

RESEARCH

Open Access



The massive 340 megabase genome of *Anisogramma anomala*, a biotrophic ascomycete that causes eastern filbert blight of hazelnut

Alanna B. Cohen^{1,2}, Guohong Cai^{3,4*}, Dana C. Price^{5,6}, Thomas J. Molnar¹, Ning Zhang^{1,2,7} and Bradley I. Hillman^{1,2*}

Abstract

Background The ascomycete fungus *Anisogramma anomala* causes Eastern Filbert Blight (EFB) on hazelnut (*Corylus* spp.) trees. It is a minor disease on its native host, the American hazelnut (*C. americana*), but is highly destructive on the commercially important European hazelnut (*C. avellana*). In North America, EFB has historically limited commercial production of hazelnut to west of the Rocky Mountains. *A. anomala* is an obligately biotrophic fungus that has not been grown in continuous culture, rendering its study challenging. There is a 15-month latency before symptoms appear on infected hazelnut trees, and only a sexual reproductive stage has been observed. Here we report the sequencing, annotation, and characterization of its genome.

Results The genome of *A. anomala* was assembled into 108 scaffolds totaling 342,498,352 nt with a GC content of 34.46%. Scaffold N50 was 33.3 Mb and L50 was 5. Nineteen scaffolds with lengths over 1 Mb constituted 99% of the assembly. Telomere sequences were identified on both ends of two scaffolds and on one end of another 10 scaffolds. Flow cytometry estimated the genome size of *A. anomala* at 370 Mb. The genome exhibits two-speed evolution, with 93% of the assembly as AT-rich regions (32.9% GC) and the other 7% as GC-rich (57.1% GC). The AT-rich regions consist predominantly of repeats with low gene content, while 90% of predicted protein coding genes were identified in GC-rich regions. Copia-like retrotransposons accounted for more than half of the genome. Evidence of repeat-induced point mutation (RIP) was identified throughout the AT-rich regions, and two copies of the *rid* gene and one of *dim-2*, the key genes in the RIP mutation pathway, were identified in the genome. Consistent with its homothallic sexual reproduction cycle, both MAT1-1 and MAT1-2 idiomorphs were found. We identified a large suite of genes likely involved in pathogenicity, including 614 carbohydrate active enzymes, 762 secreted proteins and 165 effectors.

Conclusions This study reveals the genomic structure, composition, and putative gene function of the important pathogen *A. anomala*. It provides insight into the molecular basis of the pathogen's life cycle and a solid foundation for studying EFB.

*Correspondence:

Guohong Cai
guohong.cai@usda.gov

Bradley I. Hillman
bradley.hillman@rutgers.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords Ascomycete, Plant pathogen, Repeat-induced point mutation, Mating type gene, Transposon, Effector, *Anisogramma anomala*, Eastern filbert blight

Background

The investigation of biotrophic fungi – pathogens that require living host tissue – is complex and challenging. Because of their dependency on the host organism, biotrophs are difficult to isolate and grow in artificial media. They often have strict nutritional requirements and may require certain hormones or signaling chemicals secreted by the host to induce spore germination [1, 2]. Satisfying these conditions complicates any form of manipulation under laboratory conditions. Studies of rust fungi, which are basidiomycetes, and powdery mildew fungi, which are ascomycetes, highlight many of these challenges. Despite the significant economic impact of the resulting diseases, complete life cycles of these fungi have never been witnessed outside of their natural hosts. Consequently, despite substantial effort on the parts of many scientists, many details of host–pathogen interactions in rust and powdery mildew fungi remain poorly understood [3–5].

Advances in sequencing and bioinformatic tools have led to the rapid development of genomic techniques that facilitate investigation even of recalcitrant organisms. As the number of sequenced fungal genomes expands, patterns and features that are linked to obligate biotrophy have emerged [6, 7]. Genomic features, including both coding and non-coding elements, reveal characteristics of lifestyle and pathogen biology [8, 9]. A large repertoire of species-specific secreted small cysteine-rich proteins that represent candidate effectors is typical of biotrophs that have gene specific interactions with their host [10, 11]. Large genomes inflated with repetitive elements are another hallmark of biotrophic pathogens, as amplification of such elements contributes to a flexible genomic landscape that is highly adaptable to the gene-for-gene arms race that pathogens engage in with their hosts [11–13]. Identifying these characteristics of genomic features can fill in the blanks left by a lack of experimental data [14, 15].

One such fungal pathogen whose biology lacks understanding is *Anisogramma anomala*, an ascomycete within the order *Diaporthales*. *A. anomala* causes Eastern Filbert Blight (EFB), a devastating disease of hazelnut (*Corylus* spp.). The native host of *A. anomala*, American hazelnut (*C. americana*) tolerates infection, displaying mild disease symptoms and small, non-threatening cankers [16–18]. Both host and fungus are abundant on the east coast of the U.S. However, nearly all cultivars of the commercially important

European hazelnut (*C. avellana*) are highly susceptible and develop severe perennial cankers that girdle stems, resulting in branch die-back and eventual tree death [19–21]. As such, EFB is the primary limiting factor of commercial hazelnut production in North America [22]. Historically, *C. avellana* cultivation was restricted to the Pacific Northwest region outside of the native range of *A. anomala*, limiting hazelnut cultivation to a fraction of its potential growth range [23]. Today, after an inadvertent introduction in the 1960s [24], EFB is widespread in the Pacific Northwest where it significantly impacts commercial production. Disease management costs were alleviated only recently by the release of resistant cultivars [25]. Despite the economic importance of *A. anomala* and considerable efforts now underway to breed for resistance [26], the EFB pathosystem remains poorly understood.

To support disease management and resistance breeding efforts, there is a need for a better understanding of the biology of *A. anomala* and the EFB pathosystem. However, *A. anomala* is an obligate biotroph, presenting many of the methodological difficulties as do rust fungi, powdery mildews, and other biotrophic pathogens [27]. The only useful source of tissue of *A. anomala* is ascospores extracted from the stromata of cankers of infected hazelnut, and successful subculture has not been achieved. Ascospores represent the only known spore stage of *A. anomala*; no conidial stage has been documented. Ascospores, by nature, are sexual spores and are not isogenic. While *A. anomala* ascospores can germinate and form small, branching germ hyphae, the fungus cannot be grown continuously in culture. It is predicted that *A. anomala* exhibits some form of self-inhibition, as ascospores will germinate in axenic culture only with the addition of an adsorbent such as activated charcoal or bovine serum albumin (BSA) [27]. Even with these additives, germinated ascospores exhibit poor growth and form small colonies (~0.25–0.5 mm in diameter) that yield little biomass [28]. Furthermore, the disease exhibits a complex, two-year infection cycle, which normally includes 15–18 months of latency, in which it is not feasible to visibly identify infected trees (Figure S1) [20, 21, 29–31].

Despite the challenges to performing experimental host/pathogen research, we saw the importance in understanding more about *A. anomala*, both as contributions to the U.S. hazelnut industry, and to plant

pathogen biology. Due to the lack of an experimental system by which to study *A. anomala*, we used a genomic approach to elucidate features of EFB biology and pathogenesis. An earlier draft genome of *A. anomala* was assembled and mined for sequences that would be useful as simple sequence repeat (SSR) primers to examine population biology of the fungus and assist with resistance breeding [28]. That study revealed that the genome of *A. anomala* is surprisingly large, >300 megabases (Mb) and consists of an abundance of transposons that constitute nearly 90% of the genome sequence. In this study, we present an updated and refined draft of the *A. anomala* genome sequence, its annotation, and analysis. Genomic analysis reveals characteristics of biotrophy, including a massive population of transposable elements (TEs), bimodal distribution of GC content, and a cache of genes encoding effector molecules. We also identified a number of genes that code for proteins predicted to be involved in pathogenesis and host/pathogen interactions. The annotated genome of *A. anomala* will serve as a vital resource for future research on the pathogen and EFB disease.

Results

A. anomala has a large, gene-poor genome

The mate-pair and paired-end reads of genomic DNA for *Anisogramma anomala* OR1 generated over 31 Gb of data that were assembled into a 342,525,599 nucleotide (nt) genome with an average 91× coverage (Table S1). The final assembly was distributed across 112 scaffolds with a GC content of 34.46%. Four scaffolds with a combined length of 27,247 nt were removed from further analysis as contamination, resulting in a final assembly size of 342,498,352 nt across 108 scaffolds. More than half of the assembly (N50) was on 5 scaffolds with an N50 scaffold length of 33.3 Mb. The largest scaffold was 43.9 Mb (Table 1). Nineteen major scaffolds (>1 Mb) represent over 99% of the genome. This demonstrates a marked improvement over the first version of the assembly, published in 2013 [28] (Table S2). We identified telomere sequences (repeats of TTAGGG) on both ends of the second and third largest scaffolds with lengths of 40.1 and 39.2 Mb respectively, indicating these two scaffolds represent full-length chromosomes. Telomere sequences were also found on one end of 10 other scaffolds. Of the 19 largest scaffolds, telomere sequences were found on one end or both ends in 10 scaffolds (Fig. 1). On the contig level, the N50 was 196,655 bp and the L50 was 528 (Table 1).

To evaluate the completeness of the *A. anomala* genome assembly, we performed flow cytometry using nuclei released from 8-week old mycelium. Based on flow

Table 1 Table of assembly statistics and feature summary for the haploid genome of *A. anomala*

Assembly Statistics	
Scaffold Level	
Assembly size	342,498,352 bp
Total number	108
Largest	43,949,127 bp
N50	33,254,450 bp
L50	5
N99	1,308,775 bp
L99	19
# with telomere	12
% of gap region	0.37%
Contig Level	
Total number	3,310
N50	196,655 bp
L50	538
GC content	34.46%
Annotation Statistics	
Protein coding genes	9,179
Average gene length	1,737 bp
Average protein length	481 aa
Mean exon number per gene	3.3
Mean exon length	438 bp
Eukaryotic BUSCOs	94.4%
Fungal BUSCOs	95.5%
Carbohydrate active enzymes	614
Glycoside hydrolases	298
Glycosyl transferases	154
Auxillary activities	83
Carbohydrate esterases	41
Carbohydrate-binding molecules	28
Polysaccharide lyase	10
Predicted secreted proteins	762
Candidate effectors	165

cytometry, the genome size of *A. anomala* OR1 was estimated to be 370 Mb (Figure S2), slightly more than, but consistent with the genome assembly estimate.

Using a combination of RNA-seq evidence and ab initio gene prediction, we predicted 9,179 protein coding genes in the *A. anomala* genome. This gene set includes 94.4% of eukaryotic benchmarking universal single-copy orthologs (BUSCOs) and 95.5% of fungal BUSCOs. Average gene density on major scaffolds was approximately 25.8 genes/Mb and remained relatively consistent among major scaffolds (Fig. 1).

