



**HAL**  
open science

## Genome-scale phylogeny and comparative genomics of the fungal order Sordariales

Noah Hensen, Lucas Bonometti, Ivar Westerberg, Ioana Onut Brännström, Sonia Guillou, Sandrine Cros-Arteil, Sara Calhoun, Sajeet Haridas, Alan Kuo, Stephen Mondo, et al.

► **To cite this version:**

Noah Hensen, Lucas Bonometti, Ivar Westerberg, Ioana Onut Brännström, Sonia Guillou, et al.. Genome-scale phylogeny and comparative genomics of the fungal order Sordariales. *Molecular Phylogenetics and Evolution*, 2023, 189, pp.107938. 10.1016/j.ympev.2023.107938 . hal-04709225

**HAL Id: hal-04709225**

**<https://hal.inrae.fr/hal-04709225v1>**

Submitted on 25 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



## Genome-scale phylogeny and comparative genomics of the fungal order Sordariales

Noah Hensen<sup>a</sup>, Lucas Bonometti<sup>b</sup>, Ivar Westerberg<sup>a</sup>, Ioana Onut Brännström<sup>c,d</sup>, Sonia Guillou<sup>b</sup>, Sandrine Cros-Aarteil<sup>b</sup>, Sara Calhoun<sup>e</sup>, Sajeet Haridas<sup>e</sup>, Alan Kuo<sup>e</sup>, Stephen Mondo<sup>e</sup>, Jasmyn Pangilinan<sup>e</sup>, Robert Riley<sup>e</sup>, Kurt LaButti<sup>e</sup>, Bill Andreopoulos<sup>e</sup>, Anna Lipzen<sup>e</sup>, Cindy Chen<sup>e</sup>, Mi Yan<sup>e</sup>, Chris Daum<sup>e</sup>, Vivian Ng<sup>e</sup>, Alicia Clum<sup>e</sup>, Andrei Steindorff<sup>e</sup>, Robin A. Ohm<sup>e</sup>, Francis Martin<sup>f</sup>, Philippe Silar<sup>g</sup>, Donald O. Natvig<sup>h</sup>, Christophe Lalanne<sup>g</sup>, Valérie Gautier<sup>g</sup>, Sandra Lorena Ament-Velásquez<sup>i</sup>, Åsa Kruys<sup>j</sup>, Miriam I. Hutchinson<sup>h</sup>, Amy Jo Powell<sup>k</sup>, Kerrie Barry<sup>e</sup>, Andrew N. Miller<sup>l</sup>, Igor V. Grigoriev<sup>e,m</sup>, Robert Debuchy<sup>n</sup>, Pierre Gladieux<sup>b</sup>, Markus Hiltunen Thorén<sup>a,o</sup>, Hanna Johannesson<sup>a,o,\*</sup>

<sup>a</sup> Stockholm University, Department of Ecology, Environment and Plants Sciences, Stockholm, Sweden

<sup>b</sup> University of Montpellier, PHIM Plant Health Institute, Montpellier, France

<sup>c</sup> Oslo University, Natural History Museum, Oslo, Norway

<sup>d</sup> Uppsala University, Department of Ecology and Genetics, Uppsala, Sweden

<sup>e</sup> Lawrence Berkeley National Laboratory, U.S. Department of Energy Joint Genome Institute, Berkeley, CA, USA

<sup>f</sup> Université de Lorraine, INRAE, Champenoux, France

<sup>g</sup> Université de Paris Cité, Laboratoire Interdisciplinaire des Energies de Demain, Paris, France

<sup>h</sup> University of New Mexico, Department of Biology, Albuquerque, USA

<sup>i</sup> Stockholm University, Department of Zoology, Stockholm, Sweden

<sup>j</sup> Uppsala University, Museum of Evolution, Uppsala, Sweden

<sup>k</sup> Sandia National Laboratories, Dept. of Systems Design and Architecture, Albuquerque, USA

<sup>l</sup> University of Illinois Urbana-Champaign, Illinois Natural History Survey, USA

<sup>m</sup> University of California Berkeley, Department of Plant and Microbial Biology, Berkeley, CA, USA

<sup>n</sup> Université Paris-Saclay, Institute for Integrative Biology of the Cell, Gif-sur-Yvette, France

<sup>o</sup> The Royal Swedish Academy of Sciences, Stockholm, Sweden

### ARTICLE INFO

#### Keywords:

Whole-genome phylogeny  
Podosporaceae  
Chaetomiaceae  
Sordariaceae  
Genome evolution

### ABSTRACT

The order Sordariales is taxonomically diverse, and harbours many species with different lifestyles and large economic importance. Despite its importance, a robust genome-scale phylogeny, and associated comparative genomic analysis of the order is lacking.

In this study, we examined whole-genome data from 99 Sordariales, including 52 newly sequenced genomes, and seven outgroup taxa. We inferred a comprehensive phylogeny that resolved several contentious relationships amongst families in the order, and cleared-up intrafamily relationships within the *Podosporeae*.

