ZF12	ZF13
H. sapiens	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
P. troglodyte	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
G. gorilla	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
M. mulatta	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
T. francoisi	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
P.anubis	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
C. sabaeus	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
R. roxellana	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	GGA	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
H. moloch	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
C. jacchus	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
M. musculus ZFY1	ATC	**GGC**	N/A	N/A	ATC	**CAC**	N/A	**GCC**	**GTG**	N/A	XXX	**GGC**	**ATG**
M. musculus ZFY2	ATC	**GGC**	(-)[††]	N/A	ATC	**CAC**	N/A	**GCC**	**GTG**	N/A	XXX	**GGC**	**ATG**
R. norvegicus	GTT	**GGC**	N/A	AAC	**C(G/T)C**	**CAC**	N/A	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
M. marmota	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
B. taurus	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
B. bison	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
C. hircus	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
C. elaphus	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
O. virginianus	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	XXX	**GGC**	**ATG**
S. scrofa	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	CGT	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
N. asiaeorientalis	**GXC**	N/A	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
C. lupus	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	(-)[‡‡]
M. erminea	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
L. africana	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
E. caballus	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
O. anatinus	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
M. domestica	**GXC**	**GGC**	**ATA**	**GCC**	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
G. gallus	**GXC**	**GGC**	N/A	N/A	**C(G/T)C**	**CAC**	XXX	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**
X. laevis	**GXC**	**GGC**	CTC	**GCC**	**C(G/T)C**	**CAC**	XGT	**GCC**	**GTG**	N/A	**GCC**	**GGC**	**ATG**

Table 16. ZFY domain trinucleotide prediction. ZFY domain binding partners are yet to be properly elucidated, but the prediction was made using the domain analysis via the Nearest Neighbour Prediction algorithm and using the F2F3 union data. The majority of the species have 12 zinc finger target predictions, but some species have less than 12 predictions. X indicates that the nucleotide could be any one of the four nucleotides. N/A indicates that there was no prediction at all made by the algorithm. YELLOW highlight outlines zinc fingers that have different trinucleotides to the main pattern shown, **BOLD** = conserved, and GREEN =100% conserved trinucleotides in land vertebrates. (-) indicates that either the zinc finger is absent either because of truncation or incomplete sequence.

The DNA targets shown by **Table 16** interact with the DNA binding interfaces predicted in **Table 15** and **Table 17**. **Table 17** indicated 50% of the predicted DNA binding interfaces (6 out of 12 [excluding ZF10]) across all species had 100% identity located on ZF2,ZF6, ZF8, ZF9, ZF12, and ZF13 which showed that these amino acids are well conserved throughout the evolution from the common ancestor, whereas 50% of the DNA binding interfaces had identities varying from 25-75% as ZF1,ZF3-ZF5, ZF7 and ZF11 had slight variation of amino acids at

---

[††] *M. musculus* Zfy2 shows that there was no predicted trinucleotide target as there was no predicted zinc finger 3 due to the truncation within that region of the sequence

[‡‡] *C. lupus* ZFY shows an absence of a predicted trinucleotide target as the sequence was incomplete and there was no zinc finger 13 predicted

canonical positions across all species. Although, 83% of the zinc fingers (10 out of 12 [excluding ZF10]) DNA binding interfaces had identities of 50-100%, which shows that the fingers were mostly conserved.

The general consensus from **Table 17** was that rodent DNA binding interfaces exhibited the most variation as four of the zinc fingers (ZF1, ZF4,ZF5 and ZF11) had differing amino acids at certain canonical positions. As expected, ZF10 had no DNA binding interface prediction as this particular zinc finger prediction in **Table 15** had the lowest HMMER score and **Table 16** had no predicted target.

Table 17. Predicted binding interface of each zinc finger. The table illustrates the DNA binding interface of ZF1-ZF13, which consists of 4 amino acids that are found on the helix of the zinc fingers. These 4 'canonical' positions that interact with the DNA target are in positions -1, 2, 3 and +6 relative to the zinc finger helix. Some of the species had the same DNA binding interface so the overlap in some of the zinc fingers indicate that the residues interacting with DNA are all the same. Zinc fingers that have no predicted binding residues are indicated by N/A and (-) indicates that there was truncation within the sequence. GREEN represents that the canonical position amino acid was not conserved across all species but has the same hydrophobicity (S, T, H, Q and E are hydrophilic). RED represents that the canonical position amino acid as not conserved across all species but have the same charge (R and H are basic + D and E are acidic). **BOLD** represents the amino acids are conserved throughout all the species and lowercase letters represents amino acids not conserved throughout all the land vertebrate species.

	Predicted DNA Binding Interface												
Species	ZF1	ZF2	ZF3	ZF4	ZF5	ZF6	ZF7	ZF8	ZF9	ZF10	ZF11	ZF12	ZF13
H. sapiens	SgFR	KISN	HgAt	EglR	HSeK	DSNT	DKEQ	NSDR	RSEK	N/A	QnEk	DSGR	RSEQ
P. troglodyte													
G. gorilla													
M. mulatta													
T. francoisi													
P. anubis													
C. sabaeus													
R. roxellana							DKEH						
H. moloch							DKEQ						
C. jacchus													
M. musculus ZFY1	TrFR		N/A	N/A	HSaK		N/A				QcEt		
M. musculus ZFY2			(-)[§§]										
R. norvegicus	TsFR		N/A	EtsH	HSeK						QcEk		
M. marmota	SgFR		HgAt	EglR			DKEQ				QnEk		
B. taurus													
B. bison													
C. hircus													
C. elaphus													
O. virginianus											QnDk		
S. scrofa							DKDQ				QnEk		
N. asiaeorientalis		N/A					DKEQ						
C. lupus		KISN											(-)[***]
M. erminea													RSEQ
L. africana													
E. caballus											QtEk		
O. anatinus											QgEk		
M. domestica											QnEk		
G. gallus			N/A	N/A									
X. laevis			HnAa	EglR			DKDE				QiEk		

[§§] *M. musculus* Zfy2 shows that there was no DNA binding interface as there was no predicted zinc finger 3 due to the truncation within that region of the sequence

[***] *C. lupus* ZFY shows an absence of a DNA binding interface as the sequence was incomplete and there was no zinc finger 13 predicted

3.6 Transformations

As indicated by **Table 8**, eight of the constructs were transformed into *E. coli* competent cells and the transformation of the plasmid vectors into the *E. coli* was successful as every plate developed colonies, expected as each of the vector had ampicillin resistance. However, upon observation the *E. coli* transformants expressing hZFY-long AD had a significantly lower number of colonies compared to *E. coli* transformants expressing hZFY-short AD. Subsequently after the transformation, we focused first on protein expression.

3.7 Growth of pET15b transformed *E. coli* cells

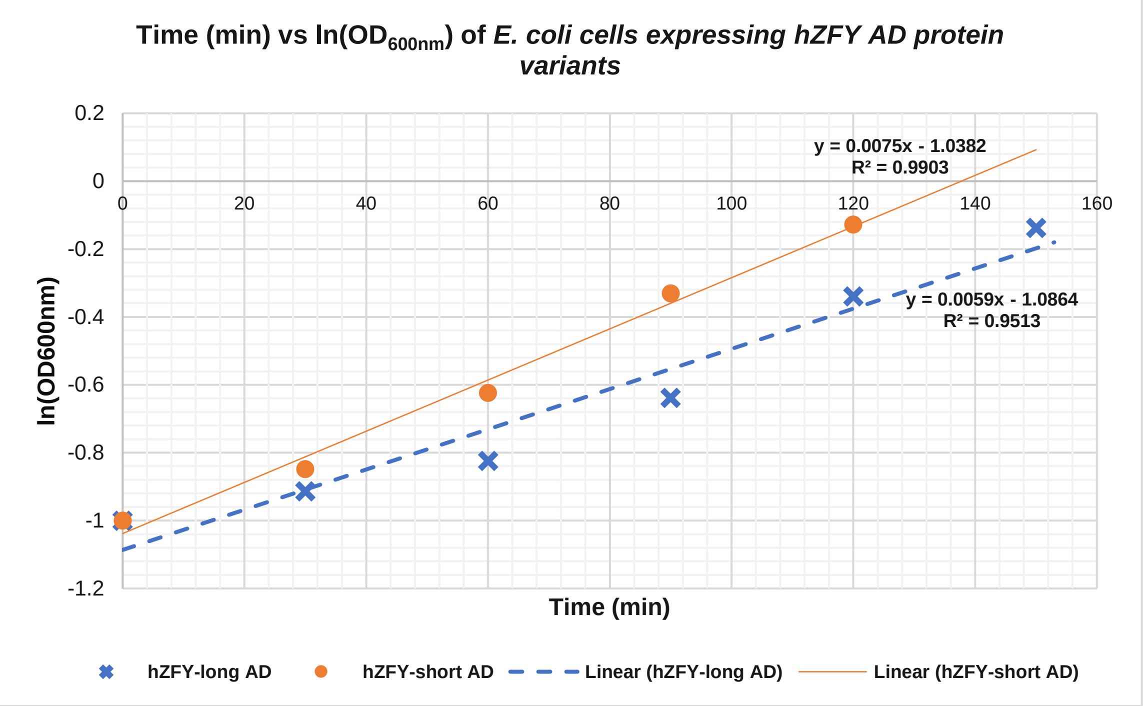


Figure 11. Natural log plot for the growth of transformed T7 E. coli expressing the recombinant hZFY AD variants (without the DBD) over a period of 120 (hZFY-short AD) and 150 minutes (hZFY-long AD). A difference was noticeable in the rate of growth as at every 30-minute interval the OD_{600} of *E. coli* expressing hZFY-short AD increased with larger increments than *E. coli* expressing hZFY-long AD. The induction began when the cultures reached $OD_{600} \geq 0.7$

To obtain the protein, we used the growth method described in **Section 2.5**. Transformed *E. coli* cells with pET15b vectors, expressing hZFY-long AD and hZFY-short AD exhibited exponential growth. Additionally, **Figure 11** indicates that

E. coli cells expressing hZFY-long AD were growing at a slower rate than the *E. coli* expressing hZFY-short AD.

.