Gene models were annotated with Gene Ontology (GO) terms merged with InterPro IDs. Eighty-eight percent of gene models had BLASTp hits against the NCBI nr database. Approximately 75% of gene models have

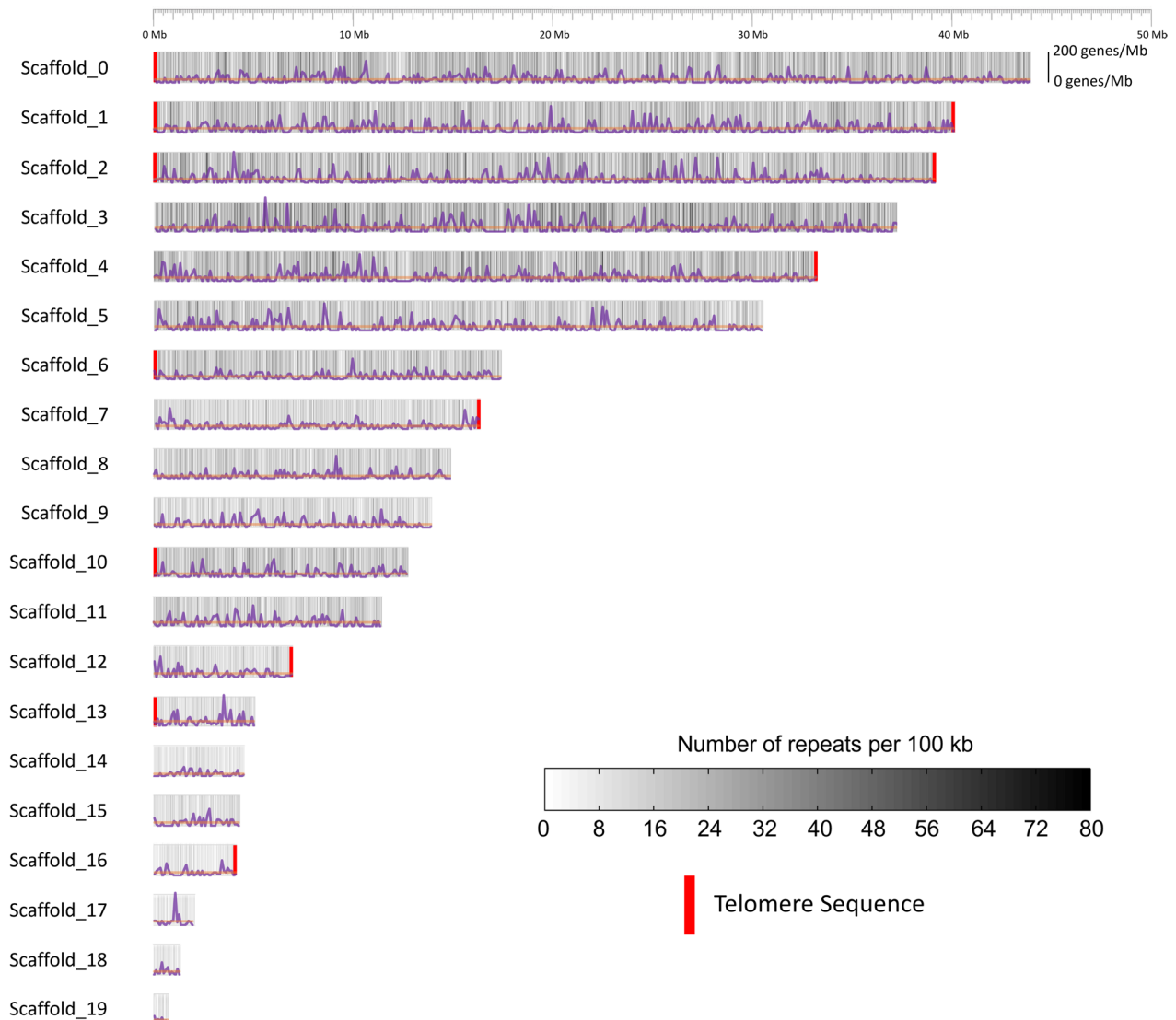


Fig. 1 Repeat content and gene density distribution across major scaffolds (> 1Mb). Size of bar reflects the length of the scaffold (x-axis). Repeat density of dispersed repeats in bins of 100kb is represented as a heat map, ranging from white to black, the darker the color indicating higher repeat density. The height of each scaffold bar (y-axis) ranges from 0 to 200 genes/Mb with the average gene density of the scaffold plotted in orange. Gene density was calculated per 100kb and plotted along each scaffold in purple

been annotated by biological process and 50% with a molecular function (Fig. 2, Table S3). Gene models were also annotated with KEGG Orthology (KO) terms, using a combination of the KEGG Automated Annotation Server and BlastKOALA. Roughly 38% of protein sequences were assigned KO identifiers, which make up 99 complete or nearly complete KEGG pathways (Table S3).

A. anomala has large arsenal of effectors and CAZymes

To identify proteins that may be involved in virulence and disease, we identified genes that code for potential

effectors, molecules that are involved in host/pathogen interactions. We first identified 762 proteins with signal peptides as evidence of a secreted protein. Those proteins were then analyzed with EffectorP 2.0 to further predict potential effector proteins. One hundred and sixty-five proteins (1.8% of total proteins. 21.7% of secreted proteins) were predicted to be effector candidates (Table 1). All effector candidates were subjected to a BLASTp search of the NCBI nr database. Over half (55%) of candidate effectors returned no BLAST hit, and of those that did return a hit, 42% were hypothetical proteins or proteins with unknown function. For those effector

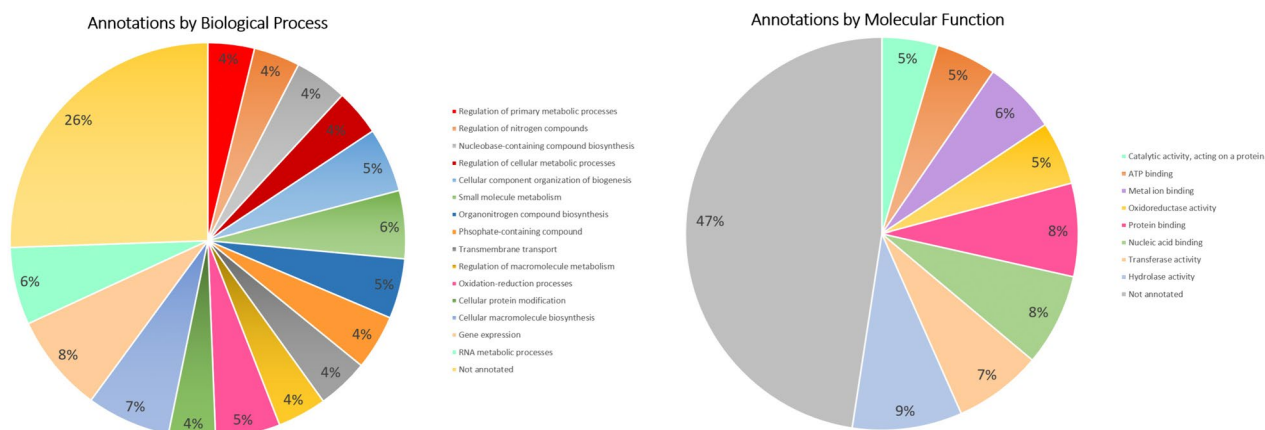


Fig. 2 Annotation of *A. anomala* predicted gene models by Gene Ontology (GO) categories. Functions of genes are shown by biological process and molecular functions

candidates that match a protein with a known function, possible roles include one glycoside hydrolase, one cutinase, and two peptidases (Table S4).

Genes encoding putative effector molecules were evaluated for their proximity to the closest repeat element and the closest large RIP affected region (LRAR) as predicted by RIPPER. BUSCOs and a random subset of all genes were included for comparison (Fig. 3). On average, effectors were approximately 1.5 kb from the nearest TE while BUSCOs and a randomized set were 3 kb and 2.5 kb respectively. The closest distance to LRARs for effectors, BUSCOs, and the randomized set did not differ significantly from each other and averaged at 9900 bp, 9400 bp, and 9700 bp respectively.

In addition to effector molecules, we also identified carbohydrate active enzymes (CAZymes) that may play a role in plant pathogenesis. Using the dbCAN3 meta server, we identified 614 potential CAZymes. These

proteins include 298 glycoside hydrolases, 154 glycosyl transferases, and 41 carbohydrate esterases (Table 1, Table S3). Finally, we identified biosynthetic gene clusters with the fungal version of antiSMASH. Twenty-five biosynthetic gene clusters were predicted, including 8 polyketide synthase (PKS), 7 terpene synthesis, 9 non-ribosomal peptide synthetase (NRPS) clusters, and 1 PKS/NRPS combination cluster (Table S5).

Genome and annotation statistics including genome size, repeat content, and different categories of protein coding genes (effectors, CAZymes, and biosynthetic gene clusters) were compared to related fungi (Table 2). Like other biotrophic fungi, *A. anomala* has a large genome with high repeat content (shown below). A large number of effectors (relative to total protein coding genes), small number of biosynthetic gene clusters and CAZymes are other hallmarks shared between *A. anomala* and related biotrophic fungi.

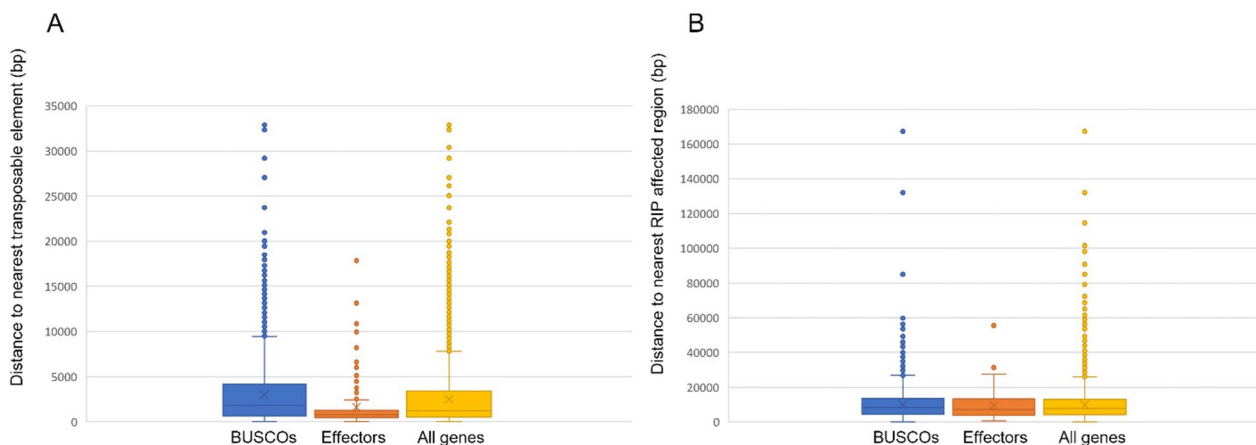


Fig. 3 **A** Comparison of distance of well-conserved genes (BUSCOs), predicted effectors, and a randomized set of gene models from the nearest repeat element or **B** large RIP affected region (LRAR) as predicted by RIPPER software

Table 2 Genomic statistics of related fungi used for comparison in this study

	Accession	Lifestyle	Host	Disease	Genome Size	GC Content	Repeat Content	Encoded proteins	CAZymes	Percent CAZymes of total genes	Secreted proteins	Percent secreted proteins of total	Predicted Effectors	Percent Effectors of total	Biosynthetic Gene Clusters	Publication
Anisogramma anomala OR1	BioProject PRJNA966177	Obligate Biotroph	Hazel-nut	Eastern Filbert Blight	342.5 Mb	34.46%	89.40%	9179	456	4.97	762	8.30	165	1.80	25	Current Study
Blumeria graminis f. sp. Tritici	GCA_905067625.1	Obligate Biotroph	Wheat	Powdery mildew	141.4 Mb	43.50%	75–90%	7,993	132	1.65	970	12.14	437	5.47	9	Wicker et al. 2013 [32]
Erysiphe neoelycopersici	GCA_003610855.1	Obligate biotroph	Tomato	Powdery mildew	42 Mb	38%	39.40%	6851	135	1.97	499	7.28	175	2.55	3	Wu et al. 2018 [33]
Erysiphe pulchra	GCA_002918395.1	Obligate biotroph	Dog-wood	Powdery mildew	63.5 Mb	38.50%	NA	6859	83	1.21	341	4.97	70	1.02	7	Wadi et al. 2019 [34]
Golovomyces orontii MGH1 v4.0		Obligate Biotroph	Wide range	Powdery mildew	211.35 Mb	42%	74.50%	12,649	668	5.28	157 (transcripts)	1.24	70 (transcripts)	0.55	24	JGI Project # 1,056,001; Micali et al. 2008 [35]
Melampsora larici populina	GCF_000204055.1	Obligate Biotroph	Poplar	Poplar leaf rust	101.1 Mb	41%	44.91%	16,257	308	1.89	1,796	11.05	1,184 (SSPs)	5.04	7	Duplessis et al. 2011 [36]
Paracaula polypora	PRJEB29715 (ENA)	Obligate biotroph	Asian Oak	Powdery mildew	28.0 Mb	46.50%	8.50%	6046	104	1.72	261	4.32	70	1.16	8	Frantzeskakis et al. 2019 [37]
Puccinia graminis tritici	GCF_000149925.1	Obligate Biotroph	Wheat/ cereals	Stem rust	88.6 Mb	43.30%	43.77%	15,979	203	1.27	1,386	8.67	1106 (SSPs)	6.92	2 ^a	Broad Institute ASMI14992V1
Ustilago maydis	GCF_000328475.2	Obligate Biotroph	Maize/ cereals	Corn smut	19.7 Mb	54%	1.10%	6780	139	2.05	426	6.28	203	2.99	13	Kamper et al. 2006 [38]
Taphrina deformans	GCA_000312925.2	Biotroph	Peach/ almond	Peach leaf curl	13.4 Mb	49.50%	1.50%	4663	231	4.95	256	5.49	24	0.51	4	Cisse et al. 2013 [39]
Fusarium gramine-arum	GCF_000240135.3	Hemibiotroph	Wheat/ barley/ maize	Fusarium head blight/ ear rot of corn	36.5 Mb	48%	0%	13,313	487	3.66	1442	10.83	190	1.43	67	Cuomo et al. 2007 [40]/King et al. 2015 [41]
Colletotrichum graminicola	GCF_000149035.1	Hemibiotroph	maize/ cereals	Anthrax-nose leaf blight	51.6 Mb	49%	22.30%	12,019	572	4.76	1650	13.73	314	2.61	42	O'Connell et al. 2012 [42]
Pyricularia grisea	GCF_004355905.1	Hemibiotroph	Rice	Rice blast	44.6 Mb	47.50%	12.61%	12,452	539	4.33	1491	11.97	490	3.94	45	Gomez-Luciano et al. 2019 [43]