Extensive comparative genomics showed that genomes from the three largest families in the dataset (*Chaetomiaceae*, *Podosporeae* and *Sordariaceae*) differ greatly in GC content, genome size, gene number, repeat percentage, evolutionary rate, and genome content affected by repeat-induced point mutations (RIP). All genomic traits showed phylogenetic signal, and ancestral state reconstruction revealed that the variation of the properties stems primarily from within-family evolution. Together, the results provide a thorough framework for understanding genome evolution in this important group of fungi.

\* Corresponding author.

E-mail address: [hanna.johannesson@su.se](mailto:hanna.johannesson@su.se) (H. Johannesson).

<https://doi.org/10.1016/j.ympev.2023.107938>

Received 26 June 2023; Received in revised form 28 September 2023; Accepted 4 October 2023

Available online 10 October 2023

1055-7903/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The order Sordariales (Ascomycota) is one of the most taxonomically diverse groups within the Sordariomycete fungi (Huhndorf et al., 2004). The order is of economic and ecological importance and contains species inhabiting a wide variety of natural habitats (Huang et al., 2021; Huhndorf et al., 2004; Hyde, 2020). The order also includes well-known model-organisms such as *Neurospora crassa* and *Podospora anserina*, both of which have been key-players in important scientific discoveries (Davis & Perkins, 2002; Gladieux et al., 2020; Roche et al., 2014; Silar, 2020). It furthermore contains species producing a diversity of biologically active secondary metabolites with interesting drug-like properties (Charria-Girón et al., 2022; Noumeur et al., 2020), and the highest known number of thermophilic species, which have large industrial relevance (Hutchinson et al., 2019; Patel & Rawat, 2021; van den Brink et al., 2015).

The order Sordariales was first described in 1960 by Chadeffaud and validated by Hawksworth and Eriksson based on morphological data (Chadeffaud & Emberger, 1963; Hawksworth & Eriksson, 1986), after which Huhndorf et al. (2004) made an initial attempt to resolve the phylogenetic relationships within the order based on LSU sequence analysis. Since then, several studies have been performed on the Sordariales in order to delimitate its families and their largest genera (e.g. Cai et al., 2006; Kruijs et al., 2015; Miller & Huhndorf, 2005; Wang et al., 2019). Past phylogenetic studies of the Sordariales have utilized many-taxa/few-genes approaches that have substantially advanced our understanding of phylogenetic relationships inside the order. Over time, however, lack of resolution has remained an issue and many parts of the tree have remained poorly resolved (see e.g. Ament-Velásquez et al., 2020; Huang et al., 2021; Marin-Felix & Miller, 2022; Wang et al., 2019).

In 2004, Huhndorf et al. (2004) divided Sordariales into three families: *Chaetomiaceae*, *Lasiosphaeriaceae* and *Sordariaceae*. Since then, several lasiosphaeriaceous taxa were reassigned to establish the additional families *Diplogelasinosporaceae*, *Naviculisporaceae*, and *Schizotheciaceae*, with the remaining species placed in *Lasiosphaeriaceae sensu lato* (Marin-Felix et al., 2020). Around the same time, the family *Podosporeae* was introduced to accommodate the *Podospora* type species and was further divided into three main clades (Wang et al., 2019). However, the branching order of these three clades and the taxonomy of the *Podospora* type species have to date remained unresolved (e.g. Ament-Velásquez et al., 2020; Silar, 2020; Vogan et al., 2021). Recently, Huang et al. (2021) introduced an additional five new families to the order (Huang et al., 2021), although ongoing debates about this taxonomic classification remain (Charria-Girón et al., 2022; Marin-Felix and Miller, 2022). Thus, obtaining and investigating additional sequence data has been a high priority to further determine the phylogenetic and taxonomic affinities in the Sordariales (Huang et al., 2021; Hyde, 2020; Marin-Felix and Miller, 2022).

In addition to resolving taxonomic ambiguities, a robust phylogenetic framework of Sordariales facilitates comparative genomic analysis, which is required to identify the key factors underlying genome evolution in Sordariales. Properties such as genome size, gene number and GC content can vary widely amongst different organisms (Li and Du, 2014). Such genomic properties are often found to correlate with lifestyles and co-vary with each other. For example, recent comparisons of trait values between subdivisions Saccharomycotina and Pezizomycotina showed that evolutionary rate, GC content, genome size, and number of protein-coding genes were highly variable (Shen et al., 2020). Furthermore, covariation is seen between numerous genomic traits and repeat-induced point mutation (RIP), a process unique to fungi that gives rise to multiple G:C → A:T mutations in repeated sequences (see e.g. Borkovich et al., 2004; Galagan and Selker, 2004; Gladyshev, 2017; Clutterbuck, 2011). Another example is the strong correlation between genome size and transposable-element content. At the dawn of large-scale sequencing, this was one of the first traits to show covariation in

eukaryotes (Gregory, 2005), and today the correlation is well established in fungi (Fouché et al., 2022; Mat Razali et al., 2019; Oggenfuss et al., 2021).

