A Genomic Resource for the Strawberry Powdery Mildew Pathogen *Podosphaera aphanis*

Thomas Heaven[1,3], Helen M. Cockerton[1,2], Xiangming Xu[1], Matthew Goddard[3], Andrew D. Armitage[4]

[1] NIAB, New Road, East Malling, Kent, ME19 6BJ, UK

[2] Kent Fungal Group, School of Biosciences, University of Kent, Canterbury, Kent, CT2 7NZ, UK

[3] School of Life Sciences, University of Lincoln, Lincoln, Lincolnshire, LN6 7DL, UK

[4] National Resources Institute, University of Greenwich, Kent, ME4 4TB, UK

Corresponding author: Andrew Armitage

Abstract

Powdery mildew is one of the most economically destructive diseases in protected strawberry production. Here we present the first genome assembly for *Podosphaera aphanis*, the causal agent of powdery mildew on strawberry. This obligate-biotrophic fungal pathogen was sampled from a naturally occurring outbreak on *Fragaria* × *ananassa* 'Malling Centenary' plants grown under cover in the UK. Assembled reads resolved a 55.6 Mb genome, composed of 12,357 contigs whose annotation led to prediction of 17,239 genes encoding 17,328 proteins. The genome is highly-complete, with 97.5 % of conserved single copy Ascomycete genes shown to be present. This annotated *P. aphanis* genome provides a molecular resource for further investigation into host-pathogen interactions in the strawberry powdery mildew pathosystem.

Additional Keywords:

Biotroph; Mycology; Horticulture; *Sphaerotheca macularis*; *Fragaria*; *Rubus*

Genome Announcement

Powdery mildew, caused by *Podosphaera aphanis* (formerly *Sphaerotheca macularis*), is an economically destructive disease affecting strawberry production around the world. Powdery mildew has been rated as the most important aerial disease for strawberries grown under protection by UK growers (Calleja, 2011; Menzel, 2021). All aerial plant tissues can be affected by the pathogen, with epidemics leading to severe yield loss as infection of leaves reduces photosynthesis and infection of fruits renders them unmarketable (Hibberd et al., 1996; Maas, 1998). Most

commercial strawberry varieties are considered highly susceptible and where resistant cultivars have been identified, this resistance frequently varies across environmental conditions (Nelson et al. 1996; Masny et al., 2016; Cockerton et al, 2018; Sargent et al., 2019; Menzel, 2021).

Disease control is primarily achieved through foliar fungicide applications. However, there is increasing pressure to find other non-fungicidal control methods, due to a desire to reduce agrochemical inputs and in response to reduced sensitivity to fungicides observed in field populations of *P. aphanis* (Palmer and Holmes, 2021). UV-C irradiation has been shown to suppress epidemics and a recent study demonstrated that a biopesticide-based approach can manage the disease effectively (Janisiewicz et al., 2016; Berrie & Xu, 2021).

Powdery mildew fungi are obligate biotrophs and thus dependent on living host cells for their survival. *P. aphanis* is understood to have a restricted host range, with evidence for host specialisation within *P. aphanis* populations (Harvey and Xu, 2010; Martin et al. 2017). *P. aphanis* is also considered the causal agent for powdery mildew of *Rubus* crops including raspberry and blackberry and has been reported on a limited number of other species (Garibaldi et al. 2005; Solano-Báez et al. 2021). However, whilst they are considered the same species, strawberry and raspberry powdery mildew isolates have been shown to be genetically distinct (Harvey and Xu, 2010). These genetic differences may reflect host-specialisation, with evidence that isolates from strawberry are unable to infect raspberry and vice versa (Martin et al. 2017).