3.8 Recombinant hZFY-long AD expression was absent, but hZFY-short AD was abundant

We tried to express both hZFY-long AD and hZFY-short AD (both without the DBD) in the bacterial expression system. We carried out SDS-PAGE to analyse the presence and expression of hZFY variants indicated by **Figure 12** after nickel column chromatography using the protocol from Thermo Scientific (available at https://www.thermofisher.com/order/catalog/product/88221#/88221). The Coomassie stained gel showed that both expression experiments shared bands roughly at 20kDa, 25kDa, 32kDa,66kDa and 90kDa. These bands were likely contaminants non-specifically bound from each hZFY isoform protein products. **Figure 12** shows that the expression of hZFY-long AD was unsuccessful as there was no expression of a band between 40-55kDa. hZFY-short AD showed two bands roughly at 35kDa and 39kDa that were specific only in the hZFY-short expression. These bands suggested hZFY-short AD protein expression had been successful as, but they were at a higher molecular weight than we expected. The expected molecular weight of the hZFY-short AD was estimated to be approximately 28.3kDa when including the hexahistidine tag and thrombin cleavage site, so to confirm that the protein bands expressed were truly hZFY-short AD, we checked by western blot and confirmed these two bands represented his-tagged products. We can therefore conclude that the expression of hZFY-long AD was unsuccessful, and hZFY-short AD was successful.

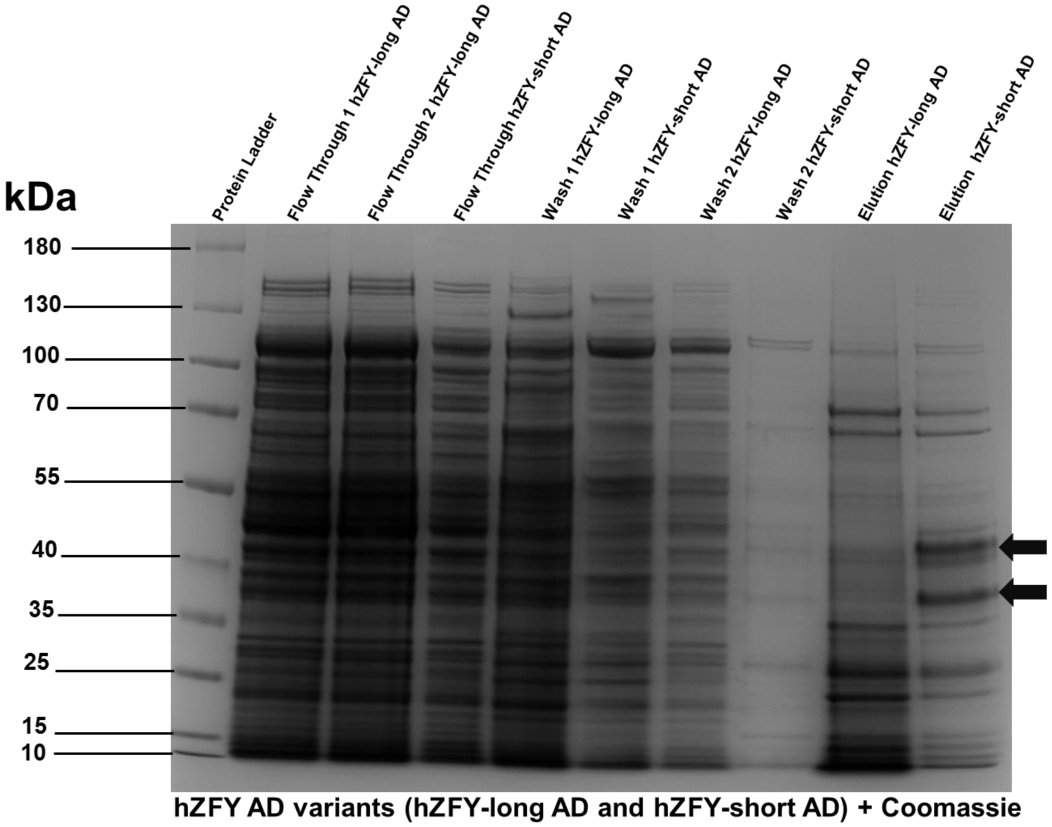


Figure 12. Expression of hZFY AD variants without (w/o) DBD. Parallel comparison of Coomassie stained gel of hZFY-long AD and hZFY-short AD expression. The expected molecular weight of hZFY-short AD was roughly 28.3kDa with the hexahistidine tag and thrombin cleavage site but the arrows indicated the molecular weight to be roughly 35kDa with an unexpected band around 39kDa as indicated by the BLACK arrows.

3.9 Western Blot analysis shows hZFY-short AD protein was present

Figure 13 shows hZFY-short protein was present and hexahistidine tagged as the capturing device detected the 2 bands, showing 2 possible molecular weights for the protein at 35kDa and 39kDa, although the size of the protein was expected to be roughly 28.3kDa. We also identified faint bands at lower molecular weights in the same elution lanes characteristic of hexahistidine tagged proteins that were likely bound non-specifically. We can therefore conclude that the expression of hZFY-short AD was successful.

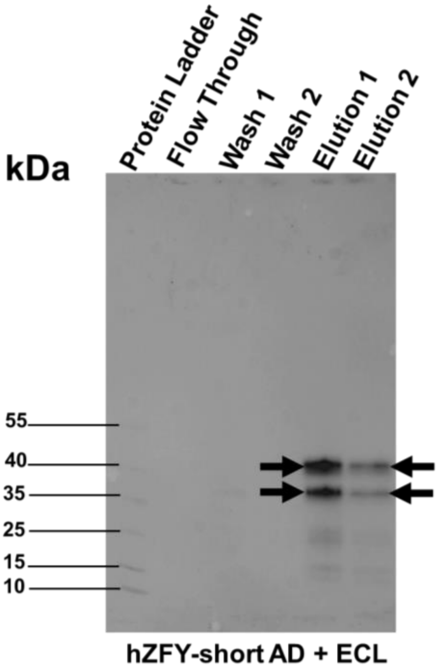


Figure 13. hZFY-short AD protein without the DBD expression shown using horse radish peroxidase tagged primary antibody. The nitrocellulose membrane showed protein expression and an anomalous band higher than hZFY-short AD expression. The protein was present due to the hexahistidine tag for isolation as nickel chromatography isolates the proteins based on the metal ion attached to the protein. Therefore, non-specific binding is removed. The capture time for the membrane image was 59 seconds.

Though the attempt of purification of hZFY-short AD was successful, expression and isolation of hZFY-long AD was unsuccessful. We attempted SDS-PAGE runs and neither runs were successful. **Figure 14A** shows a band between 35-55kDa amidst the other non-specific bands within the elution fraction which we suggested it could be our protein of interest. Therefore, for confirmation, western blot (**Figure 14B**) did not show any bands between 35-55kDa, invalidating the assumption that this was hZFY-long AD.

We lowered the imidazole concentration of the nickel column chromatography wash buffer only from 25mM to 10mM and **Figure 14C** shows the elution lane still had non-specific bands and hZFY-long AD purification was still unsuccessful due to a lack of an intense band at roughly 48.6kDa (hZFY-long AD molecular weight and the tags), but for further validation that hZFY-long AD was not present, we did a western blot (**Figure 14D**) that confirmed that the extraction was unsuccessful.