An overview of genomic features across the Sordariales will furthermore increase knowledge on the impact of fungal lifestyle on genome evolution and covariation of genomic properties. *Podospora anserina*, for example, has a coprophilic nature, believed to select for rapid sexual reproduction (Geydan et al., 2012; Griffiths, 1992; Scheckhuber and Osiewacz, 2008), leading to short generation times in the family. Many studies have found negative correlations between evolutionary rate and generation time in eukaryotes (e.g. Thomas et al., 2010; Welch et al., 2008), but no research has been done comparing the evolutionary rates of *Podosporeae* to sister groups with known longer generation times. Similarly, the *Chaetomiaceae* is well known for having the highest described number of thermophilic species (Geydan et al., 2012). Genomes of three thermophilic fungal species appear to be characterized by a relatively small size, high G:C content, and low levels of repetitive DNA (Patel and Rawat, 2021; van Noort et al., 2013), but how general these characteristics are for thermophilic versus non-thermophilic species is not yet known.

In this study, we have used whole-genome data from 99 Sordariales strains that span the diversity of nine families (Huang et al., 2021; Marin-Felix and Miller, 2022). Together with seven outgroup genomes, this creates a more comprehensive phylogenomic dataset than previously available. We used the resulting phylogeny as a framework for analyses of relatedness and evolution of genomic properties in this group of important fungi.

## 2. Materials and methods

### 2.1. Data collection

Altogether, 106 genomes were used for this study, corresponding to 99 Sordariales and seven outgroup species. Of these, 52 genomes were newly sequenced as part of JGI community sequencing programs (proposals 504394, 662/300789 and GS014564). Three of these genomes were used as outgroups from within the class Sordariomycetes: *Lollioppaia minuta* (Diaporthales), *Eutypa lata* (Xylariales) and *Phialemonium atrogisum* (Cephalothecales). Four additional Leotiomycete outgroup genomes, the closest relative class to Sordariomycetes (Shen et al., 2020), were obtained from NCBI (GCA\_002803225.1, GCA\_900074895.1, GCA\_002812745.2, GCA\_001630605.1; National Center for Biotechnology Information, n.d.). Strain information and statistics on data source and quality are summarized in Supplementary Table S1.

### 2.2. Genome sequencing and assessment of genome assemblies

A variety of different methods were used to cultivate the fungi, extract DNA, sequence, assemble and annotate the 52 genomes (Supplementary methods, Methods datafile; see Table S1 for the respective methods for each genome). After assembly and annotation of all genomes, we assessed the quality of the genome assemblies using BUSCO V5.2.2 with the Augustus gene predictor for eukaryote runs (Manni et al., 2021a, 2021b; Stanke et al., 2006). Using BUSCO, each assembly's completeness was assessed on the basis of the presence or absence of a set of 3817 predefined orthologs from the sordariomycetes\_odb10 database (Kriventseva et al., 2019).

Potential contaminations of the assemblies were determined using the Blobtools2 workflow (Challis et al., 2020) as follows: first, the FASTA assembly files were used to create basic, per-sequence statistics (e.g. length, GC proportion). Next, contigs of each assembly were queried using BLAST v2.11.0+ (Altschul, 1997) with the blastn algorithm and the settings specified in the BlobToolKitPipeline ( $E = 1 \times 10^{-25}$ , max target sequences = 10, max hsp = 1, 16 threads). The BlobTools approach provides taxonomic annotation for each sequence in an

assembly. To avoid taxonomic inference for longer contigs being dependent on a single region, the script divides contigs longer than 100 kb into chunks before running blastn. By using the setting “-taxrule bestsumorder”, the taxonomic assignment is then based on the total bitscore obtained from a single database search with the NCBI taxonomy new\_taxdump database (date: 20220129; Challis et al., 2020; Schoch et al., 2020). Based on information given by blastn searches and BUSCO analysis, the BlobToolKit viewer was used to interactively explore each assembly for contamination.

### 2.3. Phylogenetic datamatrix

To construct the phylogenetic datamatrix, we used the set of 3817 single-copy, full-length BUSCO genes from the 99 representatives of Sordariales and the seven outgroups. We removed BUSCO genes whose taxon occupancy was  $\leq 50\%$  (i.e., when the gene was present in  $< 45$  Sordariales genomes). This removal resulted in a dataset with 3800 (i.e., 99.55 % of the original 3817) BUSCO genes.

To account for the underlying codon structure of our protein-coding nucleotide sequences, the combined sequences were aligned using an amino acid guided nucleotide alignment with MACSE V2.05 (Ranwez et al., 2018). The standard translation code was used (Elzanowski and Ostell, 2019). The nucleotide sequences were individually trimmed using TrimAl V1.4 with the default settings of the “gappy-out” parameter (Capella-Gutierrez et al., 2009). To validate our ortholog selection pipeline, a random subset of 40 gene alignments (selected using the bash command shuf -n 40) was visually examined, which revealed no ambiguously aligned sequences.