Genomic approaches offer opportunity to address key questions such as the true host range, population structure and nature of fungicide resistance in *P. aphanis*. Despite the high commercial impact of pathogens from the *Podosphaera* genus, only four genomes have been sequenced to-date (Gañán et al., 2020; Kim et al. 2021; Polonio et al. 2021). Availability of wider sequence data is particularly limited for *P. aphanis*, with only 71 nucleotide sequences currently available on

NCBI (text search 'Podosphaera aphanis' against 'Nucleotide' database at www.ncbi.nlm.nih.gov, 10th March 2022), with all less than 1,500 bp in length. Genomic studies of powdery mildew species have been hampered due to the obligate biotrophic life cycle these fungi. However, this is now changing with advances in sequencing technology and new assembly methods. Powdery mildew species such as the cereal grass pathogen *Blumeria graminis*, and the cucurbit pathogen *Podosphaera xanthii* now have multiple genomes publicly available (Spanu et al. 2010; Frantzeskakis et al. 2018; Kim et al. 2021; Polonio et al. 2021). To facilitate future genomic investigation of strawberry powdery mildew, we present the first draft genome of *P. aphanis*, obtained through a whole-genome shotgun sequencing approach.

Powdery mildew material (isolate DRCT72020) was sampled from a naturally occurring outbreak of *P. aphanis* at NIAB EMR, Kent, UK in 2020. Leaves from ~30 severely affected *Fragaria* × *ananassa* cv. Malling Centenary plants were collected and immediately washed with water to remove conidia. These conidial suspensions were centrifuged at 5000 g for 5 mins and the supernatant discarded. Purified conidial samples were freeze dried overnight, transferred to 1.5 ml Eppendorf tubes and stored at -80 ºC. Genomic DNA was extracted using the protocol developed by Schwessinger and McDonald (2017), modified with extended one hour phenol/chloroform wash steps and overnight precipitation at 4 ºC. DNA concentration was assessed via Qubit dsDNA HS assay kit using a Qubit 3.0 fluorometer (Life Technologies, Waltham, MA USA). A partial sequence of the ribosomal internal transcribed spacer (ITS) region was amplified using ITS1 (5′-TCCGTAGGTGAACCTGCG-3′) and ITS4 (5′-TCCTCCGCTTATTGATATGC-3′) primers (White et al. 1990). Geneious (Kearse et al. 2012) was used to align the resulting amplicon to a reference *P. aphanis* ITS region (GenBank accession no. MF919432.1), confirming the sample identity. Genomic DNA was used for library preparation and paired-end sequencing on an Illumina

NovaSeq (insert size 350 bp, read length 150 bp) (Novogene Bioinformatics Technology Co., Ltd, Cambridge, UK).

The resulting 349,282,679 read pairs were trimmed and adapters removed using Trimmomatic v0.39 (Bolger et al. 2014), using paired end mode and -phred33 options. Trimmed reads were aligned to the *F.* × *ananassa* 'camarosa' genome (Edger et al. 2019) using bowtie2 v2.4.1, (Langmead and Salzberg 2012), with those aligning to the strawberry host omitted from further analysis. Unaligned reads were used for genome assembly via SPAdes v3.14.1 with the --isolate option and a coverage cut-off setting of 75 (Bankevich et al. 2012). Kmer analysis using kraken2 v2.1.1 (Wood et al. 2019) allowed taxonomic classification of the resulting contigs by alignment to a custom database including standard databases for archaea, bacteria, fungi, plants, protozoa, viral and mammals with the addition of the *F.* × *ananassa* 'camarosa' genome (Edger et al., 2019) and 29 powdery mildew genomes downloaded from the NCBI database (Supplementary Table S1). Only contigs assigned to the fungal class Leotiomycetes were taken forward to from the final assembly (Fig. 1).