Table 2 (continued)

	Accession	Lifestyle	Host	Disease	Genome Size	GC Content	Repeat Content	Encoded proteins	CAZymes	Percent CAZymes of total genes	Secreted proteins	Percent secreted proteins of total	Predicted Effectors	Percent Effectors of total	Biosynthetic Gene Clusters	Publication
Pyricularia oryzae	GCF_000002495.2	Hemibiotroph	Rice	Rice blast	41 Mb	51.50%	11.86%	12,989	538	4.14	752	5.79	370	2.85	30	Dean et al. 2005 [44]
Verticillium dahliae	GCA_000150675.2	Hemibiotroph	Wide range	Verticillium wilt	33.8 Mb	55.50%	5.00%	10,535	418	3.97	739	7.01	127	1.21	25	Klosterman et al. 2011 [45]
Botrytis cinerea	GCF_000143535.2	Necrotroph	wide range	Gray mold	42.6 Mb	41.50%	1%	11,698	1155	9.87	879	7.51	130	1.11	24^a	Van Kan et al. 2017 [46]; Anselem et al. 2011 [47]
Cryphonectria parasitica	GCF_011745365.1	Necrotroph	Chestnut	Chestnut blight	43.9 Mb	50.80%	12.10%	11,606	548	4.72	397	3.42	32	0.28	46	Crouch et al. 2020 [48]
Diaaporthe helianthi	GCA_0001702395.2	Necrotroph	Sunflower	Stem canker	63.7 Mb	43.50%	NA	13,139	516	3.93	1475	11.23	245	1.86	67	Baroncelli et al. 2016 [49]
Sclerotinia sclerotiorum	GCA_0001857865.1	Necrotroph	Wide range	White mold	38.9 Mb	41.50%	12.96%	11,130	411	3.69	603	5.42	57	0.51	15	Derbyshire et al. 2017 [50]
Valsa mali	GCA_000818155.1	Necrotroph	Apple	Apple canker	44.7 Mb	49%	14.10%	11,284	342	3.03	779	6.90	193	1.71	85	Yin et al. 2015 [51]
Nectria coccia	GCF_000151355.1	Saprotroph	Wide range	Root rot	51.3 Mb	50.50%	5.10%	15,708	889	5.66	1379	8.78	257	1.64	38	Coleman et al. 2009 [52]
Podospora anserina	GCF_000226545.1	Saprotroph	Non-pathic	NA	34.7 Mb	52%	5%	10,518	243	2.31	267	2.54	1064	10.12	31	Espagne et al. 2008 [52]
Stachybotrys charmarum	GCA_000730325.1	Saprotroph	Non-pathic	NA	36.9 Mb	53%	1.01%	11,530	383	3.32	1206	10.46	144	1.25	55	Semeiks et al. 2014 [53]
Neurospora crassa	GCA_000182925.2	Free-living	Non-pathic	NA	41 Mb	48%	10%	10,812	191	1.77	891	8.24	167	1.54	12	Galagan et al. 2003 [54]
Chaetomium globosum	GCF_000143365.1	Free-living	Non-pathic	NA	34.3 Mb	56%	6%	11,040	146	1.32	1094	9.91	177	1.60	45	Cuomo et al. 2015 [55]

Bold font indicates missing information from the respective study, numbers were predicted using the same methods used in the current study
^a Indicates missing CDS data, Biosynthetic gene clusters were predicted using genomes only

***A. anomala* genome hosts a large population of transposable elements (TEs)**

The *A. anomala* genome hosts a large population of TEs that accounts for approximately 88% of the final genome assembly (Table 3). Repeat content remained relatively constant at 88% across major scaffolds (Fig. 1). The TE population consists of 2,536 individual repeat families, making up over 300,000 individual interspersed elements (Table 3). The vast majority (90%) of repetitive sequences was comprised of Long Terminal Repeat (LTR) retrotransposons, mostly *Copia*-like elements, which alone account for over half of the genome assembly. Eight of the ten repeat families with the highest copy numbers (>7,000 members each) were identified as *Copia*-like elements.

***A. anomala* exhibits “two-speed” genome**

The overall distribution of GC-content across major scaffolds remained relatively constant at approximately 34%. However, measurement of proportions of GC-distribution across the entire genome reveals two peaks, indicating a bimodal genome (Fig. 4). The first peak, at 32.9% GC indicates AT-rich regions. This peak accounts for 93% of the genome and 10% (933/9,179) of the protein coding genes. These AT-rich regions are gene poor, with an average gene density of 2.93 genes/Mb. The second peak, at 57.1% GC indicates GC equilibrated regions that account for 7% of the genome and 90% (8,246/9,179) of protein coding genes. These GC-equilibrated regions are over 100-fold more gene-dense with an average of 344 genes/Mb.

We performed an enrichment analysis using the Fisher’s exact test of the gene models within AT-rich genomic regions (Table S6), sheet 1). A number of GO terms are over-represented (p -value < 0.05) including beta-glucan/cellulase metabolism, peptidase/hydrolase activity, and ion transport (Table 4). Additionally, despite these regions encoding only 10% of protein coding genes, 30% (49/165) of predicted effector coding genes were found in these AT-rich hotspots.

***A. anomala* exhibits a number of unique gene families**

We performed an Orthofinder analysis to identify gene families shared with related fungal pathogens (Table S7). A super-gene phylogeny was constructed using 34 single-copy orthologous gene families and their corresponding protein sequences. Gene family counts were used to reconstruct ancestral gene family content and gain/loss of homologous gene families with Wagner parsimony and stochastic mapping (Fig. 5).

There are 1,121 gene models that are not identified as orthologous to related fungal pathogens and are

likely specific to *A. anomala* (Table S6), sheet 2). Of these unique genes, 83 of them are predicted to code for effectors, indicating that over half of the predicted effectors are unique to *A. anomala*. GO terms overrepresented include beta-glucan and cellulose metabolism (p -value < 0.05), suggesting a role in production of plant degrading compounds (Table S8). An additional 450 GO terms are underrepresented, mostly including processes involved in central metabolism and fungal growth and development.

The Orthofinder analysis and Wagner parsimony revealed 354 genes families gained and 721 lost in *A. anomala* since diverging from its last common ancestor with *C. parasitica*. Gene families that are expanded or gained in *A. anomala* account for an additional 32 putative effector genes- meaning that approximately 70% of putative effectors are in species specific gene families or lineages of gene families that have expanded in *A. anomala*. GO terms overrepresented in gained/expanded gene families include catabolic processes and degradation of organic compounds (Table S9). The GO terms that are underrepresented include protein, organelle, and cellular biosynthetic processes.

Transposable elements show evidence of Repeat-induced point mutation (RIP)

The *A. anomala* genome encodes two genes that exhibit sequence homology and are orthologous to *rid* (RIP defective) in *Neurospora crassa*. The two genes are predicted to encode a C5-DNA methyltransferase and a modification methylase respectively. Both genes have been assigned GO terms for methyltransferase activity. *A. anomala* also encodes a homolog of *dim-2*, an additional methyltransferase identified in *N. crassa* to be involved in the RIP process (Figure S3).

Dinucleotide frequencies and RIP indices were calculated for a subset of up to 100 members for all identified repeat families (Fig. 6a). Compared to a control of non-repeat sequences, repeat sequences exhibit an over-abundance of TpA (6.7× more frequent) and TpT (4.1× more frequent) dinucleotides and under-abundance of GpC (3.9× less frequent) and CpG (2.8× less frequent) dinucleotides. RIP indices were also calculated for the same subsets of repeat families (Table 5). The mean TpA/ApT index for repetitive sequences is 1.14, while non-repeat sequences have an index of 0.48. The mean (CpA + TpG)/(ApC + CpT) index is 0.087 in repetitive sequences and 0.095 in non-repetitive sequences. There were no significant differences in dinucleotide frequencies or RIP indices between repeat classes.

An alignment-based RIP analysis of the repeat family with the highest copy number shows that *A. anomala*

Table 3 Detailed breakdown of the repeat population in the *A. anomala* genome. Repeat elements are classified by the name assigned by RepeatClassifier

	Number of Elements	Length (bp)	Percent of Genome
Class I Elements			
LTRs			
Copia-like	104,635	183,550,845	53.587
Gypsy-like	74,787	60,218,611	17.581
Other	68,679	27,882,387	8.14
LINEs			
L3/CR1	59	19,678	0.006
Tad1	4,623	5,848,916	1.708
Other	89	47,022	0.014
SINEs	136	370,136	0.108
Class II Elements			
Tc1-Mariner	47	6,543	0.002
CMC-Enspm	549	129,709	0.038
Helitron	3	3,705	0.001
Unclassified	51,711	28,213,842	8.237
Total Interspersed Repeats	305,318	306,291,394	89.42
Simple Repeats	19,952	937,149	0.27
Low-Complexity	1,402	77,435	0.02

exhibits two dominant kinds of RIP (Fig. 6b). CpA→TpA and CpT→TpT mutations were dominant over other RIP-like mutations. The top 10 repeat families with the highest copy number were also analyzed with the alignment-based RIP analysis and demonstrate the same RIP mutational preference.

A. anomala demonstrates genetic basis for homothallism

Homologs for both MAT1-1 and MAT1-2 idiomorphs have been identified in the *A. anomala* genome within the same 7 kilobase cluster (Table S6, sheet 3), consistent with evidence that the fungus is homothallic [20]. Homologs for the *mat* genes were identified through a BLASTp search of the NCBI nr database and verified by a pairwise sequence comparison to the corresponding genes in *Cryphonectria parasitica* [56]. Like the *C. parasitica* idiomorphs, three protein-coding genes are predicted to constitute MAT1-1 (MAT1-1-1, containing an alpha box motif; MAT1-1-2, a protein of unknown origin; and MAT1-1-3, containing an HMG motif), and a single protein-coding gene is predicted for MAT1-2 (MAT1-2-1, also containing an HMG motif). Within the *A. anomala* MAT locus, the gene encoding MAT1-2-1 was embedded between MAT1-1-1 and MAT1-1-2. Other genes usually associated with mating clusters in fungi, *apn2* and *sla2*, were identified in close proximity to the other MAT protein coding genes. The entire MAT cluster is largely syntenic to that of *Chrysosporthe cubensis*, a closely related homothallic fungus. The MAT loci of *A. anomala* is more compact and contains no additional genes besides those directly involved in determining mating type (Fig. 7). RNAseq data indicate that all four of these MAT genes were expressed constitutively (Figure S4).

