High-Quality Genome Assemblies of 4 Members of the Podospora anserina Species Complex

S Lorena Ament-Velásquez, Aaron A Vogan, Ola Wallerman, Fanny E Hartmann, Valérie Gautier, Philippe Silar, Tatiana Giraud, Hanna Johannesson

High-Quality Genome Assemblies of 4 Members of the *Podospora anserina* Species Complex

S. Lorena Ament-Velásquez [1,*], Aaron A. Vogan [2], Ola Wallerman[3], Fanny E. Hartmann [4], Valérie Gautier[5], Philippe Silar [5], Tatiana Giraud[4], and Hanna Johannesson [2,6,7]

[1]Division of Population Genetics, Department of Zoology, Stockholm University, 106 91 Stockholm, Sweden
[2]Systematic Biology, Department of Organismal Biology, Uppsala University, 752 36 Uppsala, Sweden
[3]Department of Medical Biochemistry and Microbiology, Comparative Genetics and Functional Genomics, Uppsala University, 752 37 Uppsala, Sweden
[4]Ecologie Systematique Evolution, CNRS, Université Paris-Saclay, AgroParisTech, 91198 Gif-sur-Yvette, France
[5]Laboratoire Interdisciplinaire des Energies de Demain (LIED), Université de Paris Cité, F-75013 Paris, France
[6]The Royal Swedish Academy of Sciences, 114 18 Stockholm, Sweden
[7]Department of Ecology, Environmental and Plant Sciences, Stockholm University, 106 91 Stockholm, Sweden

*Corresponding author: Email: slament@protonmail.com.

Accepted: February 17, 2024

Abstract

The filamentous fungus *Podospora anserina* is a model organism used extensively in the study of molecular biology, senescence, prion biology, meiotic drive, mating-type chromosome evolution, and plant biomass degradation. It has recently been established that *P. anserina* is a member of a complex of 7 closely related species. In addition to *P. anserina*, high-quality genomic resources are available for 2 of these taxa. Here, we provide chromosome-level annotated assemblies of the 4 remaining species of the complex, as well as a comprehensive data set of annotated assemblies from a total of 28 *Podospora* genomes. We find that all 7 species have genomes of around 35 Mb arranged in 7 chromosomes that are mostly collinear and less than 2% divergent from each other at genic regions. We further attempt to resolve their phylogenetic relationships, finding significant levels of phylogenetic conflict as expected from a rapid and recent diversification.

Key words: Sordariales, *Podospora bellae-mahoneyi*, *Podospora pseudoanserina*, *Podospora pseudopauciseta*, *Podospora pseudocomata*, chromosomal rearrangements, phylogenomics.

Significance

Here, we provide a data set of 28 annotated genomes from the *Podospora anserina* species complex, including chromosome-level assemblies of 4 species that lacked a reference genome. With this data set in hand, biologists can take advantage of the molecular tools available for *P. anserina* to study evolutionary dynamics at the intersection between micro- and macroevolution, with particular emphasis on trait evolution, genome architecture, and speciation.

Introduction

The filamentous fungus *Podospora anserina* (order Sordariales) holds significant importance as a model for understanding ascomycete biology and beyond (Silar 2013). It has proved particularly valuable in advancing the study of molecular biology, senescence, heterokaryon incompatibility, sexual reproduction, prion biology, meiotic drive, and plant biomass degradation (Pinan-Lucarré et al. 2007; Silar 2013, 2020; Grognet et al. 2014; Hamann and Osiewacz 2018; Hartmann et al. 2021; Vogan et al. 2022). Its reference

genome was published as early as 2008 (Espagne et al. 2008), followed by chromosome-level assemblies of several wild-type strains (Vogan et al. 2021, 2019) and short-read population genomic data from Wageningen, The Netherlands, as well as a few strains from France and other localities (Ament-Velásquez et al. 2022). However, knowledge of its diversity, geographic distribution, ecology, and evolution lags behind. It is generally agreed that *P. anserina* is an obligately sexual coprophilous fungus, but there are observations of potential asexual spores (Boucher et al. 2017; Silar 2020) and endophytic stages (Matasyoh et al. 2011). The name *P. anserina* itself has been riddled with taxonomic uncertainties (Ament-Velásquez et al. 2020; Silar 2020), leading to confusion regarding the exact identity of the fungal material used in some studies. Unsurprisingly, a phylogenetic survey showed that many strains commonly regarded as *P. anserina* actually belong to at least 6 additional species scattered around the world (Boucher et al. 2017). Representatives of all these species have been sequenced with short-read technology, which was useful to explore the dynamics of recombination suppression around the mating-type locus (Hartmann et al. 2021). However, long-read data are necessary to understand the evolution and genetic basis of many traits. For example, the development of high-quality genomic resources of 2 of these species, *Podospora comata* and *Podospora pauciseta*, already provided important insights into the evolutionary dynamics of selfish genetic elements and genome architecture (Silar et al. 2018; Vogan et al. 2021, 2019).

As originally defined, only 1 or 2 strains are known for most members of the *P. anserina* species complex (Boucher et al. 2017), many available at the Westerdijk Fungal Biodiversity Institute Collection (identified with CBS numbers). All species have a similar morphology, mating system, and coprophilous habit, with the exception of the only known strain of *Podospora pseudocomata*, which was isolated from soil (Boucher et al. 2017; Hartmann et al. 2021). Despite their similarities, they are considered biological species, since there is reproductive isolation in the form of low mating success and female sterility in the hybrids (Boucher et al. 2017). Moreover, they are identifiable by differences at the fungal barcode ITS, as well as other nuclear markers (Boucher et al. 2017). Previous genomic comparisons showed that *P. anserina*, *P. comata*, and *P. pauciseta* are more than 98% identical in genic regions (Vogan et al. 2019), confirming that they are very closely related. However, their exact relationships remain unresolved. In this study, we generated chromosome-level annotated genome assemblies of the 4 remaining species (*Podospora bellae-mahoneyi*, *Podospora pseudoanserina*, *Podospora pseudopauciseta*, and *P. pseudocomata*), as well as short-read data from additional strains. In addition, we conducted a phylogenomic analysis to provide an evolutionary framework for addressing the variety of questions for which *Podospora* is well suited.