IQ-TREE (v2.1.4\_beta) was run on a single node with three logical cores for each of the 3800 BUSCO gene alignments. The options “-m TEST -runs 10” were used to find the best-fitting substitution model, using the best-scoring ML gene tree under 10 independent tree searches (Minh et al., 2020b). A concatenation-based tree was created with IQ-TREE on a single node with three logical cores under the GTR + F + I + G4 model, as 1582 of the 3800 genes favoured this as best fitting model (Suppl. Table S2). The support for each internal branch was evaluated using 1000 ultrafast bootstrap (UFBS) replicates, the seed was manually set to 33822. Next, we inferred a coalescence-based phylogeny using the set of 3800 individual ML gene trees and ASTRAL-III V5.7.8 (Zhang et al., 2018). The reliability of each internal branch was evaluated with local posterior probability (PP; default settings).

Subsequently, gene concordance factors (gCF; setting -gcf) and site concordance factors (sCF; setting -scf 100 -T 10) were estimated for each branch using IQ-TREE (Minh et al., 2020a). Here, gCF is defined as the percentage of gene trees containing that branch, while the sCF is defined as the percentage of decisive alignment sites supporting a branch in the reference tree (Lanfear, 2018; Minh et al., 2020a). The gCF and sCF concordance factors complement ultrafast bootstrap and posterior probability estimates by offering additional information about the underlying variability in the data (Minh et al., 2020a). All measures of support were evaluated and compared between the coalescence- and concatenation-based tree. Phylogenetic trees were visualized using the online iTOL tool (Letunic and Bork, 2021; V6.5.8).

### 2.4. Genomic properties

For a given genome, the length was estimated as the total number of base pairs in the genome assembly (total scaffold length). The number of genes was calculated by using the assemblies in nucleotide format and the gff annotation files, removing genes that contained introns that overlap with CDS (Suppl. Table S6) and subsequently using the script gag (Hall et al., 2014) to create fasta files of the coding sequences. Quality of gene number estimations was analysed with BUSCO completeness of the coding regions, using the sordariomycetes\_odb10 database and default settings. The evolutionary rate is given as a number of nucleotide substitutions per site, taken as a sum of path distances from

the base of the Sordariales to the tip on the concatenation-based tree. The repeat content of each genome was assessed with RepeatModeler (V1.0.8) and RepeatMasker (V4.1.0); by first building a database with the BuildDatabase function, followed by running RepeatModeler with default settings and subsequently running RepeatMasker with the settings -pa 10 -gccalc -gff (Smit et al., 2013; Smit and Hubley, 2008). Lastly, the extent of RIP in the genomes was assessed with the web-based tool RIPper with the option “RIP profile”. In this program, a sliding window approach is used with a 1000 bp window size and 500 bp slide size in order to assess the percentage of RIP affected regions in the genomes (van Wyk et al., 2019).

Taxon sampling was most dense across the *Chaetomiaceae*, *Podosporaceae* and *Sordariaceae*. Therefore, we performed comparative genomic analyses amongst these three groups. Tests of normality were performed for each genome property using SPSS V28, and the results showed that these were non-normally distributed within each group, and over the entire dataset. As such, statistical comparisons of trait values amongst these three families were done using Kruskal Wallis Tests for independent samples with SPSS V28 (IBM Corp, 2021). Genome properties were plotted using R version 4.2.1 in Rstudio, with package ggplot2 V3.3.6 (R Core Team, 2022; Rstudio Team, 2022; Wickham, 2016).

### 2.5. Assessment of phylogenetic signal and ancestral state reconstruction

Pearson’s correlation coefficient was used to test for correlations among the trait variables on the entire phylogeny without outgroups. To correct for phylogenetic dependence of species traits, the R package ape V5.6-2 (Paradis and Schliep, 2019) was used to compute phylogenetically independent contrasts (Felsenstein, 1985). We analysed phylogenetic signal of trait variables with four indices: Abouheif’s Cmean, Moran’s I, Blomberg’s K and Pagel’s  $\lambda$ , using R-package phylosignal V1.3 with the function phylosignal (Keck et al., 2016; Münkemüller et al., 2012). Both Abouheif’s Cmean and Moran’s I are developed in the context of spatial correlation and are not useable for effect size measure. However, stronger deviations from zero indicate stronger relationships between trait values and the phylogeny. The other two measures of phylogenetic signal, Blomberg’s K and Pagel’s  $\lambda$ , specifically relate to a Brownian motion model of trait evolution. For these indices, a value close to zero indicates phylogenetic independence and values closer to one indicate strong phylogenetic signal. These four indices give information about the general presence or absence of phylogenetic signal within the phylogeny. However, traits rarely evolve similarly across the phylogeny and phylogenetic signal can be scale dependent and vary among clades. Therefore, we computed local Moran’s I values with phylosignal V1.3 to detect hotspots of phylogenetic correlation (Anselin, 1996; Keck et al., 2016).

Lastly, we computed ancestral character states for the traits across internal nodes using the R package Phytools V0.6.44 with function ContMap (Revell, 2012). The mapping was done with default settings, which estimates states at internal nodes using ML with FastAnc, and then interpolates the states along each edge using equation [2] of Felsenstein (1985). The input tree was derived from the concatenation-based tree with branch lengths, using *P. atrogriusum* as outgroup.