Discussion

The final genome assembly of *A. anomala* OR1 is approximately 343 Mb. This assembly is thought to be relatively complete based on genome size estimation compared to flow cytometry data as well as identified BUSCOs. The *A. anomala* genome is very large by fungal standards,

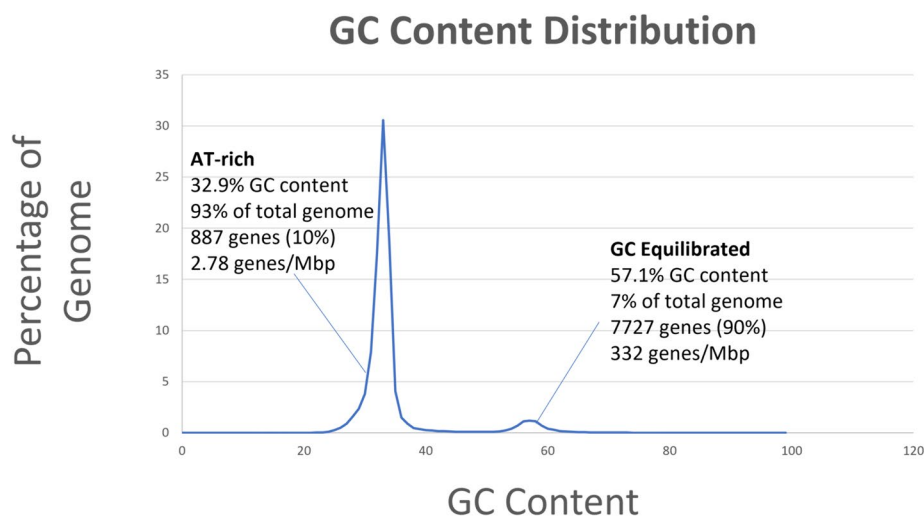


Fig. 4 Distribution of GC-content across the *A. anomala* genome. The genome was broken up into regions using Jensen-Shannon divergence, for which GC-content was calculated. Proportions of the genome were assigned to GC-content in 1 percent increments

Table 4 GO term enrichment analysis of genes within the AT-rich regions of the *A. anomala* genome

GO ID	GO Name	P-value
GO:0051275	beta-glucan catabolic process	0.02887
GO:0051273	beta-glucan metabolic process	0.00894
GO:0032986	protein-DNA complex disassembly	0.01032
GO:0032984	protein-containing complex disassembly	0.04633
GO:0005381	iron ion transmembrane transporter activity	0.02887
GO:0031572	G2 DNA damage checkpoint	0.01032
GO:0030243	cellulose metabolic process	0.02887
GO:0030245	cellulose catabolic process	0.02887
GO:0007095	mitotic G2 DNA damage checkpoint	0.01032
GO:0016825	hydrolase activity	0.02677
GO:0004843	thiol-dependent protease activity	0.02679
GO:0034755	iron ion transmembrane transport	0.02887
GO:0006564	L-serine biosynthetic process	0.01032
GO:0008236	serine-type peptidase activity	0.02677
GO:0008233	peptidase activity	0.02319
GO:0017171	serine hydrolase activity	0.02677
GO:0031498	chromatin disassembly	0.01032
GO:0016702	oxidoreductase activity	0.03590
GO:0045727	positive regulation of translation	0.02887
GO:0072503	cellular divalent inorganic cation homeostasis	0.00894
GO:0072507	divalent inorganic cation homeostasis	0.00894
GO:0015940	pantothenate biosynthetic process	0.04633
GO:0015939	pantothenate metabolic process	0.04633
GO:0034250	positive regulation of cellular amide metabolic process	0.02887
GO:0070011	peptidase activity, acting on L-amino acid peptides	0.01257
GO:0006337	nucleosome disassembly	0.01032
GO:0006826	iron ion transport	0.02887
GO:0006874	cellular calcium ion homeostasis	0.00894
GO:0046915	transition metal ion transmembrane transporter activity	0.03967
GO:0004311	farnesyltransferase activity	0.01032
GO:1,902,653	secondary alcohol biosynthetic process	0.02887
GO:0018958	phenol-containing compound metabolic process	0.02887
GO:0051560	mitochondrial calcium ion homeostasis	0.01032
GO:0043022	ribosome binding	0.01349
GO:0055074	calcium ion homeostasis	0.00894

almost 10 times the ~37 Mb size of the genome of the average ascomycete (Table 2) [32–35, 37–55, 57, 58]. However, large genomes are not uncommon amongst obligate biotrophic pathogens. Powdery mildew fungi, which are ascomycetes, have genomes in excess of 100 Mb [11], and rust fungi, which are basidiomycetes often with complex life cycles, may have genomes approaching 1 Gb [59]. Both of these unrelated fungi are subjected to the strong selective pressure imposed on biotrophic plant pathogens to maintain an intimate interaction with their host while avoiding recognition that initiates an immune response [60, 61]. The outcome of evolution driven by the

pressure of a host/pathogen arms race is parallel adaptations resulting in remarkably similar genomes among biotrophic pathogens [7, 11, 32–39].

The expansion of the *A. anomala* genome is driven by the proliferation of TEs, rather than accumulation of protein coding genes. The TE population is made primarily of LTR retrotransposons. *Copia*-like elements are by far the most abundant, which contrasts related fungi that are dominated by *Gypsy*-like repeats [48]. Despite the massive number of identified LTR retrotransposons, no single element has been determined to be intact with both 5' and 3' LTRs and the

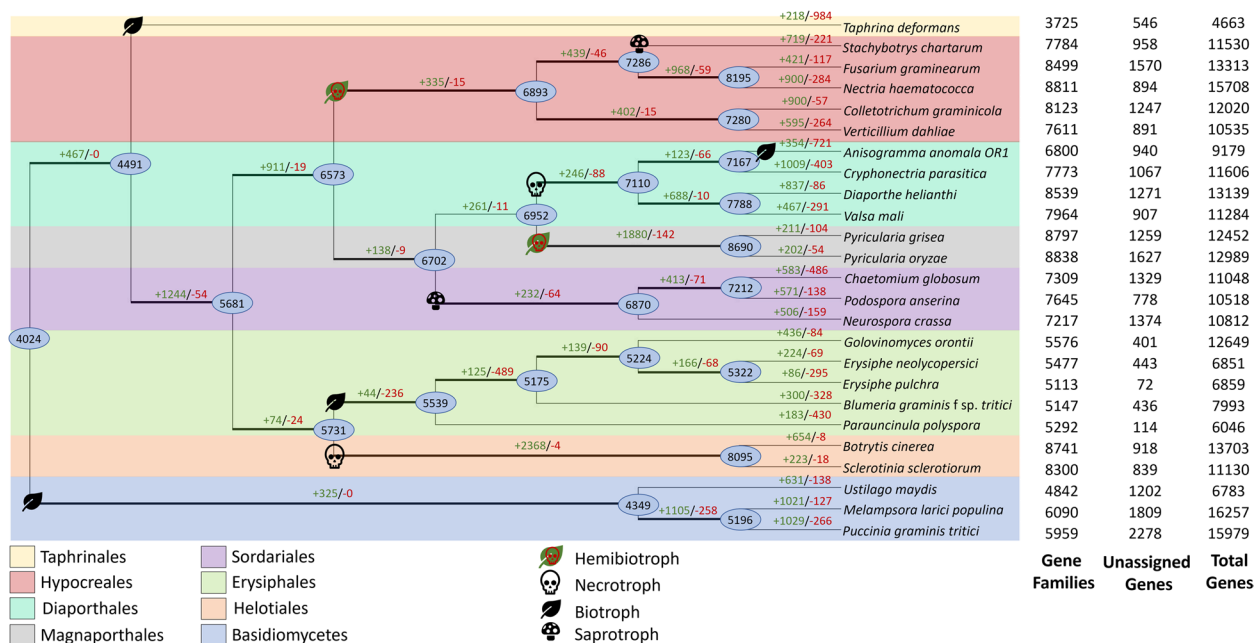


Fig. 5 Predicted pattern of gene family gain and loss in representative fungal genomes. Cladogram representation of Maximum Likelihood phylogeny of *A. anomala* and 15 related fungi based on 2,800 single copy orthologues. The total number of protein families in each species or node is estimated by Wagner parsimony and stochastic mapping. The numbers of the branches correspond to gene family gain (green) or loss (red) and inferred ancestral protein families (in oval). The numbers of gene families, unassigned genes, and total gene numbers are indicated for each species

protein domains required for autonomous transposition, namely reverse transcriptase (RT), RNase H (RH), and integrase (INT) [62]. One of the looming questions regarding the TEs in the *A. anomala* genome is how the invasion and uncontrolled replication of repetitive elements, largely of a single type of TE, was responsible for such extreme genome expansion.

Effector molecules play an important role in the colonization of biotrophic plant pathogens. Plants are able to recognize specific effector molecules through resistance (*R*) genes and activate a powerful hypersensitive response (HR) resulting in plant cell death which halts the spread of the invading pathogen [63]. If recognized, the pathogen is considered avirulent and the effector protein that triggers the HR response is characterized as an avirulence (*avr*) gene [64]. The pathogen responds by mutating or losing *avr* genes, so that they are no longer recognizable, or developing new effectors that avoid or suppress the effector-triggered immune response [65]. This relationship is the basis of the coevolutionary arms race between host plants and pathogens [66]. Most commercially available cultivars of *C. avellana* are protected by the R-gene “Gasaway”, named after the pollinizing cultivar that carried the dominant allele [67]. There is evidence that the Gasaway R-gene protects through an HR response [21]. However, Gasaway protected plants are overcome in

regions of high pathogen pressure and diversity, suggesting that effectors and *avr* genes play a role in the breakdown of resistant cultivars [68–70].

For many putative effectors, there is no known function. As the goal is to be unrecognizable, there is no benefit to maintain conserved effector genes. However, we know that effectors can play multiple roles in establishing and maintaining infection [71, 72]. The large arsenal of putative effectors encoded in the *A. anomala* genome allows for flexibility. This is also why we have observed effectors in repeat-rich regions of genome, where there are high rates of mutation and recombination [73]. The compartmentalization of effectors and genes involved in pathogenicity in repeat rich regions fits the “two-speed” model of evolution [74].