Results and Discussion

Genome Assemblies and Annotation

We isolated haploid cultures (of mating type + or −) from dikaryotic strains. From those, we selected 1 strain of each of the 4 species that lack a reference genome (hereafter, the focal strains) for Oxford Nanopore MinION and Illumina HiSeq sequencing (supplementary table S1, Supplementary Material online). In addition, we sequenced with Illumina HiSeq a known strain of *P. pauciseta* (CBS 451.62+), the type strain of *P. pseudoanserina* (CBS 253.71+), and 2 newly collected *P. comata* strains (Wageningen Collection numbers Wa132+ and Wa133−). Along with previously published assemblies and sequencing data of other members of the species complex, we assembled and annotated a total of 28 *Podospora* genomes, including the type strains of all 7 species (supplementary table S1, Supplementary Material online).

Whole genome assemblies of Oxford Nanopore MinION data from the focal strains recovered mostly chromosome-level scaffolds that are highly collinear with the reference genome of *P. anserina* (Fig. 1), although *P. pseudocomata* (strain CBS415.72−) has slightly more rearrangements. Thus, all species in the complex likely have 7 chromosomes, a similar genome size of around 35 Mb, and a repeat content ranging from more than 3% (*P. comata*) to around 7% (*P. pseudopauciseta*; supplementary table S1, Supplementary Material online) that is mostly concentrated in clusters (Fig. 1). Genome annotation using previously published RNA-seq data of *P. anserina* and *P. comata* (Vogan et al. 2021; Lelandais et al. 2022) resulted in similar protein-coding gene numbers for assemblies produced with long-read data, from 11,033 (*P. bellae-mahoneyi* strain CBS112042+) to 11,727 (*P. anserina* strain T_G+) genes, while the *P. anserina* reference itself has 10,803 predicted protein-coding genes (supplementary table S1, Supplementary Material online). The discrepancy in gene numbers with the reference is likely due to differences in the annotation method (with our pipeline, we recovered 11,660 genes in the *P. anserina* reference assembly). Similarly, the annotation of long-read assemblies gave comparable BUSCO numbers to the reference genome of *P. anserina* (supplementary table S1, Supplementary Material online), specifically 96.6% to 98.2% conserved proteins present. The annotation of short-read assemblies resulted in 90% to 93.5% conserved proteins, although assemblies alone reached BUSCO values more in line with the long-read assemblies (supplementary table S1, Supplementary Material online).

Phylogenomics and Comparative Genomics

We aimed at providing a phylogenetic context of the *P. anserina* species complex by using the 28 genome