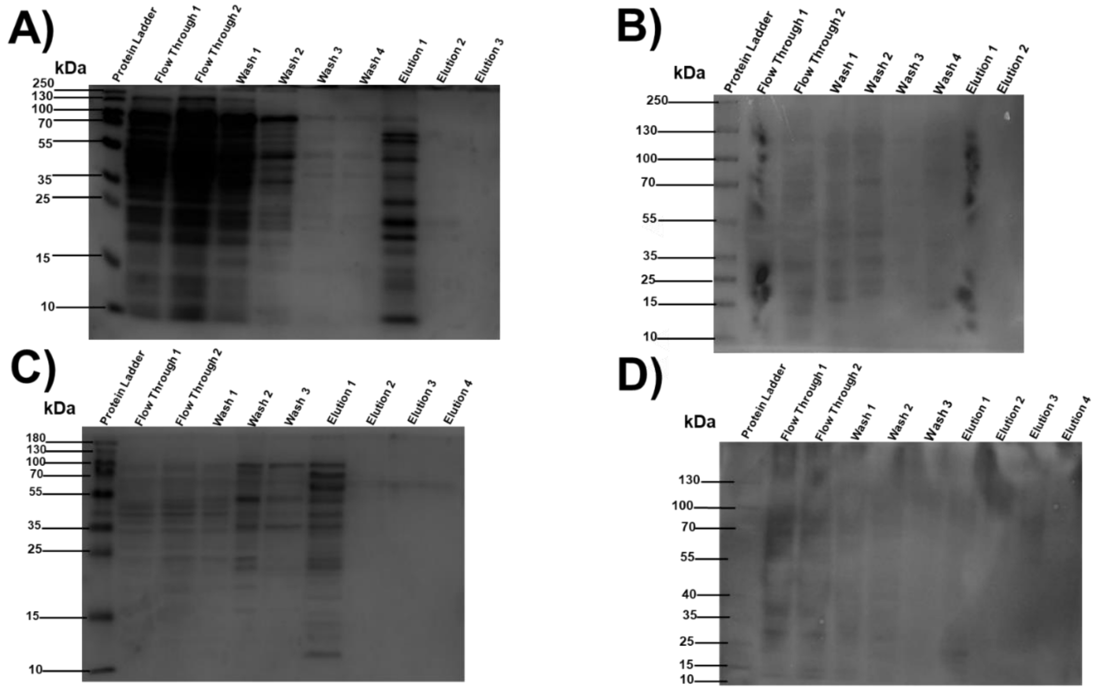


Figure 14. Expression patterns on SDS PAGE and Western blots of hZFY-long AD without DBD. hZFY was harvested as described in section 2.4. (**A**) Coomassie stained gel with wash buffer imidazole concentration of 25mM and PageRuler Plus Prestained Protein Ladder (**26619**). (**B**) Polyvinylidene fluoride (PVDF) membrane visualized in a G:Box F3 gel imaging system with a capture time of 1 minute and 37 seconds, and the same imidazole concentration as gel **A**. (**C**) Coomassie stained gel with wash buffer imidazole concentration of 10mM and PageRuler Prestained Protein Ladder (**26616**). (**D**) PVDF membrane with the same imidazole concentration as gel **C**, and a capture time of 2 minutes 29 seconds.

4 Discussion

4.1 ZFY N-terminus is less conserved than the C-terminus

Our bioinformatic analysis showed ZFY proteins had a highly conserved DNA binding domain (across the 44 species) at the C-terminal portion in comparison to the N-terminal portion, as the majority of conserved amino acids as well as the nucleotides were within the final exon of each of the species. This highlighted that the ZFY acidic activation domain likely went through a series of diverging evolutionary events in the majority of species.

The N-terminal portion of ZFY seems to have clusters of conserved amino acid patches typically ≥4 residues. This likely stabilises the interactions of the transactivation domain with transcription regulation machinery, and these specific amino acids are possibly redundant in all the species as they are well preserved across all the species (Ahmad *et al.*, 2010). Conserved residue clusters have been implied to be the ‘energy hotspots’ for protein-protein complexes, as they stabilise the protein’s core and interfaces for better interactions (Ma *et al.*, 2003; Ahmad *et al.*, 2010). Therefore, with the conservation of the N-terminal residues being relatively low in ZFY, some of these clusters likely stabilise the protein-protein complexes that arise from interaction with other transcription regulatory elements. There are several of these clusters dispersed over the span of the N-terminal activation domain likely so the ZFY interface remains flexible but maintain the critical interactions (Ma *et al.*, 2003). This pattern is typical for acidic transactivation domains, where the conserved regions represent the relatively small hydrophobic binding interface, while the less-conserved regions are highly charged “spacer” regions that act to prevent the protein from folding and remain in an unstructured configuration that presents the hydrophobic patches to the surrounding medium. As such, the precise sequence is not important for these spacer regions, only that they should be highly charged in solution.

It is important to note that some of the exon sequences used for the protein alignments were predicted. Predicted sequences are usually modelled by studying the conservation of the DNA sequences across related species and using sequence homology of a characterised sequence to identify the exon sequence. As these predicted ZFY genes display high sequence similarity to the characterised ZFY genes, the functions are likely to be similar. However, as the proteins have not been isolated or characterised in the species that have the exon sequences as predicted, any functionality inferred is to be treated cautiously.

Although fish species were included in the initial sequence alignment, there was a possibility of misidentification of ZFY as most sequences were ZFY-like isoforms and were likely generated from ZFX as they are homologous to ZFY. Therefore, it is a possibility that the reason why there is a divergence as we proposed is due to ZFY sequences of the land vertebrates being compared to the misidentified fish ZFY.

4.2 Nine amino acid TAD motif predictions are highly conserved, and two additional motifs located on the 5th and 6th exon

Although four possible (perfect) 9aa TAD regions were predicted all within the N-terminal region of the protein, only one of these regions was highly conserved across all land vertebrates from residues 63-76. It is therefore likely that this motif is of key functional significance in ZFY proteins. In addition, the motif between 338-346 (**Table 14**) was also predicted throughout all land vertebrates with the exception of three of the four rodent sequences, and it showed a percentage identity of 78%. It is a possibility that rodents had a high degree of divergent evolution in comparison to the other land vertebrates which retained this sequence.

Our data suggests residues between the sites 62-75 (**Table 13**) or 63-76 (**Table 14**) are responsible for the majority of transactivation activity of ZFY. As the 9aa TAD motif (**Table 14**) between sites 40-48 was not predicted in most mammals including *H. sapiens* and *M. musculus* Zfy2, this was suggestive that this motif was not essential for transactivation as *M. musculus* Zfy2 and *H. sapiens* full-length ZFY had putative transactivating acidic domains without this motif (Decarpentrie *et al.*, 2012). In addition, the 9aa TAD motifs between sites 167-178 (**Table 14**) and 338-346 (**Table 14** DEVYMEVIV) were predicted in *H. sapiens* ZFY but not predicted in *M. musculus* Zfy2, and the transactivating activity of *M. musculus* Zfy2 was higher when fused to Gal4-DBD.

Much remains to be understood about the biology of 9aa TAD motifs, and it is possible that they are repressor as well as activator motifs. Notably, hZFY-short also possesses the DEVYMEVIV motif between sites 316-324 (**Table 13**) motif but does not have transactivating properties when fused to Gal4-DBD (Decarpentrie *et al.*, 2012). It is thus plausible that the DEVYMEVIV motif – present in both short and long forms – acts to repress transcription while the longer SVVIQDVVEDVVIE motif present in the alternatively spliced second coding exon acts to promote transcription. The increased transactivation ability of mouse Zfy2 could therefore be due to the selective loss of the (potentially inhibitory) DEVYMEVIV motif.

Overall, hydrophobic clusters separated by hydrophilic regions within the 9 amino acids TAD motifs were suggestive that these are essential for transcriptional activation. A study by (Almlöf, Gustafsson and Wright, 1997) established a relationship that hydrophobic patches increased transactivation activity of a nuclear transcription factor, and the mechanism that was demonstrated to increase the transactivation activity was because of better interactions with coactivators. Furthermore, studies of transactivation domains of Gcn4 and p53 (Jackson *et al.*, 1996; Krois *et al.*, 2016) have also demonstrated that hydrophobic clusters within the transactivation domain were likely important for mediation of hydrophobic interactions with transcriptional machinery. Therefore, future studies addressing if there are any effects from mutagenesis of hydrophobic residue to hydrophilic residues (specifically on SVVIQDVVEDVVIE) on transactivation activity will be useful to confirm the theories. It is uncharacteristic however for 9aa TAD transcription factors to have unusually high valine content as they inactivate 9aa TAD motifs (Piskacek *et al.*, 2019), but ZFY possesses valine rich 9aa TAD motifs which could indicate a new class of 9aa TAD motifs.

Interestingly, we located possibly another conserved motif adjacent the DEVYMEVIV motif referred to as the polyalanine motif located within the acidic domain. (Poloumienko, 2004). As this motif was conserved in placental mammals other than rodents, this suggested that the motif was likely specific to placental mammals. Polyalanine tracts have been associated with transcription regulation molecular binding and transcription regulator, as motifs with 5-7 alanine residues or >7 alanine residues in human genes were mostly involved in these molecular functions, thus these motifs likely play a significant role in the various placental ZFY proteins (Lavoie *et al.*, 2003). However, further mutagenesis studies can be conducted to analyse if there are detrimental effects to ZFY function.

In addition, another motif was located in all of the land vertebrates and fish species within the acidic domain, more specifically the 6th exon and this we concluded to be an NLS due to the basic nature of the region as it contains the K(K/R)X(K/R) consensus sequence that binds an importin (α-importin) for nuclear localization (Mardon and Page, 1989; Lange *et al.*, 2007). The proposed nuclear localisation motif however was not highly conserved as the basic amino acids can be interchangeable. Nonetheless, this region remained basic across all the land vertebrates and fish species.