This yielded a final genome of 55,605,580 bp in 12,357 contigs (≥500 bp) with an N_{50} value of 11,409 and GC content of 43.06 %. A homologous sequence with 100% identity to the *P. aphanis* ITS amplicon was identified in contig_4920 through a BLASTn search against the genome. Genome completeness was assessed via the universal single copy orthologue tool BUSCO V5.0.0 (Simão et al. 2015) which identified 1,664 conserved Ascomycete genes (ascomycota_odb10 database) as complete in the *P. aphanis* genome (Table 1). A *de novo* prediction of repetitive elements was performed using RepeatModeler v2.0.2 (Flynn et al. 2020) and TransposonPSI (Haas, 2010), with 53.65% of the genome identified as repetitive elements.

In order to facilitate gene prediction RNA-Seq was performed. Infected strawberry leaves were flash frozen in liquid nitrogen prior to RNA extraction using 3 % CTAB extraction buffer as described in Yu et al. 2012 with the following modifications; chloroform:isoamyl alcohol (24:1) washing was omitted and precipitation was performed at -20 °C for four hours. The resulting RNA concentration and RNA Integrity Number (RIN) of samples was assessed using the Agilent RNA ScreenTape System with a 2,200 TapeStation (Agilent Technologies Inc., Germany) according to the manufacturer's protocols. Library construction and sequencing was performed via Illumina HiSeq at Novogene Bioinformatics Technology Co., Ltd (Cambridge, UK). The resulting 88,436,408 read pairs were subjected to a quality control check using FastQC, with sequences then trimmed and adapters removed using Trimmomatic. Reads were aligned to the draft *P. aphanis* genome assembly using STAR v2.7.3 (Dobin et al. 2013). Gene prediction was performed on the repeat-masked (softmasked) genome using BRAKER v1.9 (Hoff et al. 2019) Codingquarry v2.0 in pathogen mode (Testa et al. 2015), trained with the aligned RNA-Seq data. BRAKER gene models were used preferentially, supplemented by Codingquarry genes when these were entirely located in intergenic regions, as described in Armitage et al. (2018). A total of 17,239 genes were predicted encoding 17,328 proteins. Of these, 15,492 predicted genes originated from Braker and 1,747 from CodingQuarry. Predicted proteins were functionally annotated via Interproscan 5 v44-79.0 (Jones et al. 2014), as well as a blastp search against the Swiss-Prot database (downloaded September 2021) (Boeckmann et al. 2003), which identified homologs to 1,756 predicted proteins.

This first draft genome assembly and gene models for *P. aphanis* provides a resource that will facilitate further investigation of the genomics, transcriptomics and host-pathogen interactions in this economically important pathogen.

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAKRRZ000000000 (BioProject number PRJNA744412). The version described in this paper is GenBank accession number GCA_022627015.2. All sequencing data has been deposited at the NCBI Sequence Read Archive under the accession numbers SRR18158617 (Illumina NovaSeq raw reads) and SRR18158616 (Illumina RNAseq raw reads). Sanger sequence data for the *P. aphanis* ITS1-4 region has been deposited at GenBank under accession ON597238.

Acknowledgements

This work was funded by the Biotechnology and Biological Research Council (BBSRC) CTP PhD scheme (BB/S507180/1), Sequencing costs were supported by a BBSRC seedling grant and The Worshipful company of Fruiterers.

References

Armitage, A.**,** Taylor, A., Sobczyk, M., Baxter, L., Greenfield, B., Bates, H., et al. 2018. Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepae*. Scientific Reports. 8:13530.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477.

Berrie, A. And Xu, X. 2021. Developing biopesticide-based programmes for managing powdery mildew in protected strawberries in the UK. Crop Protection. 149:105766.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., et al. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31:365–370.

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Calleja, E.J. 2011. The potential impacts of climate change on diseases affecting strawberries and the UK strawberry industry. PhD thesis, University of Warwick.

Cockerton, H.M., Vickerstaff, R.J., Karström, A., Wilson, F., Sobcyk, M., He, J.Q., Sargent, D.J., Passey, A.J., McLeary, K.J., Pakozdi, K., Harrison, N., Lumbreras-Martinez, M., Antanaviciute, L., Simpson, D.W. and Harrison, R.J. 2018. Identification of powdery mildew resistance QTL in strawberry (Fragaria × ananassa). Theoretical and Applied Genetics, 131, 1995–2007.

Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., et al. 2019. Origin and evolution of the octoploid strawberry genome. Nat. Genet. 51:541–547.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 117:9451–9457.

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., et. al. 2018. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. BMC Genomics 19, 381

Gañán, L., White Iii, R., Friesen, M., Peever, T., and Amiri, A. 2020. A genome resource for the apple powdery mildew pathogen Podosphaera leucotricha. Phytopathology. 110:1756–1758.

Garibaldi, A., Bertetti, D., and Gullino, M. L. 2005. First Report of Powdery Mildew Caused by *Podosphaera aphanis* var. aphanis on *Potentilla fruticosa* in Italy. Plant Dis. 89:1362.

Haas, B. 2010. TransposonPSI: An Application of PSI-Blast to Mine (Retro-)Transposon ORF Homologies. Available at: http://transposonpsi.sourceforge.net/ [Accessed October 25, 2021].

Harvey, N., Xu, X. 2010. Powdery mildew on raspberry is genetically different from strawberry powdery mildew. Journal of Plant Pathology 92:775–779.

Hibberd, J.M., Richardson, P., Whitbread, R., and Farrar, J.F. 1996. Effects of leaf age, basal meristem and infection with powdery mildew on photosynthesis in barley grown in 700 µmol mol^{-1} CO2. New Phytologist. 134: 317–325.

Hoff, K. J., Lomsadze, A., Borodovsky, M., and Stanke, M. 2019. Whole-Genome Annotation with BRAKER. Methods Mol. Biol. 1962:65–95.

Janisiewicz, W.J., Takeda, F., Nichols, B., Glenn, D.M., Jurick II, W.I. and Camp, M.J. 2016. Use of low-dose UV-C irradiation to control powdery mildew caused by *Podosphaera aphanis* on strawberry plants. Canadian Journal of Plant Pathology, 38:430-439.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30:1236–1240.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649.

Kim, S., Subramaniyam, S., Jung, M., Oh, E.-A., Kim, T. H., and Kim, J.-G. 2021. Genome Resource of *Podosphaera xanthii*, the Host-Specific Fungal Pathogen That Causes Cucurbit Powdery Mildew. Mol. Plant Microbe Interact. 34:457–459.

Langmead, B., and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357–359.

Maas, J.L. 1998. Compendium of strawberry diseases (pp. 98). St. Paul, USA: American Phytopathological Society.

Martin, R. R., Ellis, M. A., Williamson, B., and Williams, R. N., eds. 2017. Compendium of raspberry and blackberry diseases and pests, second edition. The American Phytopathological Society.

Masny, A., Masny, S., Żurawicz, E., Pruski, K., and Mądry, W. 2016. Suitability of certain strawberry genotypes for breeding of new cultivars tolerant to leaf diseases based on their combining ability. Euphytica, 210, 341–366.

Menzel, M. 2021. A review of powdery mildew in strawberries: the resistance of species, hybrids and cultivars to the pathogen is highly variable within and across studies with no standard method for assessing the disease. The Journal of Horticultural Science and Biotechnology. 1:1-25.

Nelson, M.D., Gubler, W.D. and Shaw, D.V. 1996. Relative resistance of 47 strawberry cultivars to powdery mildew in California greenhouse and field environments. Plant Disease: 80, 326–328.

Polonio, Á., Díaz-Martínez, L., Fernández-Ortuño, D., de Vicente, A., Romero, D., López-Ruiz, F. J., et al. 2021. A Hybrid Genome Assembly Resource for *Podosphaera xanthii,* the Main Causal Agent of Powdery Mildew Disease in Cucurbits. Mol. Plant Microbe Interact. 34:319–324.