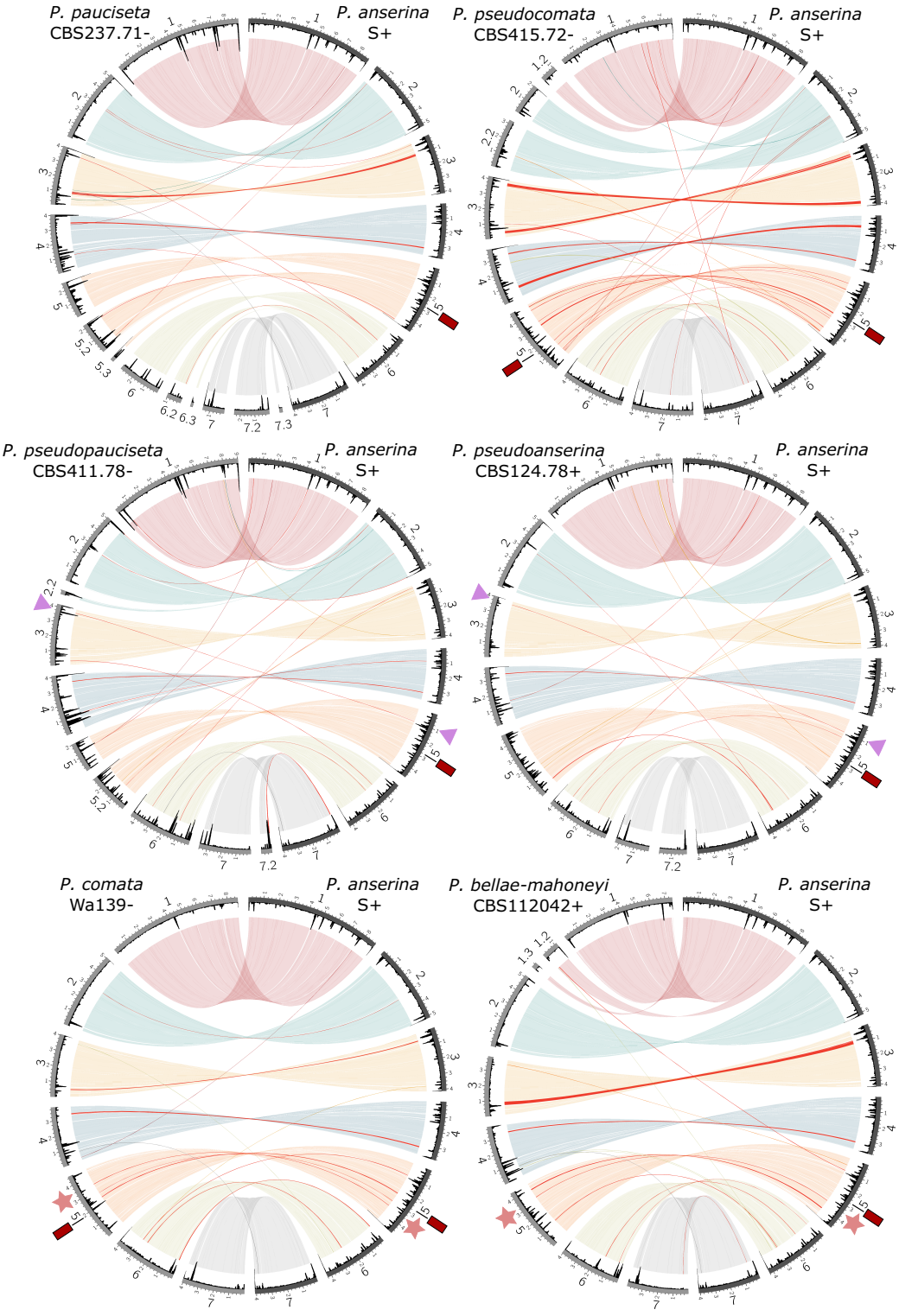

Fig. 1.—Circos plots comparing the reference assembly of *P. anserina* (strain S+, right side of each plot) to the best genome assembly of each of the other members of the species complex (left side). Light colors correspond to NUCmer alignments (larger than 5 kb) of the different chromosomes as defined in *P. anserina* (chr. 1: red; chr. 2: turquoise; chr. 3: yellow; chr. 4: blue; chr. 5: orange; chr. 6: olive green; chr 7: gray). Red solid links mark chromosome inversions or inverted translocations. The internal track in black is a histogram of repetitive element abundance calculated in sliding windows of 50 kb with steps of 10 kb. The stars and triangles mark shared structural variants (relative to *P. anserina*). The location of the insertion in chr. 5 discussed in the text is marked with a red square.

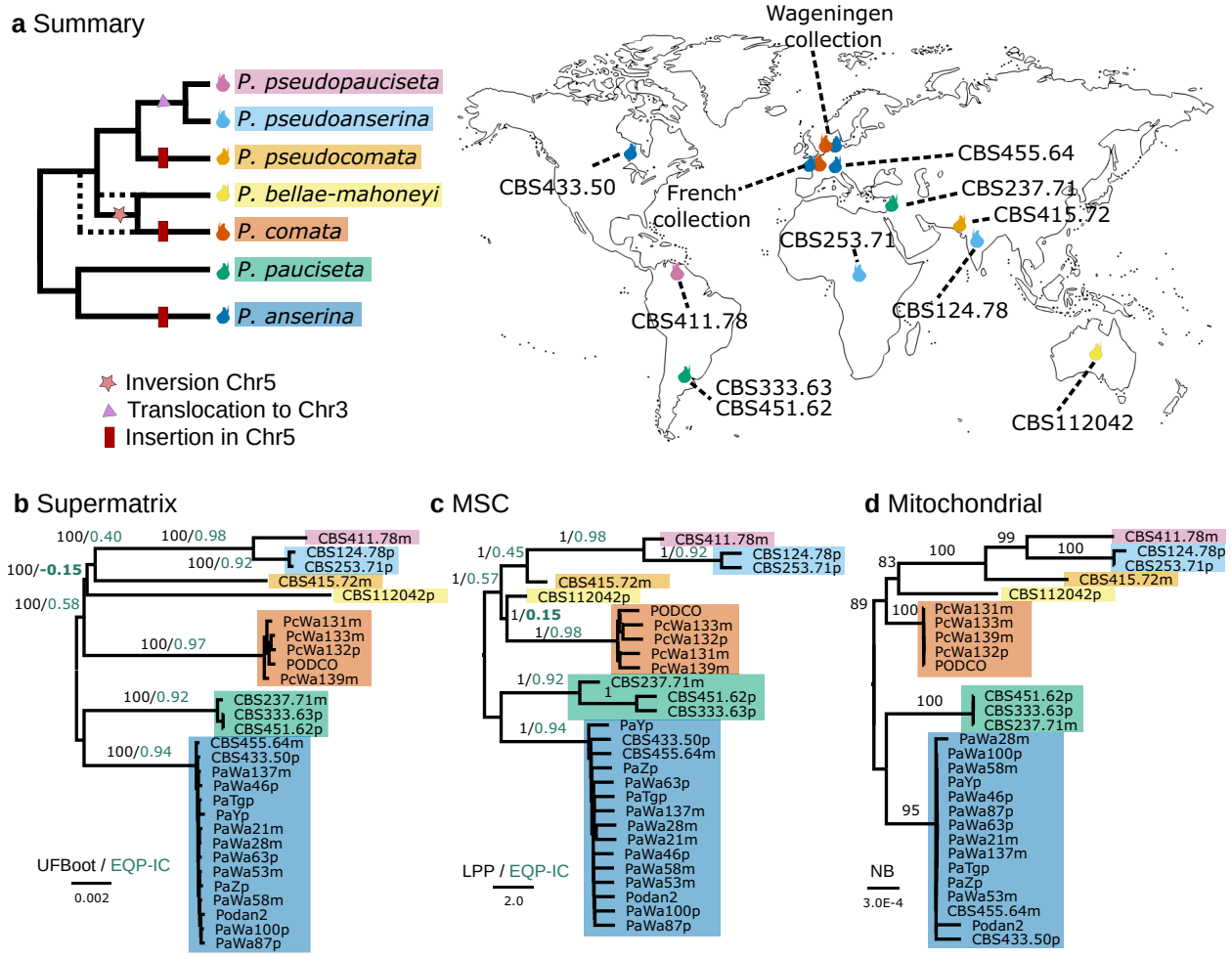

Fig. 2.—Phylogenetic relationships of the *Podospora* strains with genomic resources and their geographic distribution. a) Summary cladogram based on the phylogenomic analyses and the detected structural variants, with dotted branches illustrating an alternative topology. Fruiting body cartoons mark the country where the different strains were sampled. Phylogenetic relationships were inferred from a supermatrix ML analysis b) or a MSC analysis c) of nuclear genes, as well as a ML analysis of concatenated mitochondrial genes d). Rooting is tentative based on analyses with *C. samala* as an outgroup. Branch lengths of the phylograms are drawn to scale as indicated by the scale bar (nucleotide substitutions per site in b) and d) and coalescent units in c)). Different support metrics are shown next to their corresponding branches (within-species values are removed for clarity). The EQP-IC value of the conflicting branch is highlighted in bold. MSC, multispecies coalescent; UFBoot, ultrafast bootstrap; EQP-IC, extended quadripartition internode certainty; LPP, local posterior probability; NB, nonparametric bootstrap.

assemblies (Fig. 2; supplementary table S1, Supplementary Material online). The closest known relative of the *P. anserina* species complex is *Cercophora samala*, strain CBS 307.81 (Ament-Velásquez et al. 2020). Preliminary phylogenomic analyses using CBS 307.81 as the outgroup placed the clade of *P. anserina* and *P. pauciseta* as sister to the other *Podospora* species (supplementary fig. S1, Supplementary Material online). However, this *C. samala* strain is in fact too divergent (around 86% identity in nuclear genic regions to any *Podospora* species) relative to the species complex (>98% identical to each other), potentially creating long-branch attraction (Felsenstein 1981; Emms and Kelly 2017). Hence, only *Podospora* strains were considered below, and we tentatively rooted the phylogeny using *P. anserina* and *P. pauciseta*.