4.3 ZFY zinc fingers are highly conserved

We found the general consensus was ZFY contained 12 zinc fingers. Zinc finger 10 was not confidently predicted as a zinc finger and, it did not have predicted binding targets or DNA binding interface, contradictory of the previous (Page *et al.*, 1987; Mardon and Page, 1989) findings in which the protein was predicted to have 13 zinc fingers in humans and mouse Zfy2. In addition, zinc fingers 3,4 and 7 within most of the rodent lineage seemed to be not well retained (low ZF score) which could indicate that they were likely functionally insignificant resulting in their loss over time. The majority of the zinc fingers were conserved in most mammals which suggested that there is some constraint which enforced the DNA binding domain in exon 7 to be very well conserved.

Fish species however have two inserts in the C-terminus in the region where the linker region between ZF3 and ZF4, and also one imbedded where ZF11 was predicted for land vertebrates. As the two inserts were rich in proline, this likely has an effect on the zinc finger motif arrays as proline residues kink amino acid backbones, thus inhibiting α helix formation needed along with two β-sheets (ββα-structure) to form a zinc finger. This suggested possibly that fish species have a modular DNA binding domain that has different functions at different regions, thus not behaving as one unit but multiple units that work congruently to execute a task.

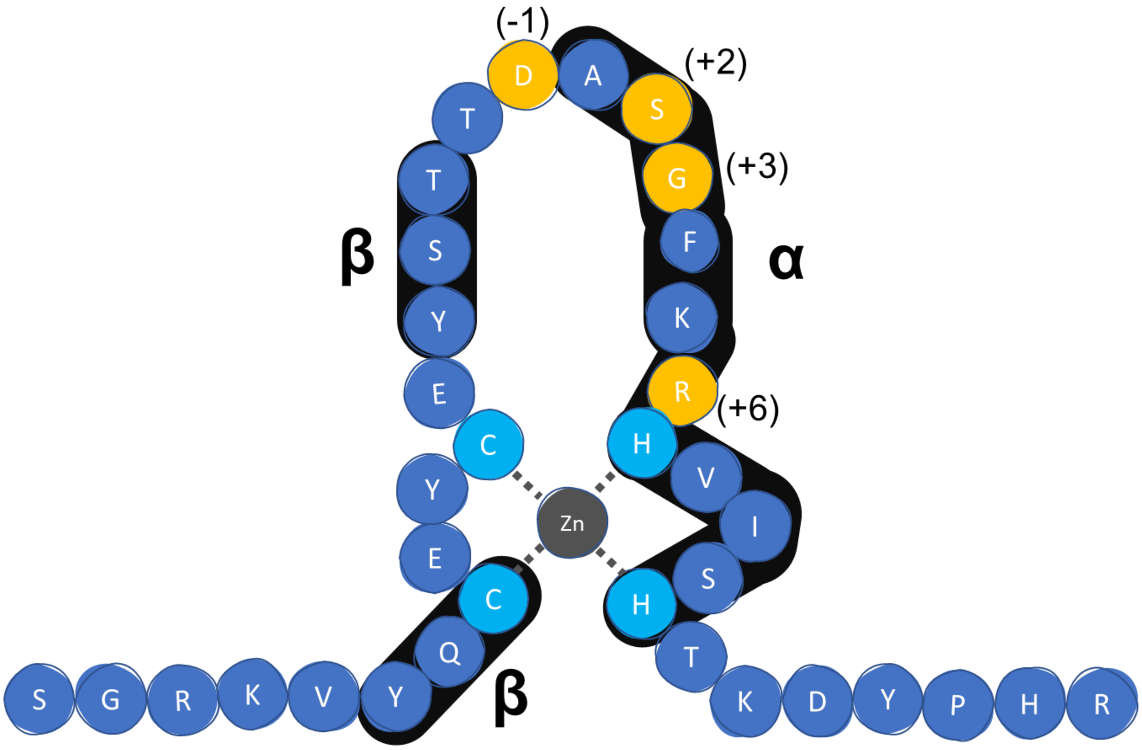


Figure 15. Zinc finger structure with a C2H2 motif. Diagram represents zinc finger 12 of the DNA binding domain of hZFY. **CYAN** represents paired cysteine (C) and histidine (H) residues that interact with the **GREY** zinc ion (Zn) which stabilises the structure of the zinc finger via a tetrahedral coordination complex. The finger contains two β-sheets and an α-helix (indicated by **BLACK**) which holds the DNA binding interface containing 4 canonical positions (-1,2,3, and 6) that are interpreted to interact with DNA indicated by **AMBER**. This image is an adaptation of (Stubbs, Sun and Caetano-Anolles, 2011)

4.4 ZFY zinc finger binding sites

It was important to investigate the conservation of the zinc fingers binding sites. The majority of the ZFY binding interfaces (**Table 14**) had high conservation between the different species which was suggestive of these amino acids having the general function of enabling the DNA binding specificities and result in high affinity binding, though great care should be taken as affinity and specificity are not coupled. Though the same amino acids were not conserved in some of the DNA binding interfaces, the properties were conserved. For example, some of the fingers would retain hydrophobicity or charge in some species as shown by **Table 14** implying that the binding of the zinc fingers is likely dependent on the properties of the amino acids rather than the shape for lock and key type interactions. As three of the four rodent sequences had the most variation in most of the DNA binding interfaces, this implied that there was a high selective pressure not to maintain the regulatory target elements of these specific zinc fingers (Sommer *et al.*, 1992).

As DNA binding of ZFY is not known exclusively, some theories suggest that all zinc fingers are involved in binding to DNA, and some explain not all of the zinc finger domains of ZFY are involved in binding DNA. But the general suggestions have implied that all zinc fingers bind to DNA but in varying ways. For instance, one mechanism for transcription factor binding has been interpreted in which some of the transcription factor zinc fingers bind DNA at nonspecific sites while the other fingers scan and direct the transcription factor to the target sites. Therefore, some of the ZFY zinc fingers could be involved in binding at nonspecific sites when the protein is scanning DNA targets sites and when the zinc fingers locate the target site, the rest of the ZFY zinc fingers bind their specific DNA target for a 'recognition' mode until dissociation of the zinc finger for the repeat of the process (Zandarashvili *et al.*, 2015; Zuo *et al.*, 2019). This implies that some zinc fingers (primary binding sites) within ZFY modulate DNA binding as they increase the affinity and stability of secondary binding sites. As ZFY zinc fingers were proposed to bind typically to 3 bases of DNA, the interactions of a single zinc finger are not likely strong as the zinc finger DNA binding interface is large so secondary binding sites stabilise existing primary binding site interactions. In addition, greater affinity allows a sufficient window for ZFY interaction with regulatory elements, resulting in efficient repression or activation (Zuo *et al.*, 2019).

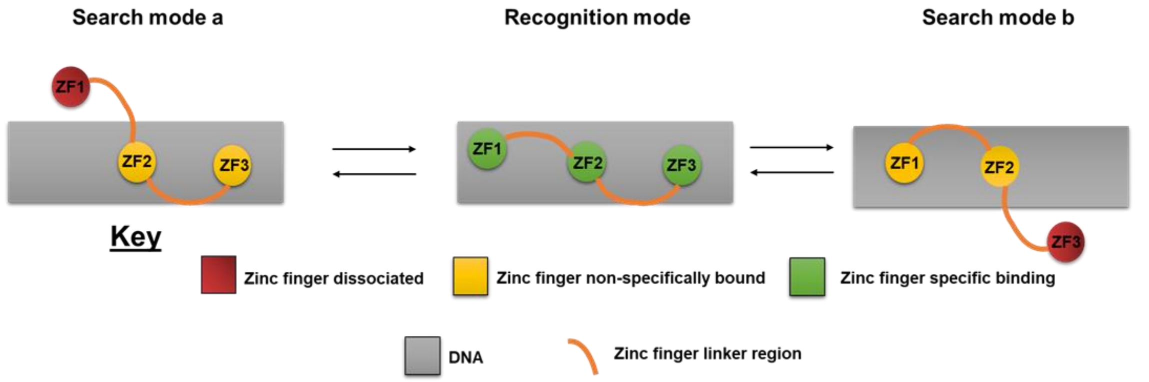


Figure 16. Proposed ZFY binding mechanism. The mechanism illustrates two search modes by which one finger is always dissociated whilst two zinc fingers are bound non-specifically. The recognition mode shows zinc fingers with a high affinity for DNA. This illustration accounts for a portion of the ZFY zinc fingers as it is an adaptation from (Zandarashvili et al., 2015).

4.5 ZFY phylogeny

The two protein phylogenetic trees showed as expected that ZFY proteins of the different primate species were consistently grouped. Unexpectedly, the majority of rodents had unique evolution pathways as they separated from *M. marmota*. This similar pattern was also observed less severely in artiodactyls as they exhibited a similar dispersal across the evolutionary tree (*N. asiaeorientalis* and *S. scrofa* dispersed). Additionally, as expected the fish species were all distantly related to the other species. The nucleotide phylogenetic tree showed however that the rodents and artiodactyls do not exhibit the unique evolution pathways as they are all grouped respectively, but the rodent lineage still displayed a long branch indicating higher nucleotide substitutions. These results indicated that there was a higher proportion of nonsynonymous substitutions, such that the correct phylogeny was recovered from the nucleotide sequence alignment and not from the protein sequence alignment. Usually, synonymous substitutions are higher in proportion to nonsynonymous substitutions. Thus, excess nonsynonymous substitutions are suggestive of positive selection in the common ancestor of rat and mouse and are evidence for the rapid evolution of the rodent protein sequence (McDonald and Kreitman, 1991). Due to time constraints limiting our experiments, future experiments should incorporate the *codeml* Phylogenetic Analysis by Maximum Likelihood (PAML) software package to quantify the ratio of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site and determine if there was selective pressure on the rodent lineage.