## 3. Results and discussion

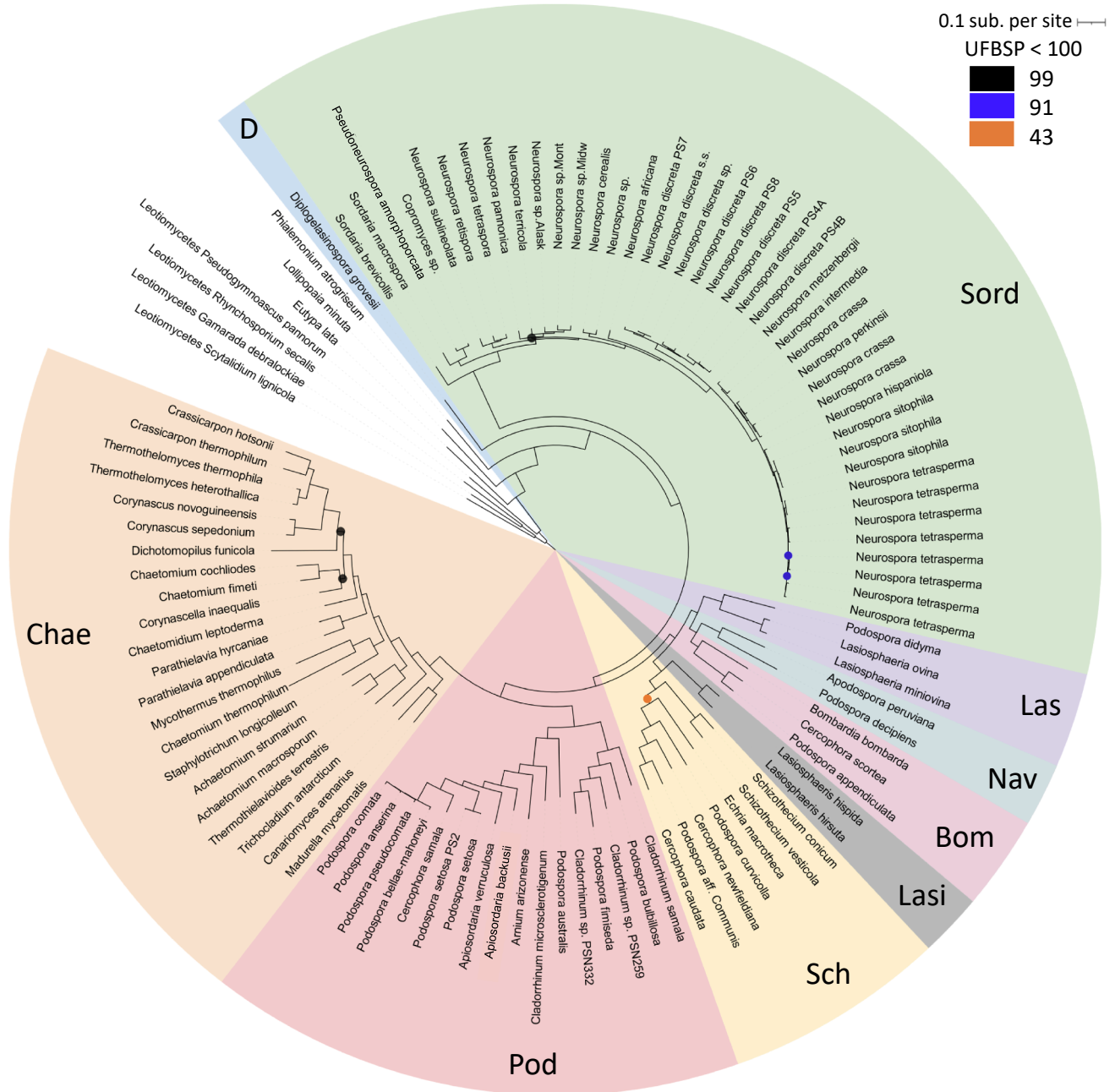
### 3.1. A genome-wide dataset of the order Sordariales

With this study we present a rich genomic resource for the order Sordariales. This dataset will prove valuable for further taxonomic investigations within the order, and provide a starting platform for comparative genomic analysis in this important group of filamentous Ascomycetes. Assembly quality statistics were obtained from the genomes and are summarized in Suppl. Table S1. Of the 106 whole genome assemblies, 100 genomes contained over 90 % of the BUSCO genes from

the sordariomycete\_odb10 dataset. In total, 17 out of the 3817 Sordariomycete BUSCO genes were found in less than 50 % of the Sordariales taxa and were deleted from the phylogenomic dataset.

BlobToolKit analysis was used to detect potential contamination in the genome assemblies. The vast majority of sequences showed no sign of contamination, but 17 genomes contained sequences of potential non-fungal origin. Manual exploration of the potentially contaminated sequences showed that, in 16 of the 17 genomes, regions without taxonomic assignment, or with best hits outside of the kingdom fungi, often showed high similarity to a broad range of taxa, including fungi. Here, we expect that the best mapping to other phyla was most likely due to the sequences being highly conserved rather than of non-fungal origin.