A summary of our results is found in Fig. 2a. We inferred groups of single-copy orthologous (SCO) genes and used them to produce 3 different phylogenetic analyses (Fig. 2b to d). All phylogenies resulted in well-supported species-level clades separated by very short internal branches, suggesting rapid diversification. Most relationships were congruent among analyses, except with regard to the relative positions of *P. comata* and *P. bellae-mahoneyi*. In the first analysis, a maximum likelihood (ML) phylogeny produced from a supermatrix of 1,000 nuclear SCO genes, *P. bellae-mahoneyi* was inferred as sister to the clade containing

P. pseudopauciseta, *P. pseudoanserina*, and *P. pseudocomata* (Fig. 2b). In contrast, a multispecies coalescent (MSC) analysis of all the 8,596 SCO genes recovered *P. comata* and *P. bellae-mahoneyi* as sister taxa (Fig. 2c). Lastly, a ML phylogeny of 8 mitochondrial genes was in support of the nuclear supermatrix tree, albeit with modest bootstrap values (Fig. 2d). To further explore this phylogenetic conflict, we obtained extended quadripartition internode certainty (EQP-IC) values (Zhou et al. 2020) for the 2 competing topologies. The EQP-IC score can range from 1 to −1. A score of 1 implies that all the SCO gene trees agree with a given branch, while a value of −1 reveals that all trees support an alternative topology. The score approaches 0 if the alternative topologies have similar frequencies among the gene trees (Zhou et al. 2020). As expected, we found intermediate positive values in most internal branches, which might be explained by incomplete lineage sorting (ILS) or introgression. The contesting branches in particular obtained values very close to 0, but there is slight support for the existence of a clade containing *P. comata* and *P. bellae-mahoneyi* (Fig. 2b and c).

Taking advantage of the extensive collinearity between genomes, we also attempted to find phylogenetically informative structural variants. In support of the MSC analysis, we found a medium-scale (33.89 kb, 13 genes) inversion in chromosome 5, relative to *P. anserina*'s genome, that is shared between *P. comata* and P. *bellae-mahoneyi* (Figs. 1 and 2a). In addition, we detected a shared translocation from chromosome 5 to chromosome 3 between *P. pseudoanserina* and *P. pseudopauciseta*, supporting their close sister relationship already observed in the phylogenetic analyses (Figs. 1 and 2a). However, in contradiction with our phylogenies, we found a region in chromosome 5 that is only present in *P. anserina*, *P. comata*, and *P. pseudocomata* (Figs. 1 and 2a). This region ranges from over 40 kb in *P. anserina* to just over 6 kb in *P. pseudocomata* and contains a number of genes and different transposable elements (TEs). Upon closer inspection, we found that the edges of this region are flanked by directed repeats of 4 bp, suggesting that this region represents a TE-mediated insertion, resembling the behavior of large TEs found in Pezizomycetes, including *Podospora* (Vogan et al. 2021; Gluck-Thaler et al. 2022). Its phylogenetic distribution might also be a consequence of ILS or introgression.

Conclusion

Here, we present high-quality annotated genome assemblies of 4 members of the *P. anserina* species complex. Together with the already available genomic resources, these new data build an evolutionary framework for further in-depth studies of this group of fungi. We provide a general idea of the genomic architecture and relationships between the sampled *Podospora* lineages, while illustrating the high levels of phylogenetic conflict that are typical of rapid species radiations. Moreover, we make available a comprehensive genomic data set of 28 strains for the study of fungal biology and evolution at shallow divergence scales. Combined with the wealth of molecular biology tools available for the model species *P. anserina*, this data set can be used to explore the evolution and function of pangenome content including metabolic clusters, selfish genetic elements like meiotic drivers and TEs, and the buildup of reproductive barriers in filamentous fungi.

Materials and Methods

All bioinformatic pipelines were done in Snakemake v. 7.25.0 or v. 7.32.3 (Mölder et al. 2021) and are available at https://github.com/SLAment/PodosporaGenomes, unless otherwise stated. Some of these pipelines rely on the Environment for Tree Exploration (ETE3) toolkit v. 3.1.3 (Huerta-Cepas et al. 2016).

Fungal Material

For detailed information about the strains used, see supplementary table S1 and Supplementary Material, Supplementary Material online. The strains with code starting with "CBS" were originally obtained from the CBS-KNAW Collection (https://wi.knaw.nl/Collection). In addition, we isolated 2 new strains of *P. comata* from rabbit dung collected in the area between Wageningen and Arnhem, The Netherlands (51°58′41.8″N, 5°50′39.6″E, locality Unksepad Oosterbeek) in September of 2016, which were deposited in the Wageningen Collection at the Laboratory of Genetics of the Wageningen University and Research (codes Wa132 and Wa133).