4.6 *E. coli* expression of recombinant hZFY proteins

As much of the research conducted for ZFY was mostly pre 2000s, there has been a gradual decline in interest since it was determined ZFY was not the testis-determining factor. Bacterial expression of hZFY has been scarce as many papers usually analyse *ZFY* expression in mouse germ cells, abnormal expression of *ZFY* in HNSCC cell lines, and hZFY transactivation activity has been conducted in *S. cerevisiae*. Therefore, there was a need for hZFY to be expressed and purified for structural and biochemical assays.

Our data suggested that the length of the acidic domain determined how bacterial colonies and bacterial cultures grow. As the **Figure 11** demonstrated, *E. coli* expressing hZFY-short AD grew approximately 1.3 times quicker than *E. coli* expressing hZFY-long AD prior IPTG induction. A similar outcome was observed by (Decarpentrie *et al.*, 2012) and colleagues, in *S. cerevisiae*, where they observed smaller colonies of yeast transformants of (highly active) mouse Zfy2-long, extended growth time in liquid media, and lower expression of Zfy2-long AD. They concluded that there possibly was a selection against high expression clones.

This similar hypothesis could be applied to our experiments as there were lower observed *E. coli* transformant colonies expressing the hZFY-long AD and higher overall growth time in the liquid culture (**Figure 11**). It was a possibility that the hZFY-long AD had a toxic effect on the bacterial cells, and this perturbed the proliferative capabilities of the cells, likely leading to apoptosis or defects in the growth mechanism of the *E. coli* (Dumon-Seignovert, Cariot and Vuillard, 2004; Saida *et al.*, 2006). As our plasmid contained a T7-based promoter and contained a lac operon, these allow leaky gene expression which is a process by which bacterial systems have basal expression of a protein (Nielsen, Willis and Lin, 2007). This is because the transcriptional control is not 100% efficient as the promoter does not turn off completely even when an inducer is not present, due to the lac repressor protein binding DNA operator sites with an efficiency lower than 100% (Nielsen, Willis and Lin, 2007).

Leaky expression of genes increases in LB-media growths as bacterial cultures enter the stationery phase likely due to scarcity of nutrients acting as a limitation (Nielsen, Willis and Lin, 2007). Therefore, we assumed that the promoter and element allowed hZFY-long AD leaky expression before induction by IPTG, and as we suggested that the gene was toxic to our *E. coli* strain, the leaky expression of the toxic gene in bacteria before induction likely eliminated some of the bacteria and induction increased the rate of elimination. Leaky expression has been associated with plasmid instability, so we also assumed that this also likely led to a poor yield in hZFY-long AD protein (Rosano and Ceccarelli, 2014).

To formally test for leaky expression of hZFY-long AD, future experiments should include small-scale expression tests (**Figure 17**). This involves growing our pET15b transformants in small quantities (i.e., 6mL) and collecting uninduced aliquots (roughly 400µl) at consistent intervals until OD_{600} of 0.7. Then, we induce the samples with varied IPTG concentration (i.e., 0.1-1.0mM), also collecting induced aliquots (roughly 400µl) at consistent intervals. For every aliquot of induced and uninduced, we prepare samples for SDS PAGE by microcentrifugation and denature samples by boiling. The samples will then be run on SDS PAGE gels (for western blot also) and analysed to identify if cells exhibit basal or ‘leaky’ expression of the hZFY-long AD. In addition, toxicity of the hZFY-long AD can be tested directly by IPTG induction of the *E. coli* cells and performing colony counts on agar plates to observe the number of cells that remain viable after a certain period.

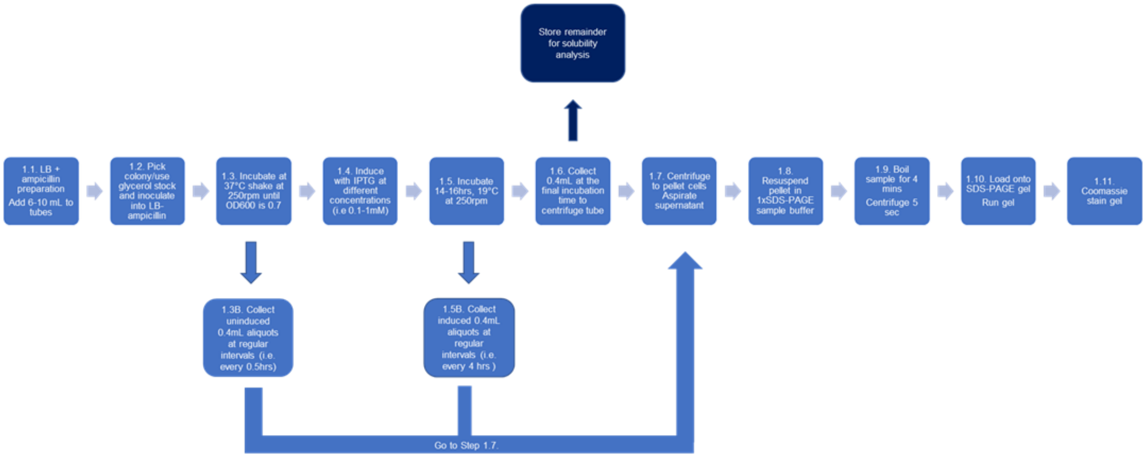


Figure 17. Flow chart for the small-scale expression test adaptation (Zerbs, Giuliani and Collart, 2014). If hZFY-long acidic domain is expressed prior induction (step 1.3B) and after IPTG induction there is lack of expression, this shows that the leaky expression has an impeding effect on *E.coli* expression. Step 1.3B and 1.5B are useful for collecting cells at varied times of growth to see progression of expression.

For the other constructs such as pFN26A or pcDNA3.1(+) (**Table 6**), the growth rates of the transformed bacteria would be expected to have no difference in growth as these contain eukaryotic promoters so will only be active in mammalian cells.

4.7 Electrophoresis and western blot analysis

Both SDS and western blot data for hZFY-short AD protein without the DBD showed evidence that the expression was successful in the elution fractions of the purified *E.coli* supernatant. High intensity bands indicated by **Figure 12** and

Figure 13 suggested that the his-tagged protein was present, with the bands seen at an estimated size of 35kDa and 39kDa. This differs from the https://web.expasy.org/protparam/ prediction (Gasteiger *et al.*, 2005), as the predicted size was estimated to be 28.3kDa including the vector-derived his-tag and thrombin cleavage site.

An explanation for an anomalously high observed molecular weight could be the high proportion of negatively charged amino acids in this protein domain, as 24% of the residues were aspartic and glutamic acid (51 of 216 amino acids). Negatively charged amino acids impede SDS binding to the protein and thus alter their migration in an SDS-PAGE gel (Graceffa, Jancsó and Mabuchi, 1992), and mutation of these residues is shown to restore migration to the normal expected.

An equation was therefore derived to account for negatively charged residues by (Guan *et al.*, 2015) and colleagues which is $y = 276.5x - 31.33$ (where x is the percentage of negatively charged amino acids and y is the average MW per amino acid). Using this equation, our ZFY-S construct is predicted to run at 35.03kDa, very similar to the lower band observed experimentally. However, a useful technique in verifying the identity of the protein band is mass spectrometry as it digests the protein and produces fragmented peptide ion peaks (Wang and Wilson, 2013).

As yet the secondary upper band at 39kDa remains unexplained. This could be potentially due to incomplete reduction of disulphide bonds in the protein, due to the reducing agent used. As β-mercaptoethanol is a volatile reducing agent, the concentration in solution usually decreases with time which likely resulted in poor ZFY reduction or reoxidation and thereby result in an extra and heavier band (https://www.bio-rad.com/webroot/web/html/lsr/tech_support_faqs/FAQ268440261.html). As dithiothreitol (DTT) is less volatile, it is probably more likely to reduce ZFY properly and provide a better result with just one distinct band. A useful experiment to distinguish which of the 2 bands is hZFY-short AD is via mass spectrometry of each band.

An explanation on two bands being present could be representative of folded and aggregated protein. As the lower band is closer to the expected molecular mass, we can assume that the lower band was the fully denatured protein, and the upper band could be an aggregated protein. This is likely because of the conditions at which the ZFY expression was performed as (Wang *et al.*, 2011) demonstrated that high expression situations lead to more aggregation of protein in inclusion bodies and produces lower bands that are less aggregated. Therefore, the 39kDa band could also be a result of protein aggregation.