CAZymes play an important role in both necrotrophic and biotrophic phytopathogenic infection. Necrotrophic fungi are known for having an arsenal of plant cell wall busting enzymes to launch an aggressive attack on their host. Necrotrophic ascomycetes code for between 600–800 CAZymes while *A. anomala* codes for 456 putative CAZymes, a typical number for biotrophic pathogens [75]. One of the most notable families of CAZymes encoded in the *A. anomala* genome is the glycoside hydrolase-18 (GH-18) family that includes all identified fungal chitinases [76]. It is predicted that both plant

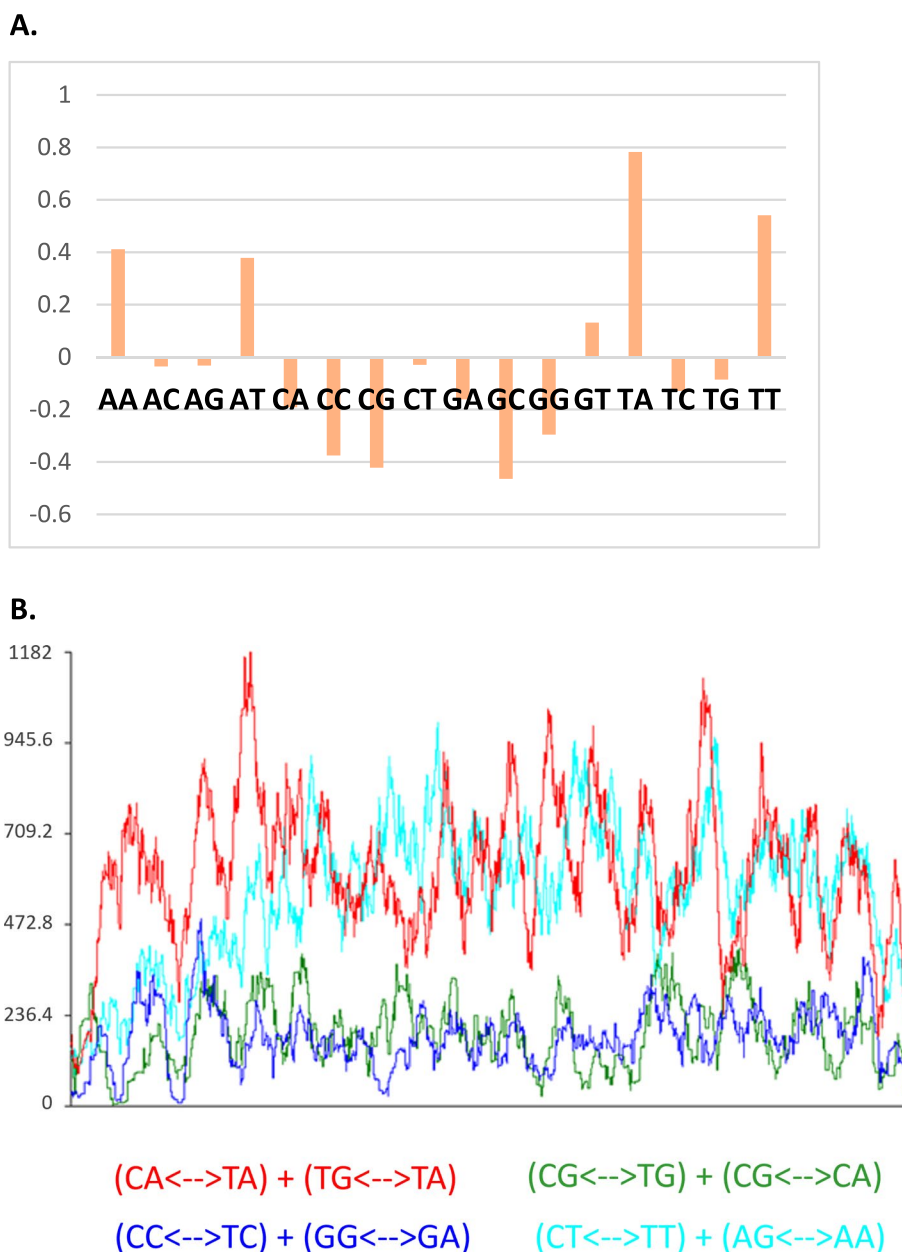


Fig. 6 a Log(10) of average fold change in dinucleotide frequencies of all repeat families compared to non-repetitive control sequences. **b** Alignment based RIP analysis of repeat family with highest copy number (rnd-1_family-0). Each type of RIP mutation is represented by a different color, demonstrating the most dominant types of RIP within this repeat family

and fungal cell wall degrading enzymes are important for establishing biotrophic infection. Histological data of early infection of *A. anomala* on *C. avellana* shows a single germ hypha penetrating the plant cell wall, followed by the formation of intracellular vesicles [21]. CAZymes are required for initial penetration of the plant cell wall as well as the reformation of the fungal cell wall at the fungal/host interface [77]. Reasonable future steps would include investigating the expression of CAZymes during

early infection to elucidate what genes are required to establish infection.

The Wagner parsimony analysis on Orthogroups of related fungal pathogens revealed the loss of 876 and gain of 285 gene families in *A. anomala* since it diverged from *C. parasitica*. Gene reduction in obligate parasites is a common trend, usually due to the loss of specific metabolic pathways as the parasite derives required compounds from their host [78]. KEGG pathway

Table 5 Calculated RIP-indices for the five repeat families with the largest copy number. A TpA/ApT index ≥ 0.89 and (CpT+TpG)/(ApC+GpT) index ≤ 1.03 indicate RIP activity. The numbers presented are the mean values calculated for a subset of 100 repeat family members

Repeat Family	Classification	Copy Number	Mean TpA/ApT	Mean (CpA + TpG)/(ApC + GpT)
rnd1_fam0	LTR-Copia	10,537	1.373688	0.6427065
rnd1_fam1	LTR-unknown	10,431	1.406927	0.644971
rnd1_fam2	LTR-Copia	8,315	1.658198	0.9923082
rnd1_fam3	LTR-Copia	7,814	1.306482	1.016627
rnd1_fam4	LTR-Copia	7,444	1.539685	0.944061

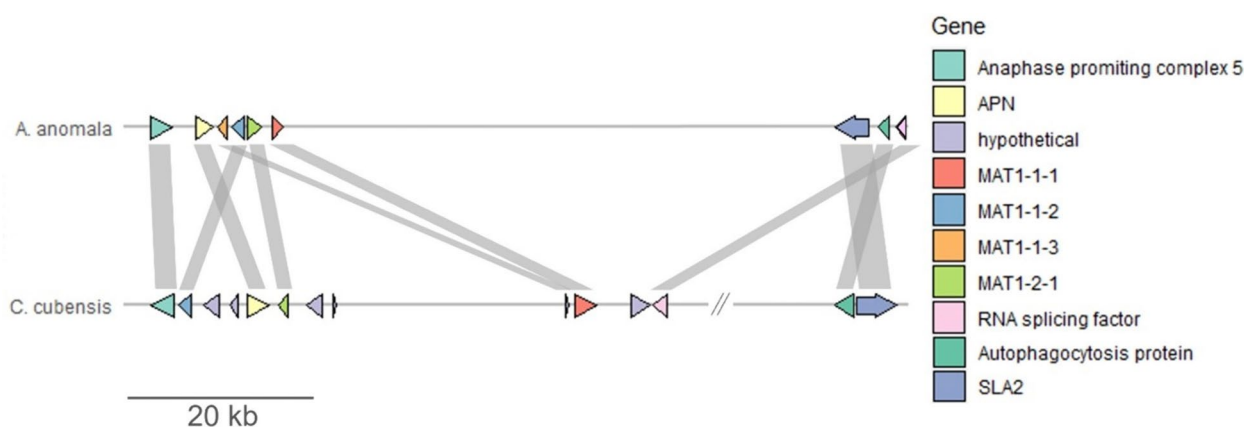


Fig. 7 Genomic region corresponding to mating-type locus in *A. anomala*. Gene models were identified as MAT homologs through BLASTp search of NCBI database and analyzed for synteny compared to *Chrysosporthe cubensis*, a homothallic fungus in the *Cryphonectriaceae*

reconstruction revealed a number of missing or incomplete pathways for the biosynthesis of several amino acids including lysine, tryptophan, and asparagine [79]. It should be noted that the culture medium used for *A. anomala* contains yeast extract as well as additional asparagine to encourage growth [27]. Genes involved in pathways involved in energy generation (NADH dehydrogenase, nitrate reduction/assimilation) are missing as well. *A. anomala* exhibits parallel evolution to unrelated obligate biotrophic fungal pathogens that have independently lost similar biosynthetic and metabolic pathways [60, 80, 81].

The exception to the trend of gene loss is genes or gene families that encode effectors. Gene families that are unique or expanded in *A. anomala* contain 70% of predicted effectors. Biotrophs use effectors to maintain an intimate signaling relationship during infection [82, 83]. The need for a large and diverse effector arsenal drives the evolution of effector diversification and expansion [84, 85] as we observed with *A. anomala*. GO terms overrepresented in unique or expanding families include transmembrane transporters that are involved in

the secretion of secondary metabolites that participate in pathogenesis. Amylases, peptidases, and catabolic activity GO terms are overrepresented in expanded families, likely aiding in adaptation to the obligate biotrophic lifestyle [10].

Despite TEs accounting for 88% of the final genome assembly, very few of these elements contain intact protein domains required for autonomous transposition. Sequences from all identified repeat families show evidence of RIP mutation. RIP is a defense mechanism that protects fungal genomes from TEs expanding unchecked [86, 87]. RIP functions by recognizing stretches over 400 bp of DNA with high (>80%) sequence identity. The DMNT-1 homologue RID (RIP defective) methylates cytosine residues, which then undergo spontaneous deamination into thymine. This induces C→T and G→A transitions in both copies of duplicated sequences, resulting in permanent mutational changes in the DNA sequence [88, 89].

Fungi that demonstrate evidence of RIP vary in the degree or effectiveness by which RIP acts on the genome. *Neurospora crassa*, in which RIP was first described

[90], has a very efficient RIP system, to the point where *N. crassa* has almost no duplicated sequences, TEs nor duplicated genes [91]. In other fungi, RIP is often demonstrable, but less effective [92]. In the related fungus *C. parasitica*, roughly 14% of the 43.9 Mb genome represented TEs, and there was some limited evidence for RIP [48, 92]. The *A. anomala* genome exhibits indications of RIP activity. The RIP indices calculated for repeat families exceed the accepted threshold for RIP activity ($\text{TpA}/\text{ApT} \geq 0.89$ and $(\text{CpT} + \text{ApT})/(\text{ApC} + \text{CpT}) \leq 1.03$) [93, 94] and the dinucleotide frequencies demonstrate a depletion of pre-RIP dinucleotides and an enrichment of post-RIP dinucleotides in repeat regions compared to non-repeat regions [92]. In spite of evidence of a functional RIP pathway, transposons have managed to overtake the *A. anomala* genome. The massive expansion of the TE population in the *A. anomala* genome underscores the observation that mere presence of an apparently functional RIP system is no guarantee that TEs will be held in check.

In addition to defending against the uncontrolled replication of TEs in a genome, RIP is a major driver of genome evolution. RIP induces mutations on duplicated sequences, but those mutations often bleed into neighboring regions, so called “leaky RIP” [86, 95, 96]. Furthermore, the G/C → T/A mutations have a major impact on GC content of a genome. The GC content of the *A. anomala* genome is relatively low, at 34%, however, it is not equally distributed across the genome. GC-proportion distribution reveals two peaks; 93% of the genome landscape has a GC-content of 32%. These stretches of GC-poor containing DNA are broken up by GC-rich blocks that are gene-rich and TE-poor. These data demonstrate that *A. anomala* fits the “two-speed” genome model [96–98].

Analysis of the mating-type locus revealed that *A. anomala* has the genes for both *MATI-1* and *MATI-2* idiomorphs, providing molecular evidence to support the previous evidence for homothallism [20]. Mating type systems, processes, and their associated genes are extraordinarily complicated in fungi, and many genes other than the MAT genes themselves may have different roles in the reproductive process [99]. In addition to controlling sexual development, MAT genes may be important in growth and virulence, including regulation of secondary metabolites and hyphal morphology.

Homothallism, such as that in *A. anomala*, is thought to be an evolutionary destination from which there is no likely return to a progenitor heterothallic state, an idea that was supported through research with *Neurospora* shifting multiple times from heterothallic to homothallic lifestyle, but never the reverse [100]. *Chyrosorthe* which is closely related to *Anisogramma*, has MAT locus features similar to *Neurospora*, including pronounced

influence of retrotransposons, but there was some evidence to suggest that the *MATI-2* and *MATI-1* idiomorphs of the heterothallic *C. austroafricana* evolved from a homothallic progenitor [101]. In both the case of *Neurospora* and *Chyrosorthe*, the evolutionary transition of mating type is facilitated by TEs within the *mat* locus. Like the rest of the *A. anomala* genome, the *mat* locus is flanked by TEs. But the core genes for each idiomorph are found within the same 7 kb block with no TEs or additional genes. The mating cluster of *C. cubensis* includes additional genes not found in *A. anomala* as well as a 200 kb insertion of DNA that contains over 60 genes not related to determination of mating type (Fig. 7). One of the roles of sex in fungi and other organisms is to bring genetic variation to the species. It seems that a combination of homothallic sex and rampant genome invasion and expansion by transposons brings sufficient variability to *A. anomala*.