DNA Extraction and Sequencing

For Illumina sequencing, we grew the haploid strains on Petri dishes of either M2 medium (CBS124.78+ and CBS307.81−) or PASM0.2 (other strains) for 3 to 4 d (Vogan et al. 2019; Silar 2020). We scraped mycelium off the plates in order to obtain about 80 to 200 mg of mycelium per strain and stored it in 1.5-mL Eppendorf tube at −80 °C for at least 24 h before extraction. For most strains, whole genome DNA was extracted with the Fungal/Bacterial Microprep kit (Zymo; https://zymoresearch.eu/). In the case of CBS124.78+ and CBS307.81−, the mycelium was lyophilized for about 20 h, and DNA was extracted using the commercial Nucleospin Soil kit from Macherey Nagel. Paired-end libraries (150 bp reads) were sequenced by either the SNP and SEQ Technology platform (SciLifeLab, Uppsala, Sweden) on the Illumina HiSeq X platform (most strains) or by the high-throughput sequencing core facility of I2BC, Université Paris-Saclay (Centre de Recherche de Gif—http://www.i2bc.paris-saclay.fr/) on the Illumina NextSeq500 platform (CBS124.78+ and CBS307.81−).

For MinION Oxford Nanopore sequencing, we grew the strains in liquid cultures of 3% malt extract solution as in Vogan et al. (2019). High-molecular-weight DNA was extracted as in Sun et al. (2017), using the Genomic Tip G-500 columns (Qiagen) and the PowerClean DNA Clean-Up kit (MoBio Labs). The strain CBS 411.78– was prepared and sequenced with the ligation kit SQK108 in a 1-pot reaction using 500 ng DNA for end-prep and ligation (NEB Ultra-II ligase) and sequenced on an R9.4.1 flow cell. In addition, a rapid barcoding (RBK004) was made for CBS 411.78– to get sufficient coverage for assembly. Similarly, CBS 415.72– was also sequenced using both the RBK004 and LSK108 kit on an R9.4.1 flow cell to maximize yield. CBS 112042+ and CBS 124.78+ were sequenced on R9.4.1 flow cells using the LSK108 kit with 3 μg DNA as input for the end-prep reaction (NEB Ultra-II EP, 20 min at 20 °C and 20 min at 65 °C). Bead purification (SpeedBeads, GE) was done before ligation to deplete short fragments, and 1.5 μg DNA per sample was ligated to 20 μL AMX 1D using Blunt/TA ligase (30 min).

Genome Assembly

The paired-end HiSeq Illumina reads were cleaned from adapters using cutadapt v. 1.13 (Martin 2011) and Trimmomatic v. 0.36 (Bolger et al. 2014) with the following options: ILLUMINACLIP:adapters.fasta:1:30:9 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:30, as in Vogan et al. (2019). We used both forward and reverse paired-end reads in downstream analyses. De novo assemblies were produced as in Vogan et al. (2019). Specifically, we assembled the MinION reads with mean Phred quality (QV) above 9 and longer than 1 kb using Minimap2 v. 2.11 and Miniasm v. 0.2 (Li 2018, 2016). Racon v. 1.3.1 (Vaser et al. 2017) was used twice to polish the resulting assemblies based on the unfiltered reads. We further polished 5 times using the Illumina reads with Pilon v. 1.22 (Walker et al. 2014), which were mapped using BWA v. 0.7.17 (Li and Durbin 2009), with PCR duplicates marked by Picard v. 2.18.11 (http://broadinstitute.github.io/picard/) and with local indel realignment from the Genome Analysis Toolkit (GATK) v. 3.7 (Van der Auwera et al. 2018). For the strains without long-read data, we ran SPAdes v. 3.12.0 (Bankevich et al. 2012) with the *k*-mers 21,33,55,77 (most strains) or 21,29,37,45,53,61,79,87 (CBS307.81–) and the *--careful* option. Due to high numbers of read pairs with read mates mapping on different chromosomes during the read mapping procedure (about 20% of mapping read pairs), we used each paired-end sequencing run as 2 independent single-end sequencing runs for read mapping and de novo assembly of CBS124.78+ and CBS307.81–, respectively (see Hartmann et al. 2021).

In the case of samples with long-read data, the scaffolds were assigned to chromosomes and reoriented by mapping them to the reference genome of the strain S (Espagne et al. 2008), which is available at the Joint Genome Institute MycoCosm website (https://mycocosm.jgi.doe.gov/Podan2/Podan2.home.html) as "Podan2". Mapping was performed with the NUCmer program from the MUMmer package v. 4.0.0beta2 (Kurtz et al. 2004) using the parameters *-b 2000 -c 200 --maxmatch*. Contigs smaller than 100 kb that contained rDNA repeats or mitochondrial sequences were discarded (except for the largest mitochondrial contig). Genome quality statistics were calculated with QUAST v. 4.6.3 (Mikheenko et al. 2016). Mean depth of coverage was obtained using Qualimap v.2.2 (Okonechnikov et al. 2016). We also used BUSCO v. 5.3.1 (Manni et al. 2021) with the 3,817 Sordariomycetes_odb10 ortholog set to assess assembly completeness. As dependencies, we used BLAST suit 2.12.0+ (Camacho et al. 2009), AUGUSTUS v. 3.4.0 (Stanke and Waack 2003), and HMMER v. 3.2.1 (Mistry et al. 2013).

We verified the correct assembly of the mitochondrial contig in the focal strains CBS 112042+, CBS 124.78+, CBS 411.78–, and CBS 415.72–, by visual inspection of long- and short-read mapping. We found a misassembly in the contig of CBS 415.72– around the first exon of the *cox1* gene. Hence, we extracted the long reads mapped to this mitochondrial contig using the *bam2fq* option of SAMtools v. 1.17 (Danecek et al. 2021) and reassembled them with Flye v. 2.9.1 (Kolmogorov et al. 2020) with the arguments *--iterations 2 --meta --keep-haplotypes*. We recovered 2 circular contigs, 1 of which proved to be formed by multiple tandem repeats of the first part of *cox1*, a configuration known as α senDNA or plDNA (Cummings et al. 1985; Hamann and Osiewacz 2018). The other contig corresponded to the full mitochondrion. We polished the mitochondrial contig 3 times using the Illumina reads as above and discarded the plDNA. We manually recircularized the mitochondrial contigs of the focal strains to avoid breaking genes.