4.8 Future Work

Due to the COVID-19 pandemic, we only touched the surface on ZFY's expression and isolation. To overcome the problems we encountered, future bacterial growth and expression experiments need to be optimized (as previously stated) to allow us to harvest pET15b constructs expressing the hZFY-long AD. Future experiments could include circular dichroism, which is a useful technique for secondary structure determination of our recombinant protein to also reveal the folding property of the protein (Greenfield, 2007). Structure and folding of proteins reveal the protein function as change in conformation of the protein likely affects the function, and in addition interacting regions/ partners are likely to be determined by assays such as pull-down assays. The structure of the acidic domain of hZFY in particular will likely reveal the conformation responsible for the transactivating properties of the hZFY-long AD.

The direction of the research should also include analysing the function of the two hZFY protein variants. These experiments include mammalian cell line expression after transfections from pCDNA3.1(+) constructs. As a pair of the constructs have eGFP tags, signals should be detected in the nuclei of mammalian cells if the protein is expressed, and it allows us to perform reporter assays using the hZFY protein variants within the mammalian cell lines. The other pair of constructs have a HA tag which is useful for western blotting, immunoprecipitation, and immunofluorescence experiments, which will allow us to isolate and purify hZFY variant proteins after detection. It is worth noting that as hZFY-long function is thought to be proapoptotic as in spermatogenesis, thus hZFY-short is likely anti-apoptotic when transfected in mammalian cells as it has no transactivation abilities and likely antagonistic.

The other pair of constructs had a pFN26A vector backbone containing a luciferase reporter, that is useful for luciferase reporter assays. As we determined that the potent 9aa TAD motif of hZFY was SVVIQDVVEDVVIE, an assumption is that when hZFY-long AD is fused to Gal4-DBD, the cells should exhibit higher luminescence than in hZFY-short AD fused to Gal4-DBD. This is because hZFY-long transactivates the Gal4-DBD, so the activity is likely significant in comparison to a hZFY-short and Gal4-DBD fusion. Therefore, mutagenesis of specific amino acids within this region can be analysed to show any change in transactivation, should the 9aa TAD motif be experimentally confirmed to be the motif responsible for transactivation.

As the *in-silico* results suggested that ZFY sequences were largely conserved across various species, further research would be useful to analyse whether the sequence conservation is related to the molecular functionality of the gene. For instance, a useful experiment would be to analyse the highly conserved zinc finger domains and create zinc finger constructs of various species to closely monitor which of the residues are involved in DNA binding. This also includes experiments to show how changes of these residues affect the binding mechanism of the zinc fingers with the target nucleotides. In addition, this also includes further investigation of the 9aa TAD and analysing the one with the highest percentage identity as usually sequences that have a high homology usually suggest similar function.

4.9 Conclusions

Whilst the project failed to express the hZFY-long AD from *E. coli* cells, this was likely because of the limitations presented by the lac operator as it is known to cause protein expression before induction by IPTG, therefore impeding the growth of the cells after induction and the outcome is no protein expression and 1.3-fold difference in the *E. coli* growth. Nonetheless, the presence of the lac operator in *E. coli* cells had no negative impact on hZFY-short AD expression as the protein was present, suggesting that the longer variant likely had a toxic phenotype on our bacterial cells. Intriguingly, hZFY-short AD was shown via SDS-PAGE and western blot analysis to have a higher observed molecular weight than the predicted, and an anomalous band was also located above the hZFY-short AD.

The acidic domain appeared to be the least conserved region of ZFY across several species. Nevertheless, the acidic domain has been shown to possess several motifs which include 9aa TAD motifs, polyalanine motifs and a nuclear localisation motif, all with varying conservation throughout several species. Most importantly, residues 63-76 are predicted to be the 9aa TAD motif responsible for transactivation in hZFY-long. The DNA binding domain of ZFY appears to be highly conserved in comparison to the acidic domain and has 12 zinc fingers that vary in conservation across several species, and these bind 3 nucleotides at specific DNA sites using specific residues at the DNA binding interface.

5 References

Ahmad, S. *et al.* (2010) ‘CCRXP : exploring clusters of conserved residues in protein structures’, *Nucleic Acids Research*, 38(May), pp. 398–401. doi: 10.1093/nar/gkq360.

Almlöf, T., Gustafsson, J. A. and Wright, A. P. (1997) ‘Role of hydrophobic amino acid clusters in the transactivation activity of the human glucocorticoid receptor.’, *Molecular and Cellular Biology*, 17(2), pp. 934–945. doi: 10.1128/mcb.17.2.934.

Bellani, M. A. *et al.* (2005) ‘SPO11 is required for sex-body formation, and Spo11 heterozygosity rescues the prophase arrest of Atm-/- spermatocytes’, *Journal of Cell Science*, 118(15), pp. 3233–3245. doi: 10.1242/jcs.02466.

Božinović, K. *et al.* (2019) ‘Genome-wide miRNA profiling reinforces the importance of miR-9 in human papillomavirus associated oral and oropharyngeal head and neck cancer’, *Scientific Reports*, 9(1), pp. 1–13. doi: 10.1038/s41598-019-38797-z.

Bray, F. *et al.* (2018) ‘Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries’, *CA: A Cancer Journal for Clinicians*, 68(6), pp. 394–424. doi: 10.3322/caac.21492.

Burgoyne, P. S., Mahadevaiah, S. K. and Turner, J. M. A. (2009) ‘The consequences of asynapsis for mammalian meiosis’, *Nature Review Genetics*, 10(3), pp. 207–216. doi: 10.1038/nrg2505.

Chua, H. H. *et al.* (2015) ‘RBMY, a novel inhibitor of glycogen synthase kinase 3β, increases tumor stemness and predicts poor prognosis of hepatocellular carcinoma’, *Hepatology*, 62(5), pp. 1480–1496. doi: 10.1002/hep.27996.

Decarpentrie, F. *et al.* (2012) ‘Human and mouse ZFY genes produce a conserved testis-specific transcript encoding a zinc finger protein with a short acidic domain and modified transactivation potential’, *Human Molecular Genetics*, 21(12), pp. 2631–2645. doi: 10.1093/hmg/dds088.

Domingo-vidal, M. *et al.* (2019) ‘Cigarette Smoke Induces Metabolic Reprogramming of the Tumor Stroma in Head and Neck Squamous Cell Carcinoma’, *American Association for Cancer Research*, 17(9), pp. 1893–1910. doi: 10.1158/1541-7786.MCR-18-1191.

Dorak, M. T. and Karpuzoglu, E. (2012) ‘Gender differences in cancer susceptibility : an inadequately addressed issue’, *Frontiers in Genetics*, 3(November), pp. 1–11. doi: 10.3389/fgene.2012.00268.

Dreumont, N. *et al.* (2009) ‘Human RBMY regulates germline-specific splicing events by modulating the function of the serine / arginine-rich proteins 9G8 and Tra2- β’, *Journal of Cell Science*, 123, pp. 40–50. doi: 10.1242/jcs.055889.

Dumon-Seignovert, L., Cariot, G. and Vuillard, L. (2004) ‘The toxicity of recombinant proteins in Escherichia coli: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3)’, *Protein Expression and Purification*, 37(1), pp. 203–206. doi: 10.1016/j.pep.2004.04.025.

Edgren, G., Liang, L. and Chang, E. T. (2012) ‘Enigmatic sex disparities in cancer incidence’, *European Journal of Epidemiology*, 27, pp. 187–196. doi: 10.1007/s10654-011-9647-5.

Elliott, D. J. *et al.* (1997) ‘Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm’, *Proceedings of the National Academy of Sciences of the United States of America*, 94(8), pp. 3848–3853. doi: 10.1073/pnas.94.8.3848.

Elliott, D. J. *et al.* (2000) ‘A mammalian germ cell-specific RNA-binding protein interacts with ubiquitously expressed proteins involved in splice site selection’, *Proceedings of the National Academy of Sciences of the United States of America*, 97(11), pp. 5717–5722. doi: 10.1073/pnas.97.11.5717.

Emerson, R. O. and Thomas, J. H. (2009) ‘Adaptive evolution in zinc finger transcription factors’, *PLoS Genetics*, 5(1). doi: 10.1371/journal.pgen.1000325.

Finn, R. D. *et al.* (2009) ‘The Pfam protein families database’, *Nucleic Acids Research*, 38(SUPPL.1), pp. 211–222. doi: 10.1093/nar/gkp985.

Gasteiger, E. *et al.* (2005) ‘The Proteomics Protocols Handbook’, *The Proteomics Protocols Handbook*, pp. 571–608. doi: 10.1385/1592598900.

Guan, Y. *et al.* (2015) ‘An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide’, *Scientific Reports*. Nature Publishing Group, 5(1), pp. 1–11. doi: 10.1038/srep13370.

Guindon, G. and Boisclair, D. (2003) *Past, Current and Future Trends in Tobacco Use*, *UCSF: Center for Tobacco Control Research and Education*. Available at: https://escholarship.org/uc/item/4q57d5vp.

Graceffa, P., Jancsó, A. and Mabuchi, K. (1992) ‘Modification of acidic residues normalizes sodium dodecyl sulfate-polyacrylamide gel electrophoresis of caldesmon and other proteins that migrate anomalously’, *Archives of Biochemistry and Biophysics*, 297(1), pp. 46–51. doi: 10.1016/0003-9861(92)90639-E.