Conclusions

At nearly 350 Mb, the *A. anomala* genome represents the largest ascomycete genome yet characterized. Gene number and putative functions are typical of fungal plant pathogens, but runaway amplification of repeat sequences has led to a massively bloated genome, despite hallmarks of functional genome surveillance by RIP. The *A. anomala* genome characterization will serve as a resource for others investigating this economically important plant pathogen, and for those interested in fungal genome evolution.

Methods

Fungal strain

A. anomala is an obligate biotroph that has not been grown in continuous culture, so tissue is scarce and not clonal. Based on knowledge of EFB epidemiology [102, 103], the *A. anomala* population in Oregon is believed to be decedents of a single introduction event from east of the Rocky Mountains and belong to a single lineage. But mycelium from different trees in fields is not clonal, and DNA or RNA extracted from a collection of germinated ascospores is also not clonal. The closest approximation we have to homogeneous tissue is to harvest ascospores from a single canker on a single tree, with the understanding that it most likely represents the result of a single infection. We collected ascospores from individual cankers from infected branches harvested from hazelnut plants growing at the Oregon State University Smith Horticultural Research Farm, Corvallis, OR. These plants had been inoculated 18 months prior in the greenhouse using local diseased plant material as inoculum source. We designate the strain presented here Oregon1, OR1.

Ascospores were extracted following the protocol we used previously [28]. Briefly, the branches were cut into pieces 5–7 cm in length, and surface-sterilized for 3 min in 10% bleach (0.525% sodium hypochlorite) followed by 1 min in 70% ethanol. After rinsing with sterile H₂O, the stromata were hydrated in sterile H₂O for 30 min and air-dried. The top of a canker was cut off with a sterile razor blade to expose the necks of the perithecia, and another sterile razor blade was inserted under the perithecia to provide pressure from below and push ascospores out of perithecial neck. The spores from individual cankers were suspended in sterile H₂O containing 10 ppm rifampicin and 100 ppm streptomycin and quantified with a hemocytometer. We found one canker that produced approximately 5.5 M ascospores and these spores were used in this study unless noted otherwise.

To generate primary mycelium, a portion of the ascospores was adjusted to 1×10^5 spores per ml and used to inoculate plates of culture medium overlaid with cellophane. The rest of the ascospores were stored at -80 °C. Half a milliliter of the spore suspension was spread on the cellophane surface in individual 9-cm diameter petri dishes. The medium contained (per liter) 2.7 g modified Murashige and Skoog basal salt mixture; 20 g sucrose; 2 g yeast extract; 2 g L-Asparagine; 15 g Bacto agar; 0.25 g activated charcoal; and 10 mg Rifampicin [27]. The cultures were grown at 18 °C in the dark for 8 weeks, by which time many spores had germinated and grown into opaque, whitish colonies approximately 0.25–0.5 mm in diameter. Mycelium was harvested by rinsing the cellophane with sterile H₂O. A subset of plates was kept for four more weeks. By then the small colonies were turning grey and black, and the senescent mycelium was harvested as described above.

Nucleic acid extraction, genome sequencing and assembly

Mycelium from 8-week-old cultures were used for DNA extraction using Genra Puregene kit (Qiagen) following the fungi protocol. One paired-end DNA library with insert size approximately 350 bp (excluding adapters) was constructed using the TruSeq DNA Sample Prep kit (Illumina). Three mate-pair DNA libraries with insert sizes approximately 3 kb, 6 kb, and 10 kb, respectively, were constructed using the Nextera Mate Pair Library Prep kit (Illumina) following manufacturer's instructions. All libraries were sequenced on the Illumina MiSeq platform.

The paired-end reads were trimmed with Trimmomatic v0.32 [104] in paired-end mode to remove adapter sequences and reads shorter than 100 bp after trimming were dropped. The mate-pair reads were first trimmed with Trimmomatic in paired-end mode to remove external adapters, then trimmed with Trimmomatic in

single-end mode to remove internal adapters at ligation junctions. Reads shorter than 35 bp after trimming were dropped. The resulting reads were processed with a custom Perl script and only read pairs meeting the following conditions were retained for genome assembly: 1) both reads must have survived adapter trimming; 2) for read pairs in which external adapters were found, the junction adapter must be found in both reads; 3) for read pairs in which external adapters were not found, junction adapter must be found in at least one read. After data processing, the sequence reads were assembled using AllPaths-LG release 52,155 with default settings. [105]. Assembled scaffolds were subjected to a BLASTn search of the GenBank database release 258 [106]. Any scaffolds where the top hit was not fungal were removed as contamination.

Flow cytometry

One hundred micrograms of freshly harvested 8-week old mycelium were cut into fine pieces with a sterile razor blade in 500 µl LB01 buffer on ice to release the nuclei [107]. The mixture was passed through a 40 µm filter and washed with 200 µl LB01 buffer. Nuclei from 50 mg young radish leaf, which has a 2C genome size of 1.1 Gb, were released the same way and used as control. Nuclei solutions were treated with RNase A and stained with propidium Iodide at room temperature for 20 min in darkness and run through a Beckman Cytoflex flow cytometer. The experiment was repeated three times.

Repeat identification and masking

The assembled genome was soft-masked prior to gene prediction [108]. A comprehensive, non-redundant repeat library was created by integrating output from RepeatModeler [109, 110], TransposonPSI [111], and LTRharvest [112]. RepeatModeler v1.0.11 and TransposonPSI were run using default parameters to generate the first two repeat libraries. The third repeat library was built using LTRharvest. False positives were removed from the LTRharvest library by running LTRdigest with protein HMMs from Pfam [113] and GyDB [114] databases. LTR retrotransposons without domain hits were removed from the LTRharvest repeat library.

Each of the three repeat libraries was classified using RepeatClassifier, part of the RepeatModeler program suite, with Replibase version 23.08 [115], for consistency in identification and naming of repeat elements. The three repeat libraries were then merged and clustered with CD-HIT [116] at $\geq 80\%$ identity to create a non-redundant library [117]. This custom library was used to soft-mask the *A. anomala* genome using RepeatMasker with the "xsmall" argument and default parameters [118].

Transcriptome sequencing, gene prediction and annotation

Ascospores, 8-week old mycelium and 12-week old senescent mycelium were used for RNA extraction using the Plant RNeasy kit (Qiagen) following manufacturer's instructions. Three mRNA libraries, one for each sample, were prepared using the TruSeq RNA Preparation kit (Illumina) following manufacturer's instructions. The libraries were sequenced on the Illumina MiSeq platform.

Gene models were predicted using the BRAKER2 annotation pipeline [119], incorporating GeneMark-ET [120, 121] and Augustus [122] for ab initio and evidence-based gene prediction. RNA-Seq reads were mapped to the genome assembly using STAR [123]. The RNA-Seq mapping results were used as evidence for gene prediction in the BRAKER2 pipeline, using the “fungus” argument for fungal gene prediction. Genome completeness was assessed through a BUSCO analysis of benchmarking eukaryotic and fungal single-copy orthologs [124].

Blast2GO v5.2.5 [125] was used to perform a BLASTp search of the NCBI nr database with E-value cutoff of 1e-3. Interproscan v5.53 [126] results were imported in to Blast2GO and merged with GO annotations. KEGG annotation terms [79, 127] were assigned using a combination of BlastKOALA v2.2 [128] and the KEGG Automatic Annotation Server (KAAS) [129] searched against eukaryote and prokaryote KEGG GENES databases (release v89.1), with the single-directional best hit method.

Secreted proteins were predicted using SignalP 5.0 [130] to identify signal peptides sequences. Predicted secreted proteins were then analyzed with EffectorP 2.0 [131, 132] to predict genes encoding for potential effectors. Evidence including protein size and cysteine content was used for effector prediction. Potential function of effectors was evaluated by a BLASTp search [133] of the GenBank nr database (release 239) [106] with an e-value cutoff of 0.001. Functional domains were assigned using CD-Search webserver with default settings against the Conserved Domain Database v3.20 [134, 135]. Carbohydrate active enzymes were predicted using the dbCAN3 meta server [136–138] which integrates HMMER [139], DIAMOND [140], and Hotpep [141] searches of the CAZy database [142]. Biosynthetic gene clusters were predicted and identified using antiSMASH v5.1.2 [143].

GC-content distribution

Analysis of GC-content was performed by segmenting genomic sequences into regions of differing GC-content using the Jensen-Shannon divergence at each sequence position calculated using OcculterCut v1.1 [144]. Gene models associated with AT-rich genomic regions were used as a test set in a GO term enrichment analysis test

using a two-tailed Fisher's Exact Test with a filter value of 0.05 with BLAST2GO v5.2.5 [125].

Fungal super-gene phylogeny

We collected proteomes from 24 related ascomycete species to identify orthologous gene families. OrthoFinder v2.2.6 [145] was used under default settings to build orthogroups. Thirty-four single-copy orthologous gene families and their corresponding protein sequences were retrieved and aligned with MUSCLE v3.8.31 [146] and alignments were trimmed with TrimAl v1.4 [147] using the automated feature to select the best method. The trimmed alignments were automatically concatenated and partitioned using IQ-TREE v1.7-beta17 [148, 149]. The maximum likelihood tree was reconstructed with IQ-TREE under the LG+I+G model as selected using ModelFinder [150].

Gene family counts from the Orthofinder analysis were used to reconstruct ancestral gene family content and gain/loss of homologous gene families. These traits were reconstructed using Wagner parsimony in the Count software package [151] as well as stochastic mapping with GLOOME [152].

RIP analysis

RIP indices of individual repeat copies were calculated in RStudio v1.1.414 [153] using a custom R (v4.1.2) script (file S1) and the Biostrings package v2.62.0 [154]. RIP-CAL v2 [155] was used for alignment based analysis of repeat families and calculations of mutation frequencies. Large RIP affected regions (LRARs) were identified by a minimum of seven consecutive sliding windows (window size=1000 bp, slide size=500 bp) with a minimum RIP product value of 1.1, maximum RIP substrate value of 0.75 and minimum composite (product – substrate) value of 0.01. RIP product, substrate, and composite values and LRAR analysis was performed using The RIPper [156].

Abbreviations

BUSCO	Benchmarking single-copy orthologs
CAZymes	Carbohydrate-active enzymes
EFB	Eastern Filbert Blight
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
LRAR	Large RIP-affected region
RIP	Repeat-induced point mutation
TEs	Transposable elements

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10198-1>.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.
Supplementary Material 9.
Supplementary Material 10.
Supplementary Material 11.
Supplementary Material 12.
Supplementary Material 13.
Supplementary Material 14.

Acknowledgements

We thank Dr. Shawn Mehlenbacher, Oregon State University, for his generous gift of the infected hazelnut stems from which the fungal spores and DNA were isolated.

Authors' contributions

B.I.H., T.J.M., and G.C. conceived of the project and designed research; A.B.C. and G.C. performed the wet lab experiments; A.B.C., D.C.P. and G.C. assembled and annotated the genome; A.B.C. and N.Z. performed the evolutionary analyses; A.B.C., G.C., and B.I.H. wrote the paper with contribution from all authors. All author(s) read and approved the final manuscript.

Funding

Funding from the following is very gratefully acknowledged: the USDA-NIFA through the Specialty Crop Research Initiative Competitive Grants Program to TJM and BIH (Grant #2016–51181–25412); the Rutgers Microbial Biology Graduate Program for an initial PhD Research Fellowship to ABC; The New Jersey Agricultural Experiment Station for partial salary support to GC, DCP, TJM, NZ and BIH.

Availability of data and materials

The *Anisogramma anomala* OR1 genome sequence and assembly have been submitted to NCBI SRA and have been given the BioProject reference number PRJNA966177. Protein and coding sequence fasta files can be found on Fig-Share under identifiers 24,905,166.v1 and 24,898,656.v1 respectively.

Declarations

Ethics approval and consent to participate

Not applicable.