Genome Annotation

The annotation of all genomes was done with a modified version of a previous pipeline (Vogan et al. 2021). Briefly, we used previously produced (Vogan et al. 2019) training files for SNAP release 2013-11-29 (Lomsadze et al. 2005) and GeneMark-ES v. 4.38 (Lomsadze et al. 2005; Ter-Hovhannisyan et al. 2008) within the program MAKER v. 3.01.04 (Holt and Yandell 2011; Campbell et al. 2014) to generate gene models for all species. MAKER was run with the following dependencies: BLAST suit 2.13.0+, tRNAscan-SE v. 1.3.1 (Lowe and Eddy 1997), Exonerate v. 2.4.0 (Slater and Birney 2005), and RepeatMasker v. 4.1.0 (http://www.repeatmasker.org/). As external protein evidence, we used the sequences from the PODANS_v2016 annotation (Lelandais et al. 2022) and from the reference

genome of *P. comata* (PODCO; Silar et al. 2018), along with a small set of manually curated proteins (available in the GitHub repository). We also used transcript models as external evidence from 2 sources: the curated set of transcripts with defined transcription starts and ends from PODANS_v2016 and transcript models of published RNA-seq data of *P. anserina* and *P. comata* (Vogan et al. 2021, 2019). The latter were produced as in Vogan et al. (2021), using STAR v. 2.7.10b (Dobin et al. 2013), Cufflinks v. 2.2.1 (Trapnell et al. 2010), and TransDecoder v. 5.7.0 (Haas et al. 2013). Additionally, we used the custom library "PodoTE-1.00" (https://github.com/johannessonlab/SpokBlockPaper/blob/master/Annotation/data/) to annotate repeated elements in all species using RepeatModeler v. 1.0.8 (http://www.repeatmasker.org/RepeatModeler/).

The gene models produced by the MAKER pipeline were functionally annotated with Funannotate v. 1.8.15 (Palmer and Stajich 2020) using the *annotate* function with the dependencies HMMER 3.3.2 (Mistry et al. 2013), Diamond v. 2.1.6 (Buchfink et al. 2021) with the UniProt DB version 2023_01, InterProScan v. 5.62-94.0 (Jones et al. 2014; Blum et al. 2021), bedtools v. 2.30.0 (Quinlan and Hall 2010), Eggnog-mapper v. 2.1.10 (Cantalapiedra et al. 2021) with the database emapperdb-5.0.2, and the fungal version of antiSMASH v. 6.1.1 (Blin et al. 2021). The input assemblies were soft masked with RepeatModeler as above. Mitochondrial annotation was done with the online version of MFannot (Lang et al. 2023), setting the genetic code to 4 (https://megasun.bch.umontreal.ca/apps/mfannot/; consulted in July 2023; see Supplementary Material for details.

Comparative Genomics

In order to explore collinearity among the *Podospora* species, we used NUCmer with parameters *-b 2000 -c 2000 --maxmatch* to align the long-read assembly of a representative of each species against Podan2 (Fig. 1). We further calculated the coverage distribution of repetitive elements along chromosomes using the RepeatMasker annotation and by dividing the genome in windows of 50 kb with steps of 10 kb using the utilities *makewindows* and *coverage* of BEDTools v. 2.29.0 (Quinlan and Hall 2010; Quinlan 2014). Both the alignments and coverage distributions were plotted using Circos v. 0.69.6 (Krzywinski et al. 2009). We removed all alignments smaller than 5 kb to exclude the most abundant TEs.

Phylogenomic Analyses

In order to resolve the relationships between the *Podospora* species, we inferred SCO groups by running OrthoFinder v. 2.5.2 (Emms and Kelly 2019) with a single representative per species: strains S+ (Podan2), PODCO, CBS 237.71–, CBS 124.78+, CBS 411.78–, CBS 415.72, and CBS 112042+. OrthoFinder was run with the proteins predicted for the chosen strains. However, the low level of divergence within the species complex makes the proteins largely uninformative. Hence, once SCO groups were defined, we used the ortholog of the reference S+ as a BLASTn query to retrieve the nucleotide sequences of the corresponding homologs (including introns) in all the *Podospora* strains in our data set (supplementary table S1, Supplementary Material online). We kept only the orthogroups with a single sequence per strain, resulting in a total of 8,596 SCO groups. These were then aligned with MAFFT v. 7.407 (Katoh and Toh 2008) with the options *--adjustdirection --anysymbol --maxiterate 1000 --retree 1 --localpair*. We inferred ML trees of each alignment using IQ-TREE v. 2.2.3 (Nguyen et al. 2015; Hoang et al. 2017) with parameters *-m MFP -seed 1234 -bnni --keep-ident -bb 1000*. To reduce noise in our data set, we collapsed branches with ultrafast bootstraps (UFBoots) support lower than 95% into polytomies using Newick utilities v. 1.6 (Junier and Zdobnov 2010). These trees were then given to ASTRAL v. 5.7.3 (Zhang et al. 2018) to construct a MSC phylogeny. In addition, we randomly selected 1,000 of the SCO to form a concatenated alignment (supermatrix) of 1,732,364 sites (39,771 [2.4%] informative) and produced a ML phylogeny with IQ-TREE as above. In order to evaluate the level of conflict within the SCO trees with respect to both the supermatrix ML and MSC phylogenies, we calculated the EQP-IC score with the program QuartetScores v. 1.0 (Zhou et al. 2020) using the unrooted reference trees and the SCO trees with collapsed low-support branches from above as the evaluation set (see Supplementary Material for details).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We acknowledge the support given by the National Genomics Infrastructure (NGI)/Uppsala Genome Center on massive parallel DNA sequencing. The computations were performed on resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) partially funded by the Swedish Research Council through grant agreement no. 2022-06725 and no. 2018-05973.

Funding

This work was supported by the European Research Council (ERC) grant ERC-2014-CoG (project 648143 SpoKiGen) to H.J. and advanced grant ERC-2019-ADV (project 832352 EvolSexChrom), as well as the Louis D. Foundation

(Institut de France) to T.G., by the Agence Nationale de la Recherche (PIA grant # ANR-20-IDEES-0002) to F.H., and support from Stiftelsen Lars Hierta Minne, the Nilsson-Ehle Endowments of the Kungliga Fysiografiska Sällskapet i Lund, and the Swedish Research Council (grant 2022-00341) to S.L.A.-V.