Greenfield, N. J. (2007) ‘Using circular dichroism spectra to estimate protein secondary structure’, *Nature Protocols*, 1(6), pp. 2876–2890. doi: 10.1038/nprot.2006.202.

Hashibe, M. *et al.* (2009) ‘Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the INHANCE consortium’, *Cancer Epidemiol Biomarkers Prev*, 18(2), pp. 541–550. doi: 10.1158/1055-9965.EPI-08-0347.Interaction.

Jackson, B. M. *et al.* (1996) ‘Identification of seven hydrophobic clusters in GCN4 making redundant contributions to transcriptional activation.’, *Molecular and Cellular Biology*, 16(10), pp. 5557–5571. doi: 10.1128/mcb.16.10.5557.

Jan, S. Z. *et al.* (2018) ‘Distinct prophase arrest mechanisms in human male meiosis’, *Development (Cambridge)*, 145(16). doi: 10.1242/dev.160614.

Jane, J. E. *et al.* (2018) ‘Horse Y chromosome assembly displays unique evolutionary features and putative stallion fertility genes’, *Nat Commun*, p. 15. doi: 10.1038/s41467-018-05290-6.

Jethwa, A. R., Khariwala, S. S. and Surgery, N. (2018) ‘Tobacco-related carcinogenesis in head and neck cancer’, *Cancer and Metastasis Reviews*, 36(3), pp. 612–626. doi: 10.1007/s10555-017-9689-6.

Jiang, H. *et al.* (2012) ‘Knockdown of zinc finger protein X-linked inhibits prostate cancer cell proliferation and induces apoptosis by activating caspase-3 and caspase-9’, *Cancer Gene Therapy*, 19(10), pp. 684–689. doi: 10.1038/cgt.2012.53.

Kido, T., Fai, Y. and Lau, C. (2015) ‘Roles of the Y chromosome genes in human cancers’, *Asian Journal of Andrology*, 17(3), pp. 373–380. doi: 10.4103/1008-682X.150842.

Klein, S. L. (2012) ‘Immune Cells Have Sex and So Should Journal Articles’, *Endocrinology*, 153(June), pp. 2544–2550. doi: 10.1210/en.2011-2120.

Koopman, P., Ashworth, A. and Lovell-Badge, R. (1991) ‘The ZFY gene family in humans and mice’, *Trends in Genetics*, 7(4), pp. 132–136. doi: 10.1016/0168-9525(91)90458-3.

Kot, M. C. and Handel, M. A. (1990) ‘Spermatogenesis in XO , Sxr Mice : Role of the Y Chromosome’, *Journal of Experimental Biology*, 256(1), pp. 92–105. doi: 10.1002/jez.1402560112.

Krois, A. S. *et al.* (2016) ‘Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein’, *Proceedings of the National Academy of Sciences of the United States of America*, 113(13), pp. E1853–E1862. doi: 10.1073/pnas.1602487113.

Lavoie, H. *et al.* (2003) ‘Polymorphism , shared functions and convergent evolution of genes with sequences coding for polyalanine domains’, *Human Molecular Genetics*, 12(22), pp. 2967–2979. doi: 10.1093/hmg/ddg329.

Lichtenstein, P. *et al.* (2000) ‘Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland’, *The New England Journal of Medicine*, 343(2), pp. 78–85. doi: 10.1056/NEJM200007133430201.

Liu, Y. *et al.* (2009) ‘The Germ Cell Nuclear Proteins hnRNP G-T and RBMY Activate a Testis-Specific Exon’, *PLoS Genetics*, 5(11). doi: 10.1371/journal.pgen.1000707.

Ma, B. *et al.* (2003) ‘Protein – protein interactions : Structurally conserved residues distinguish between binding sites and exposed protein surfaces’, *Proceedings of the National Academy of Sciences of the United States of America*, 100(10), pp. 5772–5777. doi: 10.1073/pnas.1030237100.

Mahadevaiah, S. K. *et al.* (1998) ‘Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids , and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities’, *Oxford University Press*, 7(4), pp. 715–727. doi: 10.1093/hmg/7.4.715

Mahadevaiah, S. K. *et al.* (2001) ‘Recombinational DNA double-strand breaks in mice precede synapsis’, *Nature Genetics*, 27(3), pp. 271–276. doi: 10.1038/85830.

Mahadevaiah, S. K. *et al.* (2012) ‘Spermatid development in XO male mice with varying Y chromosome short-arm gene content: evidence for a Y gene controlling the initiation of sperm morphogenesis ``’, *Reproduction (Cambridge, England)*, 144(4), pp. 433–445. doi: 10.1530/REP-12-0158.

Mardon, G. and Page, D. C. (1989) ‘The Sex-Determining Region of the Mouse Y Chromosome Encodes a Protein with a Highly Acidic Domain and 13 Zinc Fingers’, *Cell*, 56(5), pp. 765–770. doi: 10.1016/0092-8674(89)90680-6.

Mardon, G. *et al.* (1990) ‘Mouse Zfx protein is similar to Zfy-2: each contains an acidic activating domain and 13 zinc fingers.’, *Molecular and Cellular Biology*, 10(2), pp. 681–688. doi: 10.1128/mcb.10.2.681.

McDonald, J. H. and Kreitman, M. (1991) ‘Adaptive protein evolution at the Adh locus in Drosophila’, *Nature*, 351(6328), pp. 652–654. doi: 10.1038/351652a0.

Mckee, B. D. and Handel, M. A. (1993) ‘Sex chromosomes , recombination , and chromatin conformation’, *Chromosoma*, 102(2), pp. 71–80. doi: 10.1007/BF00356023.

Murr, R. *et al.* (2007) ‘Orchestration of chromatin-based processes: Mind the TRRAP’, *Oncogene*, 26(37), pp. 5358–5372. doi: 10.1038/sj.onc.1210605.

Nagamine, C. M. *et al.* (1990) ‘The two candidate testis-determining Y genes (Zfy-1 and Zfy-2) are differentially expressed in fetal and adult mouse tissues’, *Genes and Development*, 4(1), pp. 63–74. doi: 10.1101/gad.4.1.63.

Nagamine, C. M. *et al.* (1989) ‘Chromosome mapping and expression of a putative testis-determining gene in mouse’, *Science*, 243(4887), pp. 80–83. doi: 10.1126/science.2563174.

Nakasuji, T. *et al.* (2017) ‘Complementary Critical Functions of Zfy1 and Zfy2 in Mouse Spermatogenesis and Reproduction’, *PLoS Genetics*, 13(1), pp. 1–20. doi: 10.1371/journal.pgen.1006578.

Nielsen, B. L., Willis, V. C. and Lin, C. (2007) ‘Western Blot Analysis to Illustrate Relative Control Levels of the lac and ara Promoters in Escherichia coli’, *Biochemistry and Molecular Biology Education*, 35(2), pp. 133–137. doi: 10.1002/bmb.25.

Page, D. C. *et al.* (1987) ‘The sex-determining region of the human Y chromosome encodes a finger protein’, *Cell*, 51(6), pp. 1091–1104. doi: 10.1016/0092-8674(87)90595-2.

Persikov, A. V. and Singh, M. (2011) ‘An expanded binding model for Cys2His2 zinc finger protein-DNA interfaces’, *Physical Biology*, 8(3). doi: 10.1088/1478-3975/8/3/035010

Persikov, A. V. *et al.* (2015) ‘A systematic survey of the Cys2His2 zinc finger DNA-binding landscape’, *Nucleic Acids Research*, 43(3), pp. 1965–1984. doi: 10.1093/nar/gku1395.

Piskacek, S. *et al.* (2007) ‘Nine-amino-acid transactivation domain: Establishment and prediction utilities’, *Genomics*, 89(6), pp. 756–768. doi: 10.1016/j.ygeno.2007.02.003.

Piskacek, M. *et al.* (2017) ‘The 9aaTAD is exclusive activation domain in Gal4’, *PLoS ONE*, 12(1), pp. 1–13. doi: 10.1371/journal.pone.0169261.

Piskacek, M. *et al.* (2019) ‘The evolution of the 9aaTAD domain in Sp2 proteins: Inactivation with valines and intron reservoirs’, *bioRxiv*. doi: 10.1101/570598.

Poloumienko, A. (2004) ‘Cloning and comparative analysis of the bovine, porcine, and equine sex chromosome genes ZFX and ZFY’, *Genome*, 47(1), pp. 74–83. doi: 10.1139/g03-099.

Praud, D. *et al.* (2016) ‘Cancer incidence and mortality attributable to alcohol consumption’, *International Journal of Cancer*, 138(6), pp. 1380–1387. doi: 10.1002/ijc.29890.

Rogakou, E. P. *et al.* (1999) ‘Megabase chromatin domains involved in DNA double-strand breaks in vivo’, *Journal of Cell Biology*, 146(5), pp. 905–915. doi: 10.1083/jcb.146.5.905.

Rosano, G. L. and Ceccarelli, E. A. (2014) ‘Recombinant protein expression in Escherichia coli : advances and challenges’, *Frontiers in Microbiology*, 5(172), pp. 1–17. doi: 10.3389/fmicb.2014.00172.