Infected hazelnut stems for laboratory use were imported under USDA Permit # P526P-07–03455 to BIH. The authors comply with relevant institutional guidelines for plant studies.

This paper does not involve human or animal experiments. This paper does not include any personal information. The dataset of this study is being submitted to GenBank and should be available in time for the review process.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Plant Biology, Rutgers The State University of New Jersey, New Brunswick, NJ 08901, USA. ²Graduate Program in Microbial Biology, Rutgers The State University of New Jersey, New Brunswick, NJ 08901, USA. ³Crop Production and Pest Control Research Unit, USDA-ARS, West Lafayette, IN 47907, USA. ⁴Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA. ⁵Department of Entomology, Rutgers The State University of New Jersey, New Brunswick, NJ 08901, USA. ⁶Center for Vector

Biology, Rutgers The State University of New Jersey, New Brunswick, NJ 08901, USA. ⁷Department of Biochemistry and Microbiology, Rutgers The State University of New Jersey, New Brunswick, NJ 08901, USA.

Received: 12 May 2023 Accepted: 7 March 2024

Published online: 05 April 2024

References

- Giovannetti M, Avio L, Sbrana C. Fungal spore germination and pre-symbiotic mycelial growth—physiological and genetic aspects. In: Koltai H, Kapulnik Y, editors. Arbuscular mycorrhizas: physiology and function. Dordrecht: Springer; 2010. p. 3–32.
- Chanclud E, Morel JB. Plant hormones: a fungal point of view. *Mol Plant Pathol.* 2016;17(8):1289–97.
- Lorrain C, Goncalves Dos Santos KC, Germain H, Hecker A, Duplessis S. Advances in understanding obligate biotrophy in rust fungi. *New Phytol.* 2019;222(3):1190–206.
- Lorrain DA. The powdery mildews: a review of the world's most familiar (yet poorly known) plant pathogens. *Annu Rev Phytopathol.* 2008;46:27–51.
- Bélanger RR, Bushnell WR, Dik AJ, Carver TL. The powdery mildews: a comprehensive treatise. St. Paul: American Phytopathological Society (APS Press); 2002.
- Spanu P, Kamper J. Genomics of biotrophy in fungi and oomycetes—emerging patterns. *Curr Opin Plant Biol.* 2010;13(4):409–14.
- Spanu PD. The genomics of obligate (and nonobligate) biotrophs. *Annu Rev Phytopathol.* 2012;50:91–109.
- Kemen E, Jones JD. Obligate biotroph parasitism: can we link genomes to lifestyles? *Trends Plant Sci.* 2012;17(8):448–57.
- Tang C, Xu Q, Zhao M, Wang X, Kang Z. Understanding the lifestyles and pathogenicity mechanisms of obligate biotrophic fungi in wheat: The emerging genomics era. *The Crop Journal.* 2018;6(1):60–7.
- Liang P, Liu S, Xu F, Jiang S, Yan J, He Q, et al. Powdery mildews are characterized by contracted carbohydrate metabolism and diverse effectors to adapt to obligate biotrophic lifestyle. *Front Microbiol.* 2018;9:3160.
- Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science.* 2010;330(6010):1543–6.
- Grandaubert J, Lowe RG, Soyer JL, Schoch CL, Van de Wouwe AP, Fudal I, et al. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans*-*Leptosphaeria biglobosa* species complex of fungal pathogens. *BMC Genomics.* 2014;15(1):1–27.
- Oliver KR, Greene WK. Transposable elements: powerful facilitators of evolution. *BioEssays.* 2009;31(7):703–14.
- Duplessis S, Bakkeren G, Hamelin R. Advancing knowledge on biology of rust fungi through genomics. *Adv Bot Res.* 2014;70:173–209.
- Bindschedler LV, Panstruga R, Spanu PD. Mildew-omics: how global analyses aid the understanding of life and evolution of powdery mildews. *Front Plant Sci.* 2016;7:123.
- Fuller A. The filbert or hazelnut. The Nut Culturist, Orange Judd Company, NY. 1908:118–46.
- Weschcke C. Hazels and filberts. Growing nuts in the north Webb, St Paul, MN. 1954:24–38.
- Farr DF, Bills GF, Chamuris GP, Rossman AY. Fungi on plants and plant products in the United States. St. Paul: APS Press; 1989.
- Pinkerton J, Johnson K, Mehlenbacher S, Pscheidt J. Susceptibility of European hazelnut clones to eastern filbert blight. *Plant Dis.* 1993;77(3):261–6.
- Gottwald T, Cameron H. Studies in the morphology and life history of *Anisogramma anomala*. *Mycologia.* 1979;71(6):1107–26.
- Pinkerton J, Stone J, Nelson S, Johnson K. Infection of European hazelnut by *Anisogramma anomala*: Ascospore adhesion, mode of penetration of immature shoots, and host response. *Phytopathology.* 1995;85(10):1260–8.
- Thompson M, HB L, SA M. Hazelnuts. Fruits Breeding (Edited by Jules Janick and James N. Moore). Volume III Chapter 3. 1996;184:125.
- Pinkerton J, Johnson K, Theiling K, Griesbach J. Distribution and characteristics of the eastern filbert blight epidemic in western Oregon. *Plant Dis.* 1992;76(11):1179–82.

24. Davison A, Davidson R. Apioporthes and Monochaetia cankers reported in western Washington. *Plant Disease Reporter*. 1973.
25. Julian J, Seavert C, Olsen J, editors. An economic evaluation of the impact of Eastern Filbert Blight resistant hazelnut cultivars in Oregon, Usa. VII International Congress on Hazelnut. 2008;845.
26. Snelling J, Mehlenbacher S, Heilsnis B, Mooneyham R, editors. Breeding hazelnuts resistant to eastern filbert blight. XXXI International Horticultural Congress (IHC2022): International Symposium on Breeding and Effective Use of Biotechnology and 1362;2022.
27. Stone JK, Pinkerton J, Johnson K. Axenic culture of *Anisogramma anomala*: Evidence for self-inhibition of ascospore germination and colony growth. *Mycologia*. 1994;86(5):674–83.
28. Cai G, Leadbetter CW, Muehlbauer MF, Molnar TJ, Hillman BI. Genome-wide microsatellite identification in the fungus *Anisogramma anomala* using Illumina sequencing and genome assembly. *PLoS ONE*. 2013;8(11): e82408.
29. Gottwald TR, Cameron HR. Infection site, infection period, and latent period of canker caused by *Anisogramma anomalum* in European Filbert. *Phytopathology*. 1980;70(11):1083–7.
30. Pinkerton J, Johnson K, Stone J, Ivors K. Maturation and seasonal discharge pattern of ascospores of *Anisogramma anomala*. *Phytopathology*. 1998;88(11):1165–73.
31. Gottwald TR. Infection Site, Infection Period, and Latent Period of Canker Caused by *Anisogramma anomalum* in European Filbert. *Phytopathology*. 1980;70(11):1083–7.
32. Wicker T, Oberhaensli S, Parlange F, Buchmann JP, Shatalina M, Roffler S, et al. The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. *Nat Genet*. 2013;45(9):1092–6.
33. Wu Y, Ma X, Pan Z, Kale SD, Song Y, King H, et al. Comparative genome analyses reveal sequence features reflecting distinct modes of host-adaptation between dicot and monocot powdery mildew. *BMC Genomics*. 2018;19:1–20.
34. Wadl PA, Mack BM, Beltz SB, Moore GG, Baird RE, Rinehart TA, et al. Development of genomic resources for the powdery mildew. *Erysiphe pulchra* Plant Dis. 2019;103(5):804–7.
35. Micali C, Göllner K, Humphry M, Consonni C, Panstruga R. The powdery mildew disease of *Arabidopsis*: a paradigm for the interaction between plants and biotrophic fungi. *Arabidopsis Book*. 2008;6:e0115.
36. Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci U S A*. 2011;108(22):9166–71.
37. Frantzeskakis L, Németh MZ, Barsoum M, Kusch S, Kiss L, Takamatsu S, et al. The *Parauncinula polyspora* draft genome provides insights into patterns of gene erosion and genome expansion in powdery mildew fungi. *MBio*. 2019;10(5): 01692–19. <https://doi.org/10.1128/mbio>.
38. Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature*. 2006;444(7115):97–101.
39. Cissé OH, Almeida JM, Fonseca A, Kumar AA, Salojärvi J, Overmyer K, et al. Genome sequencing of the plant pathogen *Taphrina deformans*, the causal agent of peach leaf curl. *MBio*. 2013;4(3):00055–13. <https://doi.org/10.1128/mbio>.
40. Cuomo CA, Gueldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, et al. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science*. 2007;317(5843):1400–2.
41. King R, Urban M, Hammond-Kosack MC, Hassani-Pak K, Hammond-Kosack KE. The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics*. 2015;16(1):544.
42. O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, et al. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat Genet*. 2012;44(9):1060–5.
43. Gómez Luciano LB, Tsai JJ, Chuma I, Tosa Y, Chen Y-H, Li J-Y, et al. Blast fungal genomes show frequent chromosomal changes, gene gains and losses, and effector gene turnover. *Mol Biol Evol*. 2019;36(6):1148–61.
44. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND. The genome sequence of the rice blast fungus *Magnaporthe oryzae*. *Nature*. 2005;434(7036):980–6.
45. Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BP, et al. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Pathog*. 2011;7(7): e1002137.
46. Van Kan JA, Stassen JH, Mosbach A, Van Der Lee TA, Faino L, Farmer AD, et al. A gapless genome sequence of the fungus *Botrytis cinerea*. *Mol Plant Pathol*. 2017;18(1):75–89.
47. Amselem J, Cuomo CA, van Kan JA, Viaud M, Benito EP, Couloux A, et al. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet*. 2011;7(8): e1002230.
48. Crouch JA, Dawe A, Aerts A, Barry K, Churchill AC, Grimwood J, et al. Genome sequence of the chestnut blight fungus *Cryphonectria parasitica* EP155: a fundamental resource for an archetypical invasive plant pathogen. *Phytopathology*. 2020;110(6):1180–8.
49. Baroncelli R, Scala F, Vergara M, Thon MR, Ruocco M. Draft whole-genome sequence of the *Diaporthe helianthi* 7/96 strain, causal agent of sunflower stem canker. *Genomics Data*. 2016;10:151–2.
50. Derbyshire M, Denton-Giles M, Hegedus D, Seifbarghy S, Rollins J, van Kan J, et al. The complete genome sequence of the phytopathogenic fungus *Sclerotinia sclerotiorum* reveals insights into the genome architecture of broad host range pathogens. *Genome Biol Evol*. 2017;9(3):593–618.
51. Yin Z, Liu H, Li Z, Ke X, Dou D, Gao X, et al. Genome sequence of *Valsa* canker pathogens uncovers a potential adaptation of colonization of woody bark. *New Phytol*. 2015;208(4):1202–16.
52. Coleman JJ, Rounsley SD, Rodríguez-Carres M, Kuo A, Wasmann CC, Grimwood J, et al. The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet*. 2009;5(8): e1000618.
53. Semeiks J, Borek D, Otwinowski Z, Grishin NV. Comparative genome sequencing reveals chemotype-specific gene clusters in the toxigenic black mold *Stachybotrys*. *BMC Genomics*. 2014;15(1):1–16.
54. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, et al. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*. 2003;422(6934):859–68.
55. Cuomo CA, Untereiner WA, Ma L-J, Grabherr M, Birren BW. Draft genome sequence of the cellulolytic fungus *Chaetomium globosum*. *Genome Announc*. 2015;3(1):e00021–e115.
56. McGuire IC, Marra RE, Turgeon BG, Milgroom MG. Analysis of mating-type genes in the chestnut blight fungus. *Cryphonectria parasitica* *Fungal Genet Biol*. 2001;34(2):131–44.
57. Mohanta TK, Bae H. The diversity of fungal genome. *Biol Proced Online*. 2015;17:8.
58. Espagne E, Lespinet O, Malagnac F, Da Silva C, Jaillon O, Porcel BM, et al. The genome sequence of the model ascomycete fungus *Podospora anserina*. *Genome Biol*. 2008;9(5):R77.
59. Tavares S, Ramos AP, Pires AS, Azinheira HG, Caldeirinha P, Link T, et al. Genome size analyses of Pucciniales reveal the largest fungal genomes. *Front Plant Sci*. 2014;5:422.
60. Kemen AC, Agler MT, Kemen E. Host–microbe and microbe–microbe interactions in the evolution of obligate plant parasitism. *New Phytol*. 2015;206(4):1207–28.
61. Gómez-Pérez D, Kemen E. Predicting lifestyle from positive selection data and genome properties in oomycetes. *Pathogens*. 2021;10(7):807.
62. Muszewska A, Hoffman-Sommer M, Grynberg M. LTR retrotransposons in fungi. *PLoS ONE*. 2011;6(12): e29425.
63. Wu L, Chen H, Curtis C, Fu ZQ. Go in for the kill: How plants deploy effector-triggered immunity to combat pathogens. *Virulence*. 2014;5(7):710–21.
64. Jaswal R, Kiran K, Rajarammohan S, Dubey H, Singh PK, Sharma Y, et al. Effector biology of biotrophic plant fungal pathogens: Current advances and future prospects. *Microbiol Res*. 2020;241: 126567.
65. De Wit PJ. Pathogen avirulence and plant resistance: a key role for recognition. *Trends Plant Sci*. 1997;2(12):452–8.
66. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal effectors and plant susceptibility. *Annu Rev Plant Biol*. 2015;66:513–45.
67. Mehlenbacher SA, Thompson MM, Cameron HR. Occurrence and Inheritance of Resistance to Eastern Filbert Blight in Gasaway Hazelnut. *HortScience*. 1991;26(4):410–1.