Schwessinger, B., and McDonald, M. 2017. High quality DNA from Fungi for long read sequencing e.g. PacBio, Nanopore MinION V.2. Available at: https://www.protocols.io/view/high-quality-dna-from-fungi-for-long-read-sequenci-hadb2a6 [Accessed October 25, 2021]

Sargent, D.J., Buti, M., Šurbanovski, N., Brurberg, M.B., Alsheikh, M., Kent, M.P. and Davik, J. 2019. Identification of QLTs for powdery mildew (*Podosphaera aphanis*; syn. *Sphaerotheca macularis* f. sp. *fragariae*) susceptibility in cultivated strawberry (*Fragaria* × *ananassa*). PLoS One, 14, e0222829.

Solano-Báez, A. R., Leyva-Mir, S. G., Camacho-Tapia, M., Victoria, A. A., Rodríguez-Bautista, G., Sánchez-Rosas, C. S., et al. 2021. First Report of *Podosphaera aphanis* Causing Powdery Mildew on Wild Blackberry Species (Rubus spp.) in Mexico. Plant Dis.

Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stüber, K., et al. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science. 330:1543–1546.

Testa, A. C., Hane, J. K., Ellwood, S. R., and Oliver, R. P. 2015. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. BMC Genomics. 16:170.

Wood, D. E., Lu, J., and Langmead, B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol. 20:257.

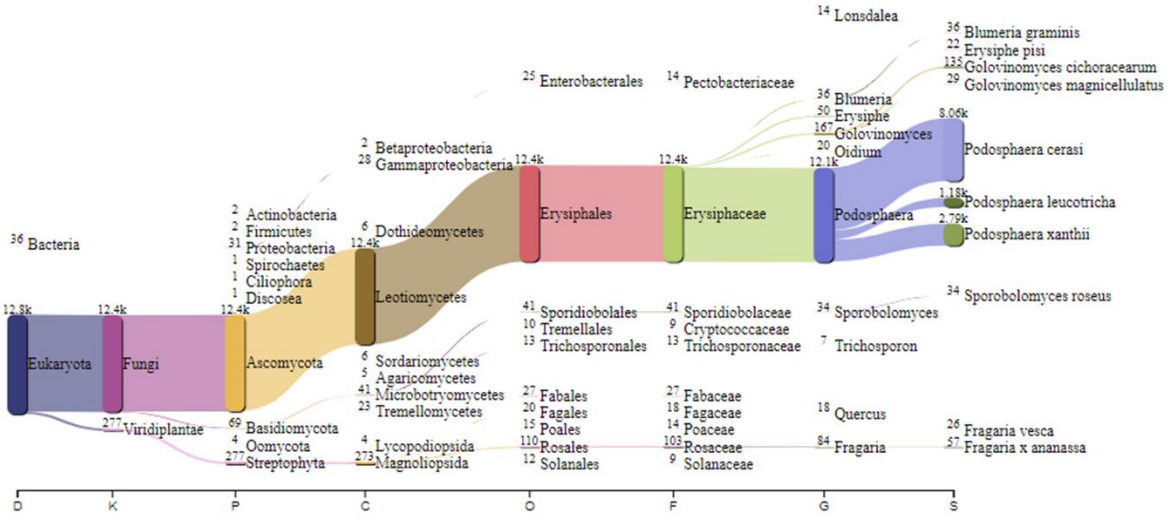
White, T.J., Bruns, T., Lee, S., and Taylor, J. 1990 Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-322 in: PCR Protocols: a guide to methods and applications, Academic Press, New York, USA

Yu, D., Tang, H., Zhang, Y., Du, Z., Yu, H., and Chen, Q. 2012. Comparison and Improvement of Different Methods of RNA Isolation from Strawberry (*Fragria* × *ananassa*). JAS. 4.

Table 1. Summary of genome assembly statistics for *Podosphaera* species, from left to right; *Podosphaera aphanis*, *Podosphaera leucotricha* (Gañán et al. 2020), *Podosphaera xanthii* (Polonio et al. 2021) and *Podosphaera xanthii* (Kim et al. 2021).