The total sequence length of these hits was < 0.5 % of the total genome length in all 16 genomes (Suppl. Table S1). In the seventeenth potentially contaminated genome, however, that of *Arniium arizonense*, 12.6 % of the genome assembly was classified as non-fungal. The majority of the blastn hits were to Firmicutes bacteria, without additional hits to the kingdom fungi. BUSCO completeness for the *A. arizonense* genome was >90 % for both genome assembly and coding region sequences, indicating good quality and estimates of gene numbers. Furthermore, all potential contamination in this genome was located outside of identified BUSCO genes. Therefore, *A. arizonense* was retained in the dataset, though we note that e.g. gene numbers may be influenced by contamination in this genome.



**Fig. 1.** Genome-wide maximum likelihood phylogeny of the fungal order sordariales. The concatenation-based ML phylogeny was inferred from 3800 single copy BUSCO genes found in over 50 % of the investigated Sordariales genomes. The tree is rooted at the Leotiomyces, and represents 99 Sordariales and seven outgroup strains. Clade colors represent different families of the order Sordariales, based on the described families by Huang et al (2021) and Marin-Felix and Miller (2022). UFBS values <100 % are shown. Note that low UFBS (<95 %) values were found on three occasions, all within-family. Family abbreviations: Sor: Sordariaceae; Las: Lasiosphaeriaceae; Nav: Naviculisporiaceae; Bom: Bombardiaceae; Lasi: Lasiosphaeridaceae; Sch: Schizotheciaceae; Pod: Podosporaceae; Chae: Chaetomiaceae.

### 3.2. A highly supported phylogeny of the order Sordariales

Phylogenetic inference using both concatenation- and coalescent-based approaches of the 3800 genes yielded a robust, well-resolved and comprehensive phylogeny of the Sordariales order (Fig. 1). The vast majority of internodes in both the concatenation-based and the coalescent-based phylogenies received strong ( $\geq 95\%$  UFBS and PP) support (98% of the nodes; three nodes without strong support in both trees). The three cases of low UFBS and PP support were all found on intrafamily branches. The majority of the nodes were congruent between the phylogenies inferred using both approaches, with 92 congruent internodes (92.4%; Fig. S1). There were no incongruences between the two inference methods amongst the branches leading to the nine families. The majority of nodes (70 nodes, 67%) received strong support from the majority of single gene trees (high gCF values). Only six out of 105 branches showed low sCF values, all low sCF values were found within families. ( $< 33\%$ ; Figs. S2-S6). The high interfamily gCF and sCF values indicate that the family phylogeny is well supported, and that only low levels of conflicting signals are present. The higher levels of gene tree discordance and site discordance were found within the genus *Neurospora* (Sordariaceae), within the Chaetomiaceae and amongst deep interfamily nodes in the phylogeny. Such discordance can have multiple biological sources, including incomplete lineage sorting and introgression. In places where branch lengths are extremely short, such as along the short branches within the genus *Neurospora*, either technical or biological processes can increase the discordance (Mallet et al., 2016; Minh et al., 2020a, 2020b).

Due to the high levels of support, our phylogeny has robustly resolved relationships that were only weakly supported in previous analyses based on smaller datasets. Our phylogeny strongly supported the division of Sordariales into at least nine clades corresponding to the families accepted by Huang et al. (2021). Four interfamily relationships were strongly supported with all measures of branch support (100% UFBS and PP, with the majority of single gene trees supporting the branching pattern). Congruent with Huang et al. (2021) our phylogeny consistently favoured placement of Chaetomiaceae and Podosporaceae as sister groups. In congruence with Marin-Felix and Miller (2022), our analysis confirms the grouping of Lasiosphaeridaceae and Schizotheciaceae as sister clades. The interfamily relationships among Bombardiaceae, Naviculisporiaceae and Lasiosphaeriaceae and the relationship between Sordariaceae and the rest of the tree were supported by high levels of both UFBS and PP (100% UFBS and PP).

Our phylogeny has furthermore cleared up intrafamily relationships within the Podosporaceae. Previous studies of this family have utilized four marker genes (rpb2, tub2, ITS and LSU) and divided the family into three main clades. These three main clades, characterized in our tree by *Podospora fimiseda*, *P. australis*, and *P. anserina* (Suppl. Fig. S2) are strongly supported (Ament-Velázquez et al., 2020; Wang et al., 2019; this study). However, in the four-marker phylogenies, the relationship of the clades was dependent on the variation in only one marker with strong phylogenetic signal (rpb2; Ament-Velázquez et al., 2020). With our genomic approach used herein, we were able to clarify the relationships amongst the clades. In short, our analysis confirmed the sister relationship of the clades containing *P. australis* and *P. anserina* (Suppl. Fig. S2) as initially described by Wang et al. (2019). The degree of conflict between markers for this internode does not seem to be dependent on a single gene (rpb2) with strong phylogenetic signal. Instead, we find an overall strong support for the branching relationships of all three clades, with 100% UFBS and PP. In total 84.84% of gene trees support the branching pattern between the two clades, indicating almost no conflicting signal in the underlying gene trees. This robust resolution of the relationships amongst the genera provides a better basis for discussions to minimize taxonomic conflict within Podosporaceae.

We note that the genome sequences of the taxa sampled are unevenly distributed across the families of Sordariales (specifically, e.g.