Royo, H. *et al.* (2010) ‘Evidence that meiotic sex chromosome inactivation is essential for male fertility’, *Current Biology*, 20(23), pp. 2117–2123. doi: 10.1016/j.cub.2010.11.010.

Saida, F. *et al.* (2006) ‘Expression of Highly Toxic Genes in E. coli: Special Strategies and Genetic Tools’, *Current Protein and Peptide Science*, 7(1), pp. 47–56. doi: 10.2174/138920306775474095.

Skrisovska, L. *et al.* (2007) ‘The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction’, *EMBO Reports*, 8(4), pp. 372–379. doi: 10.1038/sj.embor.7400910.

Sommer, R. J. *et al.* (1992) ‘Evolutionary conservation pattern of zinc-finger domains of Drosophila segmentation genes’, *Proceedings of the National Academy of Sciences of the United States of America*, 89(22), pp. 10782–10786. doi: 10.1073/pnas.89.22.10782.

Stoyanov, G. S. *et al.* (2017) ‘Demographics of Head and Neck Cancer Patients : A Single Institution Experience’, *Cureus*, 9(7), p. E1418. doi: 10.7759/cureus.1418.

Stubbs, L., Sun, Y. and Caetano-Anolles, D. (2011) ‘Function and Evolution of C2H2 Zinc Finger Arrays’, in Hughes, T. R. (ed.) *A Handbook of Transcription Factors*. Dordecht: Springer Netherlands, pp. 75–94. doi: 10.1007/978-90-481-9069-0_4.

Swann, J. B. and Smyth, M. J. (2007) ‘Immune surveillance of tumors’, *The Journal of Clinical Investigation*, 117(5), pp. 1137–1146. doi: 10.1172/JCI31405.antigens.

Tricoli, J. V. and Bracken, R. B. (1993) ‘ZFY gene expression and retention in human prostate adenocarcinoma’, *Genes, Chromosomes and Cancer*, 6(2), pp. 65–72. doi: 10.1002/gcc.2870060202.

Tsuei, D. J. *et al.* (2004) ‘RBMY, a male germ cell-specific RNA-binding protein, activated in human liver cancers and transforms rodent fibroblasts’, *Oncogene*, 23(34), pp. 5815–5822. doi: 10.1038/sj.onc.1207773.

Tsuei, D. J. *et al.* (2011) ‘Male germ cell-specific RNA binding protein RBMY: A new oncogene explaining male predominance in liver cancer’, *PLoS ONE*, 6(11), pp. 1–9. doi: 10.1371/journal.pone.0026948.

Turner, J. M. A. (2007) ‘Meiotic sex chromosome inactivation’, *Development*, 134(10), pp. 1823–1831. doi: 10.1242/dev.000018.

Venables, J. P. *et al.* (1999) ‘T-STAR/ETOILE: A novel relative of SAM68 that interacts with an RNA-binding protein implicated in spermatogenesis’, *Human Molecular Genetics*, 8(6), pp. 959–969. doi: 10.1093/hmg/8.6.959.

Venables, J. P. (2000) ‘RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing’, *Human Molecular Genetics*, 9(5), pp. 685–694. doi: 10.1093/hmg/9.5.685.

Vernet, N. *et al.* (2011) ‘The Y-encoded gene Zfy2 acts to remove cells with unpaired chromosomes at the first meiotic metaphase in male mice’, *Current Biology*, 21(9), pp. 787–793. doi: 10.1016/j.cub.2011.03.057.

Vernet, N. *et al.* (2014) ‘Mouse Y-Linked Zfy1 and Zfy2 Are Expressed during the Male-Specific Interphase between Meiosis I and Meiosis II and Promote the 2nd Meiotic Division’, *PLoS Genetics*, 10(6). doi: 10.1371/journal.pgen.1004444.

Vernet, N. *et al.* (2016) ‘Mouse Y-encoded transcription factor Zfy2 is essential for sperm head remodelling and sperm tail development’, *PLoS ONE*, 11(1), pp. 1–15. doi: 10.1371/journal.pone.0145398.

Vernet, N. *et al.* (2016) ‘Zfy genes are required for efficient meiotic sex chromosome inactivation (MSCI) in spermatocytes’, *Human Molecular Genetics*, 25(24), pp. 5300–5310. doi: 10.1093/hmg/ddw344.

Vigneswaran, N. and Williams, M. D. (2014) ‘Epidemiologic Trends in Head and Neck Cancer and Aids in Diagnosis’, *Oral Maxillofac Surgery Clinics of North America*, 26(2), pp. 123–141. doi: 10.1016/j.coms.2014.01.001.

Wang, P. and Wilson, S. R. (2013) ‘Mass spectrometry-based protein identification by integrating de novo sequencing with database searching.’, *BMC bioinformatics*. BioMed Central Ltd, 14 Suppl 2(Suppl 2), p. S24. doi: 10.1186/1471-2105-14-S2-S24.

Wang, Z. *et al.* (2011) ‘Optimizing expression and purification of an ATP-binding gene gsiA from Escherichia coli k-12by using GFP fusion’, *Genetics and Molecular Biology*, 34(4), pp. 661–668. doi: 10.1590/S1415-47572011005000043.

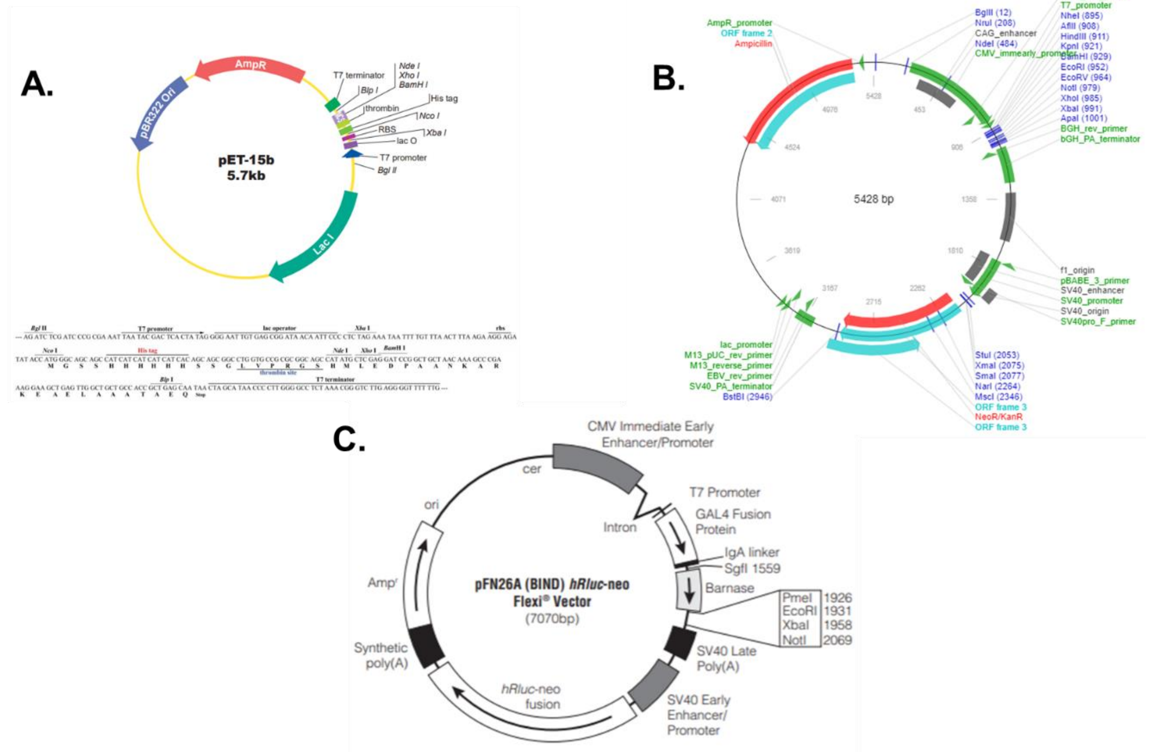
Yamauchi, Y. *et al.* (2015) ‘Mouse Y-Encoded Transcription Factor Zfy2 Is Essential for Sperm Formation and Function in Assisted Fertilization’, *PLoS Genetics*, 11(12), pp. 1–18. doi: 10.1371/journal.pgen.1005476.

Zandarashvili, L. *et al.* (2015) ‘Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble’, *Proceedings of the National Academy of Sciences of the United States of America*, 112(37), pp. E5142–E5149. doi: 10.1073/pnas.1507726112.

Zerbs, S., Giuliani, S. and Collart, F. (2014) *Small-scale expression of proteins in E. coli*, *Methods in Enzymology*. Edited by J. Lorsch. Elsevier Inc. doi: 10.1016/B978-0-12-420070-8.00011-8.

Zuo, Z. *et al.* (2019) *Why Do Long Zinc Finger Proteins have Short Motifs?*, *bioRxiv*. doi: 10.1101/637298.

6 Supplementary data



Supplementary Figure 1. Vector plasmid maps used for cloning hZFY genes. (**A.**) pET15b plasmid map, (**B.**)pcDNA3.1(+) plasmid map, (**C.**) pFN26A (BIND) plasmid map