Data Availability

The genome assemblies and raw sequencing data are available at GenBank under the BioProject PRJNA685103 (see supplementary table S1, Supplementary Material online). The reads of the strains sequenced with long-read technology in this study have BioSample accessions SAMN17076437 (CBS124.78+), SAMN17076438 (CBS411.78–), SAMN17076439 (CBS415.72–), and SAMN17076440 (CBS112042+). The annotated genome assemblies are available in GenBank under accession numbers JAFFHC000000000, JAFFHB000000000, JAFFHA000000000, and JAFFGZ000000000. The strains newly sequenced with Illumina have BioSample accessions SAMN37845441 (Wa132+), SAMN37845442 (Wa133–), SAMN37845443 (CBS 451.62+), and SAMN37845444 (CBS 253.71+). In addition, all genome assemblies, annotations in gff3 format, and alignments are available in Dryad Digital Repository (https://doi.org/10.5061/dryad.1vhhmgr0j). Scripts and Snakemake pipelines are available at https://github.com/SLAment/PodosporaGenomes.

Literature Cited

Ament-Velásquez SL, Johannesson H, Giraud T, Debuchy R, Saupe SJ, Debets AJM, Bastiaans E, Malagnac F, Grognet P, Peraza-Reyes L, et al. The taxonomy of the model filamentous fungus *Podospora anserina*. MycoKeys., 2020:75:51–69. https://doi.org/10.3897/mycokeys.75.55968.

Ament-Velásquez SL, Vogan AA, Granger-Farbos A, Bastiaans E, Martinossi-Allibert I, Saupe SJ, de Groot S, Lascoux M, Debets AJM, Clavé C, et al. Allorecognition genes drive reproductive isolation in *Podospora anserina*. Nat Ecol Evol. 2022:6(7):910–923. https://doi.org/10.1038/s41559-022-01734-x.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 2012:19(5):455–477. https://doi.org/10.1089/cmb.2012.0021.

Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 2021:49(W1): W29–W35. https://doi.org/10.1093/nar/gkab335.

Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021:49(D1):D344–D354. https://doi.org/10.1093/nar/gkaa977.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014:30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Boucher C, Nguyen T-S, Silar P. Species delimitation in the *Podospora anserina/P. pauciseta/P. comata* species complex (Sordariales). Cryptogam Mycol. 2017:38(4):485–506. https://doi.org/10.7872/crym/v38.iss4.2017.485.

Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021:18(4): 366–368. https://doi.org/10.1038/s41592-021-01101-x.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics 2009:10(1):421. https://doi.org/10.1186/1471-2105-10-421.

Campbell MS, Holt C, Moore B, Yandell M. Genome annotation and curation using MAKER and MAKER-P. Curr. Protoc. Bioinform. 2014:48(1). https://doi.org/10.1002/0471250953.bi0411s48.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021:38(12):5825–5829. https://doi.org/10.1093/molbev/msab293.

Cummings DJ, MacNeil IA, Domenico J, Matsuura ET. Excision-amplification of mitochondrial DNA during senescence in *Podospora anserina*. DNA sequence analysis of three unique "plasmids.". J Mol Biol. 1985:185(4):659–680. https://doi.org/10.1016/0022-2836(85)90052-X.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. GigaScience 2021:10(2): giab008. https://doi.org/10.1093/gigascience/giab008.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-Seq aligner. Bioinformatics 2013:29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.

Emms DM, Kelly S. STRIDE: species tree root inference from gene duplication events. Mol Biol Evol. 2017:34(12):3267–3278. https://doi.org/10.1093/molbev/msx259.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:20(1):238. https://doi.org/10.1186/s13059-019-1832-y.

Espagne E, Lespinet O, Malagnac F, Da Silva C, Jaillon O, Porcel BM, Couloux A, Aury J-M, Ségurens B, Poulain J, et al. The genome sequence of the model ascomycete fungus *Podospora anserina*. Genome Biol. 2008:9(5):R77. https://doi.org/10.1186/gb-2008-9-5-r77.

Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 1981:17(6):368–376. https://doi.org/10.1007/BF01734359.

Gluck-Thaler E, Ralston T, Konkel Z, Ocampos CG, Ganeshan VD, Dorrance AE, Niblack TL, Wood CW, Slot JC, Lopez-Nicora HD, et al. Giant *Starship* elements mobilize accessory genes in fungal genomes. Mol Biol Evol. 2022:39(5):msac109. https://doi.org/10.1093/molbev/msac109.

Grognet P, Bidard F, Kuchly C, Chan L, Tong H, Coppin E. Maintaining two mating types : structure of the mating type locus and its role in heterokaryosis in *Podospora anserina*. Genetics. 2014:197:421–432. https://doi.org/10.1534/genetics.113.159988.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013:8(8): 1494–1512. https://doi.org/10.1038/nprot.2013.084.

Hamann A, Osiewacz HD. Conn's handbook of models for human aging. 2nd ed. London, San Diego, Cambridge, Oxford: Elsevier; 2018. p. 431–444.

Hartmann FE, Ament-Velásquez SL, Vogan AA, Gautier V, Le Prieur S, Berramdane M, Snirc A, Johannesson H, Grognet P, Malagnac F,

et al. Size variation of the nonrecombining region on the mating-type chromosomes in the fungal *Podospora anserina* species complex. Mol Biol Evol. 2021:38(6):2475–2492. https://doi.org/10.1093/molbev/msab040.

Hoang DT, Chernomor O, Haeseler AV, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2017:35(2):518–522. https://doi.org/10.1093/molbev/msx281.

Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011:12(1). https://doi.org/10.1186/1471-2105-12-491.

Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol. Biol. Evol. 2016:33(6):1635–1638. https://doi.org/10.1093/molbev/msw046.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics 2014:30(9):1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Junier T, Zdobnov EM. The Newick utilities: high-throughput phylogenetic tree processing in the U NIX shell. Bioinformatics. 2010:26(13):1669–1670. https://doi.org/10.1093/bioinformatics/btq243.

Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinformatics. 2008:9(4):81–92. https://doi.org/10.1093/bib/bbn013.

Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods. 2020:17(11):1103–1110. https://doi.org/10.1038/s41592-020-00971-x.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009:19(9):1639–1645. https://doi.org/10.1101/gr.092759.109.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. Genome Biol. 2004:5(2):R12. https://doi.org/10.1186/gb-2004-5-2-r12.

Lang BF, Beck N, Prince S, Sarrasin M, Rioux P, Burger G. Mitochondrial genome annotation with MFannot: a critical analysis of gene identification and gene model prediction. Front. Plant Sci. 2023:14:1222186. https://doi.org/10.3389/fpls.2023.1222186.

Lelandais G, Remy D, Malagnac F, Grognet P. New insights into genome annotation in *Podospora anserina* through re-exploiting multiple RNA-seq data. BMC Genomics. 2022:23(1):859. https://doi.org/10.1186/s12864-022-09085-4.

Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics. 2016:32(14):2103–2110. https://doi.org/10.1093/bioinformatics/btw152.

Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:34(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009:25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

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):6494–6506. https://doi.org/10.1093/nar/gki937.

Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997:25(5):955–964. https://doi.org/10.1093/nar/25.5.955.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021:38(10):4647–4654. https://doi.org/10.1093/molbev/msab199.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011:17(1):10. https://doi.org/10.14806/ej.17.1.200.

Matasyoh JC, Dittrich B, Schueffler A, Laatsch H. Larvicidal activity of metabolites from the endophytic *Podospora* sp. against the malaria vector *Anopheles gambiae*. Parasitol Res. 2011:108(3):561–566. https://doi.org/10.1007/s00436-010-2098-1.

Mikheenko A, Valin G, Prjibelski A, Saveliev V, Gurevich A. Icarus: visualizer for *de novo* assembly evaluation. Bioinformatics. 2016:32(21):3321–3323. https://doi.org/10.1093/bioinformatics/btw379.

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 2013:41(12):e121. https://doi.org/10.1093/nar/gkt263.

Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, et al. Sustainable data analysis with Snakemake. F1000Res. 2021:10:33. https://doi.org/10.12688/f1000research.29032.2.

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):268–274. https://doi.org/10.1093/molbev/msu300.

Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 2016:32(2):292–294. https://doi.org/10.1093/bioinformatics/btv566.

Palmer JM, Stajich J. 2020. Funannotate v1.8.1: Eukaryotic genome annotation. [Computer software]. Zenodo. Version v1.8.1. https://doi.org/10.5281/zenodo.4054262.

Pinan-Lucarré B, Paoletti M, Clavé C. Cell death by incompatibility in the fungus *Podospora*. Semin Cancer Biol. 2007:17(2):101–111. https://doi.org/10.1016/j.semcancer.2006.11.009.

Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. Curr Protoc Bioinform. 2014:47:11.12.1–11.12.34. https://doi.org/10.1002/0471250953.bi1112s47.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010:26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033.

Silar P. *Podospora anserina*: from laboratory to biotechnology. In: Horwitz BA Mukherjee PK Mukherjee M, Kubicek CP, editors. Genomics of soil- and plant-associated fungi, soil biology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 283–309.

Silar P. 2020. Podospora anserina. France: HAL, Centre pour la Communication scientifique directe Villeurbanne (CCSD), ISBN n°978-2-9555841-2-5.

Silar P, Dauget JM, Gautier V, Grognet P, Chablat M, Hermann-Le Denmat S, Couloux A, Wincker P, Debuchy R. A gene graveyard in the genome of the fungus *Podospora comata*. Mol Genet Genomics. 2018:294(1):177–190. https://doi.org/10.1007/s00438-018-1497-3.

Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 2005:6(1):31. https://doi.org/10.1186/1471-2105-6-31.

Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 2003:19(suppl_2):ii215–ii225. https://doi.org/10.1093/bioinformatics/btg1080.

Sun Y, Svedberg J, Hiltunen M, Corcoran P, Johannesson H. Large-scale suppression of recombination predates genomic

rearrangements in *Neurospora tetrasperma*. Nat Commun. 2017:8(1):1140. https://doi.org/10.1038/s41467-017-01317-6.
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): 1979–1990. https://doi.org/10.1101/gr.081612.108.
Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010:28(5): 511–515. https://doi.org/10.1038/nbt.1621.
Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinform. 2018. https://doi.org/10.1002/0471250953.bi1110s43.
Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017:27(5):737–746. https://doi.org/10.1101/gr.214270.116.
Vogan AA, Ament-Velásquez SL, Bastiaans E, Wallerman O, Saupe SJ, Suh A, Johannesson H. The Enterprise, a massive transposon carrying Spok meiotic drive genes. Genome Res. 2021:31(5):789–798. https://doi.org/10.1101/gr.267609.120.
Vogan AA, Ament-Velásquez SL, Granger-Farbos A, Svedberg J, Bastiaans E, Debets AJ, Coustou V, Yvanne H, Clavé C, Saupe SJ, et al. Combinations of Spok genes create multiple meiotic drivers in *Podospora*. eLife 2019:8:e46454. https://doi.org/10.7554/eLife.46454.
Vogan AA, Martinossi-Allibert I, Ament-Velásquez SL, Svedberg J, Johannesson H. The spore killers, fungal meiotic driver elements. Mycologia 2022:114(1):1–23. https://doi.org/10.1080/00275514.2021.1994815.
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014:9(11): e112963. https://doi.org/10.1371/journal.pone.0112963.
Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 2018:19(S6). https://doi.org/10.1186/s12859-018-2129-y.
Zhou X, Lutteropp S, Czech L, Stamatakis A, Looz MV, Rokas A. Quartet-based computations of internode certainty provide robust measures of phylogenetic incongruence. Syst Biol. 2020:69(2): 308–324. https://doi.org/10.1093/sysbio/syz058.

Associate editor: Li-Jun Ma