*Neurospora* spp. are over-represented compared to other taxa). Next to increasing sequence length, extensive taxon sampling is one of the most important determinants of accurate phylogenetic estimation (Heath et al., 2008). Previous studies of the order have focused on many taxa, but few loci, and produced highly conflicting results of the Sordariales phylogeny. While our dataset contained 99 Sordariales species and 3800 genes, future studies may add data from additional taxa that are under-represented in our study, and it is possible that the phylogeny may change as a result. This caveat notwithstanding, utilizing a genome-wide dataset has allowed us to establish a well-resolved phylogeny of the Sordariales. Between the concatenation-based and coalescence-based approaches only a few incongruences were found. All of the incongruences were found within families, mainly within species (complexes) and were accompanied by higher gene tree discordance. The vast majority of branches were highly supported with all phylogenetic reconstructions. Very few relationships were not fully resolved, with conflicting branches mainly present within one species or species complex.

### 3.3. Analysis of genomic properties

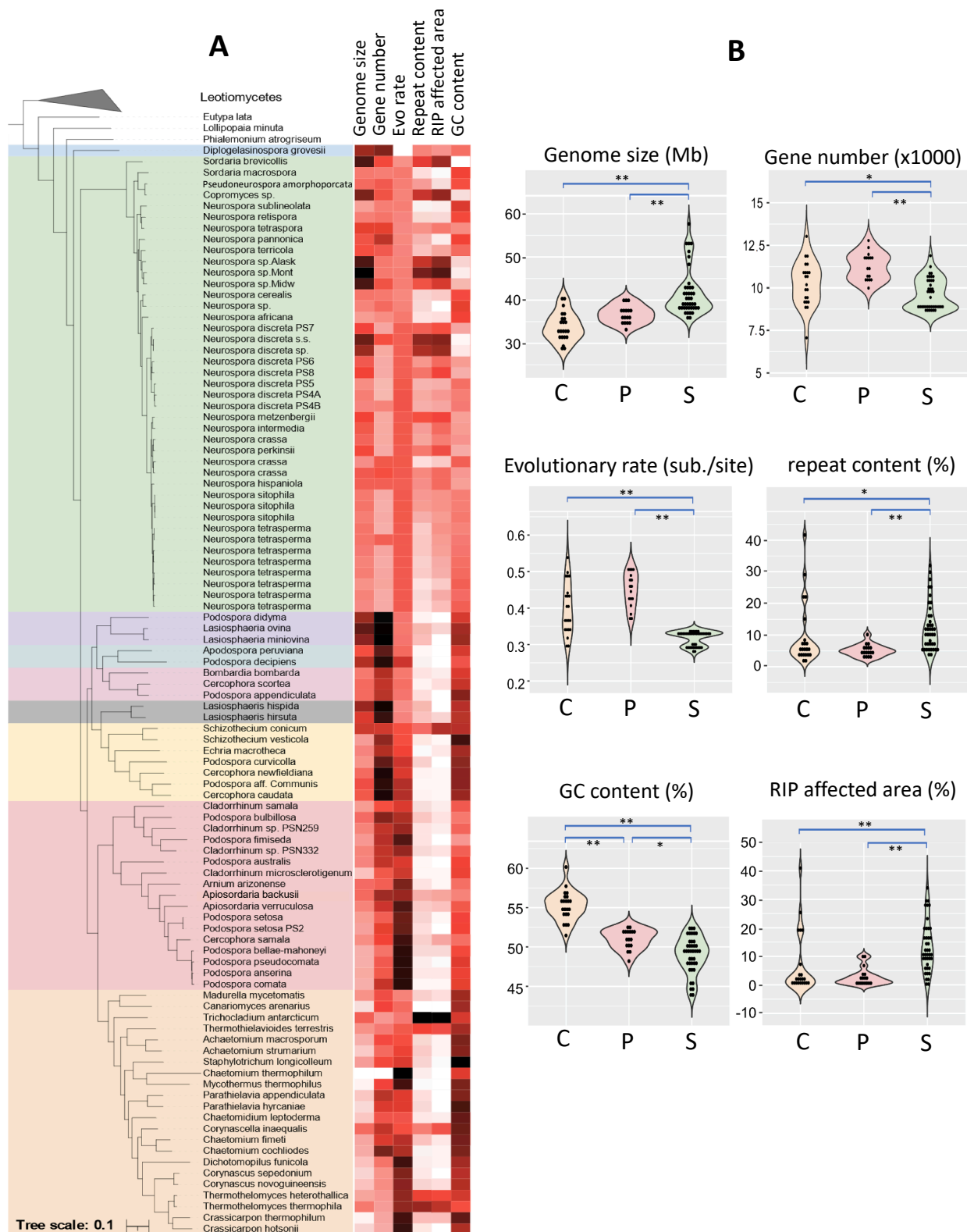
We performed comparative genomics for six genomic traits: genome size, gene number, GC content, evolutionary rate, repeat content and RIP-affected genome content. To this end, we first controlled for the possible effect of differences in genome assembly quality by looking at the correlation between N50 and each of the genomic properties. The only correlation we found was between N50 and gene number (Suppl. Table S4). Accordingly, to further assess the quality of the gene number analysis, we assessed BUSCO completeness of the coding regions for Sordariales genomes (Suppl. Table S1). From the coding sequences, 97 of the 101 examined genomes contained more than 90% of the complete single-copy BUSCO genes, indicating reliable gene number estimations. Only a few duplicated or fragmented BUSCO genes were found, indicating high quality of the genome assemblies in general.