68. Molnar TJ, Goffreda JC, Funk CR. Survey of *Corylus* Resistance to Anisogramma anomala from Different Geographic Locations. *HortScience*. 2010;45(5):832–6.
69. Molnar TJ, Capik J, Zhao S, Zhang N. First Report of Eastern Filbert Blight on *Corylus avellana* “Gasaway” and “VR20-11” Caused by Anisogramma anomala in New Jersey. *Plant Dis*. 2010;94(10):1265.
70. Sathuvalli VR, Mehlenbacher SA, Smith DC. Response of Hazelnut Accessions to Greenhouse Inoculation with Anisogramma anomala. *HortScience*. 2010;45(7):1116–9.
71. Toruño TY, Stergiopoulos I, Coaker G. Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. *Annu Rev Phytopathol*. 2016;54:419–41.
72. Sharpee WC, Dean RA. Form and function of fungal and oomycete effectors. *Fungal Biol Rev*. 2016;30(2):62–73.
73. Plissonneau C, Benevenuto J, Mohd-Assaad N, Fouché S, Hartmann FE, Croll D. Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. *Front Plant Sci*. 2017;8:119.
74. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol*. 2012;10(6):417–30.
75. Zhao Z, Liu H, Wang C, Xu J-R. Erratum to: comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*. 2014;15(1):1–15.
76. Gruber S, Seidl-Seiboth V. Self versus non-self: fungal cell wall degradation in Trichoderma. *Microbiology*. 2012;158(1):26–34.
77. Lyu X, Shen C, Fu Y, Xie J, Jiang D, Li G, et al. Comparative genomic and transcriptional analyses of the carbohydrate-active enzymes and secretomes of phytopathogenic fungi reveal their significant roles during infection and development. *Sci Rep*. 2015;5(1):15565.
78. Scott K. Obligate parasitism by phytopathogenic fungi. *Biol Rev*. 1972;47(4):537–72.
79. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 1999;27(1):29–34.
80. Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, et al. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biol*. 2011;9(7): e1001094.
81. McDowell JM. Genomes of obligate plant pathogens reveal adaptations for obligate parasitism. *Proc Natl Acad Sci*. 2011;108(22):8921–2.
82. Pendleton AL, Smith KE, Feau N, Martin FM, Grigoriev IV, Hamelin R, et al. Duplications and losses in gene families of rust pathogens highlight putative effectors. *Front Plant Sci*. 2014;5:299.
83. Bourras S, Praz CR, Spanu PD, Keller B. Cereal powdery mildew effectors: a complex toolbox for an obligate pathogen. *Curr Opin Microbiol*. 2018;46:26–33.
84. Kamoun S. Groovy times: filamentous pathogen effectors revealed. *Curr Opin Plant Biol*. 2007;10(4):358–65.
85. De Jonge R, Bolton MD, Thomma BP. How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Curr Opin Plant Biol*. 2011;14(4):400–6.
86. Gladyshev E. Repeat-Induced Point Mutation and Other Genome Defense Mechanisms in Fungi. *Microbiol Spectr*. 2017;5(4):687–99.
87. Galagan JE, Selker EU. RIP: the evolutionary cost of genome defense. *Trends Genet*. 2004;20(9):417–23.
88. Selker EU, Garrett PW. DNA sequence duplications trigger gene inactivation in *Neurospora crassa*. *Proc Natl Acad Sci*. 1988;85(18):6870–4.
89. Cambareri EB, Jensen BC, Schabtach E, Selker EU. Repeat-induced G-C to A-T mutations in *Neurospora*. *Science*. 1989;244(4912):1571–5.
90. Singer MJ, Marcotte BA, Selker EU. DNA methylation associated with repeat-induced point mutation in *Neurospora crassa*. *Mol Cell Biol*. 1995;15(10):5586–97.
91. Wang L, Sun Y, Sun X, Yu L, Xue L, He Z, et al. Repeat-induced point mutation in *Neurospora crassa* causes the highest known mutation rate and mutational burden of any cellular life. *Genome Biol*. 2020;21:1–23.
92. Clutterbuck AJ. Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. *Fungal Genet Biol*. 2011;48(3):306–26.
93. Selker EU, Tountas NA, Cross SH, Margolin BS, Murphy JG, Bird AP, et al. The methylated component of the *Neurospora crassa* genome. *Nature*. 2003;422(6934):893–7.
94. Margolin BS, Garrett-Engle PW, Stevens JN, Fritz DY, Garrett-Engle C, Metznerberg RL, et al. A methylated *Neurospora* 5S rRNA pseudogene contains a transposable element inactivated by repeat-induced point mutation. *Genetics*. 1998;149(4):1787–97.
95. Selker EU, Stevens JN. DNA methylation at asymmetric sites is associated with numerous transition mutations. *Proc Natl Acad Sci*. 1985;82(23):8114–8.
96. Frantzeskakis L, Kusch S, Panstruga R. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Mol Plant Pathol*. 2019;20(1):3–7.
97. Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GC, Wittenberg AH, et al. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res*. 2016;26(8):1091–100.
98. Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev*. 2015;35:57–65.
99. Lee SC, Corradi N, Doan S, Dietrich FS, Keeling PJ, Heitman J. Evolution of the sex-related locus and genomic features shared in microsporidia and fungi. *PLoS ONE*. 2010;5(5): e10539.
100. Gioti A, Mushegian AA, Strandberg R, Stajich JE, Johannesson H. Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. *Mol Biol Evol*. 2012;29(10):3215–26.
101. Kanzi AM, Steenkamp ET, Van der Merwe NA, Wingfield BD. The mating system of the Eucalyptus canker pathogen *Chrysosporium austroafricanum* and closely related species. *Fungal Genet Biol*. 2019;123:41–52.
102. Muehlbauer MF, Tobia J, Honig JA, Zhang N, Hillman BI, Gold KM, et al. Population differentiation within *Anisogramma anomala* in North America. *Phytopathology*. 2019;109(6):1074–82.
103. Tobia J, Muehlbauer MF, Honig JA, Pscheidt JW, Capik JM, Molnar TJ. Genetic Diversity Analysis of *Anisogramma anomala* in the Pacific Northwest and New Jersey. Manuscript submitted for publication. 2022.
104. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
105. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A*. 2011;108(4):1513–8.
106. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the National Center for Biotechnology Information in 2023. *Nucleic Acids Res*. 2023;51(D1):D29–38.
107. Loureiro J, Rodriguez E, DOLEŽEL J, Santos C. Comparison of four nuclear isolation buffers for plant DNA flow cytometry. *Annals of Botany*. 2006;98(3):679–89.
108. Haridas S, Salamov A, Grigoriev IV. Fungal genome annotation. *Fungal Genomics*: Springer; 2018. p. 171–84.
109. Smit A, Hubley R. RepeatModeler Open-10.2008–2015: <http://www.repeatmasker.org>.
110. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci*. 2020;117(17):9451–7.
111. Haas B. TransposonPSI: an application of PSI-Blast to mine (retro-) transposon ORF homologies. Broad Institute, Cambridge, MA, USA. 2007.
112. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*. 2008;9(1):18.
113. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47(D1):D427–32.
114. Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viú JM, Tamarit D, et al. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucl Acids Res*. 2010;39(suppl_1):D70–D4.
115. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*. 2005;110(1–4):462–7.
116. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658–9.
117. Coghlan A, Tsai IJ, Berriman M. Creation of a comprehensive repeat library for a newly sequenced parasitic worm genome. *Protoc Exch*. 2018. <https://doi.org/10.1038/protex.2018.054>.

118. Smit A, Hubley R, Green P. RepeatMasker Open-40.2013–2015:<<http://www.repeatmasker.org>>.
119. Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: incorporating protein homology information into gene prediction with GeneMark-EP and AUGUSTUS. *Plant and Animal Genomes XXVI*. 2018.
120. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res*. 2005;33(20):6494–506.
121. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res*. 2008;18(12):1979–90.
122. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucl Acids Res*. 2006;34(suppl_2):W435–W9.
123. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
124. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–2.
125. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21(18):3674–6.
126. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. *Nucl Acids Res*. 2005;33(suppl_2):W116–W20.
127. Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*. 2005;21(19):3787–93.
128. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol*. 2016;428(4):726–31.
129. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucl Acids Res*. 2007;35(suppl_2):W182–W5.
130. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8(10):785–6.
131. Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mole Plant Pathol*. 2018;19(9):2094–110.
132. Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, et al. EffectorP: predicting fungal effector proteins from secretomes using machine learning. *New Phytol*. 2016;210(2):743–61.
133. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:1–9.
134. Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, et al. The conserved domain database in 2023. *Nucleic Acids Res*. 2023;51(D1):D384–8.
135. Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. *Nucl Acids Res*. 2004;32(suppl_2):W327–W31.
136. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46(W1):W95–101.
137. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2012;40(W1):W445–51.
138. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucl Acids Res*. 2023;51(W1):W115–21.
139. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucl Acids Res*. 2011;39(suppl_2):W29–37.
140. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60.
141. Busk PK, Pilgaard B, Lezyk MJ, Meyer AS, Lange L. Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function. *BMC Bioinformatics*. 2017;18(1):214.
142. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 2014;42(D1):D490–5.
143. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucl Acids Res*. 2019;47(W1):W81–W7.
144. Testa AC, Oliver RP, Hane JK. OcculterCut: A Comprehensive Survey of AT-Rich Regions in Fungal Genomes. *Genome Biol Evol*. 2016;8(6):2044–64.
145. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 2015;16(1):157.
146. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792–7.
147. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–3.
148. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37(5):1530–4.
149. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–74.
150. Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587–9.
151. Csűös M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics*. 2010;26(15):1910–2.
152. Cohen O, Ashkenazy H, Belinky F, Huchon D, Pupko T. GLOOME: gain loss mapping engine. *Bioinformatics*. 2010;26(22):2914–5.
153. Team R. RStudio: integrated development for R. RStudio, Inc, Boston, MA URL <http://www.rstudio.com>. 2015;42:14.
154. Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings. R package version. 2017;2(0).
155. Hane JK, Oliver RP. RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics*. 2008;9(1):478.
156. Van Wyk S, Harrison CH, Wingfield BD, De Vos L, van Der Merwe NA, Steenkamp ET. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. *PeerJ*. 2019;7: e7447.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.