Species	***P. aphanis***	***P. leucotricha***	***P. xanthii***	***P.xanthii***
Isolate	DRCT72020	PuE-3	2086	Wanju2017
Total length (bp)	55,613,046	43,868,508	142,114,041	209,067,775
Number of contigs	12,357	8,921	1,727	1,112
Number of contigs ≥ 1000 bp	7,859	8,921	1,598	1,112
Size of largest contig (bp)	77,136	60,133	947,834	2,325,138
Contig N_{50} (bp)	11,409	8,371	163,173	581,650
Contig N_{75} (bp)	5,053	4,117	84,907	252,521
GC content (%)	43.06	43.69	43.23	44.25
Repeat elements (%)	53.77	77.8	76.16	63.41
BUSCO complete (%)	97.5	96.1	95.5	97.9
BUSCO duplicated (%)	0.3	0.4	1	20.8
BUSCO fragmented (%)	0.5	1.7	1.2	0.2
Number of predicted genes	17,284	9,372	16,030	12,834

Figure 1 Kmer-based assignment of assembled contigs from the environmental *P. aphanis* sample to reference genomes. Contigs assigned to Leotiomycetes were taken forward to form the final *P. aphanis* assembly.



Supplementary Table S1. Mildew genomes used in contig classification.

Species	Strain/isolate	Host	GenBank assembly accession
Blumeria graminis f. sp. *hordei*	RACE1	Barley	GCA_900237765.1
Blumeria graminis f. sp. *hordei*	DH14	Barley	GCA_900239735.1
Blumeria graminis f. sp. *hordei*	K1	Barley	GCA_900638725.1
Blumeria graminis f. sp. *hordei*	A6	Barley	GCA_000401675.1
Blumeria graminis f. sp. *triticale*	THUN-12	Triticale	GCA_905067625.1
Blumeria graminis f. sp. *tritici*	96224	Wheat	GCA_000418435.1
Blumeria graminis f. sp. *tritici*	Bgt#70	Wheat	GCA_000441875.1
Blumeria graminis f. sp. *tritici*	94202	Wheat	GCA_000417865.1
Blumeria graminis f. sp. *tritici*	JIW2	Wheat	GCA_000417025.1
Blumeria graminis f. sp. *tritici*		Wheat	GCA_900519115.1
Erysiphe necator	c	Grapevine	GCA_000798715.1
Erysiphe necator	NAFU1	Grapevine	GCA_016906895.1
Erysiphe necator	e1-101	Grapevine	GCA_000798795.1
Erysiphe necator	lodi	Grapevine	GCA_000798775.1
Erysiphe necator	ranch-9	Grapevine	GCA_000798755.1
Erysiphe necator	branching	Grapevine	GCA_000798735.1
Erysiphe pisi		Pea	GCA_000208805.1
Erysiphe pisi		Pea	GCA_000214055.1
Erysiphe pulchra	Cflorida	Flowering Dogwood	GCA_002918395.1
Golovinomyces cichoracearum	UMSG1	Sow thistle	GCA_003611235.1
Golovinomyces cichoracearum	UMSG3	Tobacco	GCA_003611195.1
Golovinomyces cichoracearum	UCSC1	Arabidopsis	GCA_003611215.1
Golovinomyces magnicellulatus	FPH2017-1	*Phlox paniculata*	GCA_006912115.1
Oidium heveae	HO-73	Rubber tree	GCA_003957845.1
Oidium neolycopersicim	UMSG2	Tomato	GCA_003610855.1
Podosphaera cerasi	MH	Sweet Cherry	GCA_018398735.1
Podosphaera leucotricha	PuE-3	Apple	GCA_013170925.1
Podosphaera xanthii	Wanju2017	Cucumber	GCA_010015925.1
Podosphaera xanthii	2086	Courgette	GCA_014884795.1