In our comparative genomic analysis, we found a wide distribution of trait values among the investigated taxa (Fig. 2, specific data of each species is given in Suppl. Table S3). Genome size ranged from 28 to 58 Mb, gene numbers from 7066 to 14970, GC content from 44% to 60%, evolutionary rate from 0.27 to 0.63 substitutions per site. Repeat content ranged from 1.58% for *Canariomyces arenarius* to 41.64% in *T. antarcticum* (both members of Chaetomiaceae), while the RIP-affected genome content was estimated to range from below 1% in a number of taxa, to 40.96% in *Trichocladium antarcticum*.

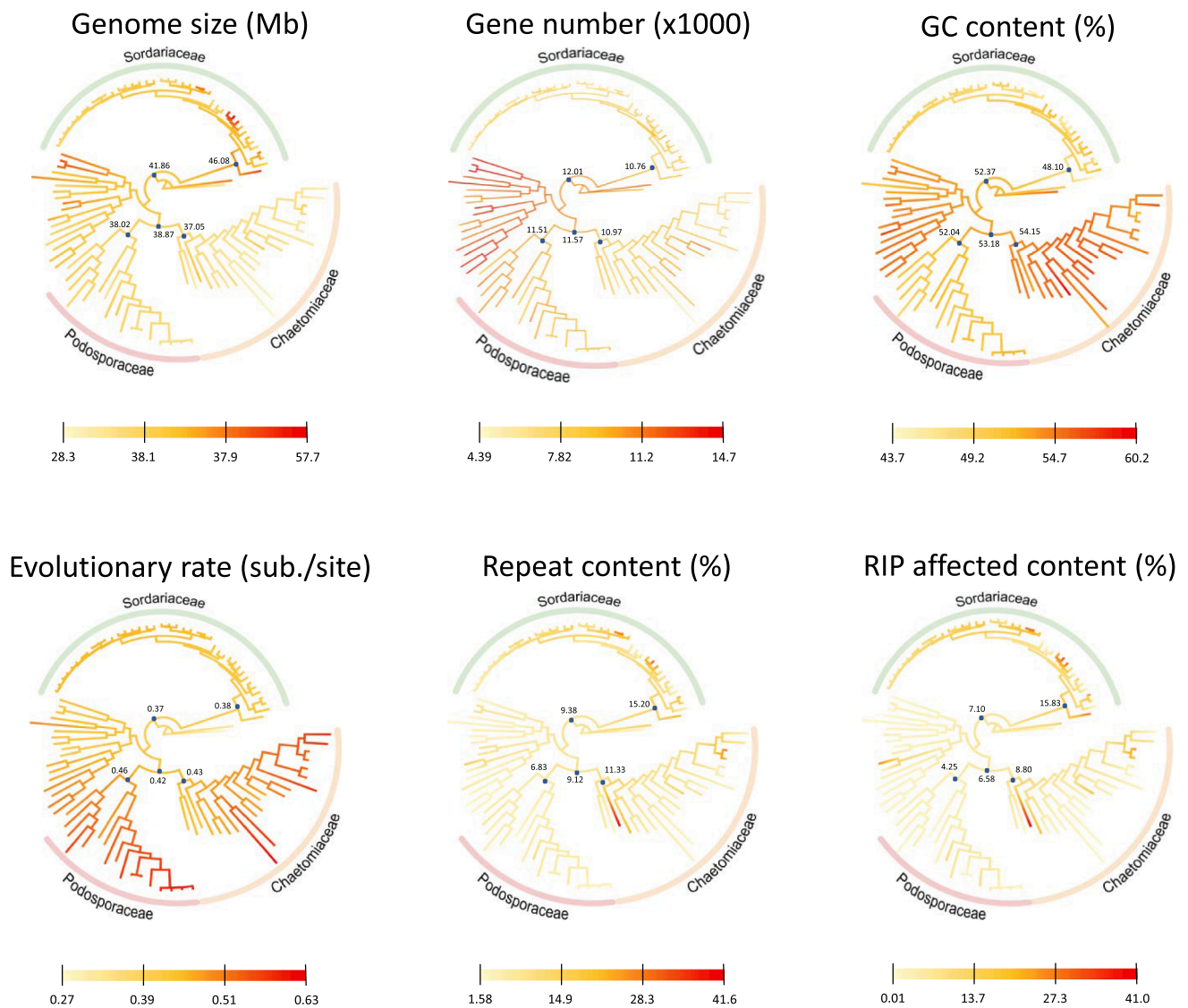
#### 3.3.1. Comparative genomics in Chaetomiaceae, Podosporaceae and Sordariaceae

Despite its large taxonomical diversity, genome sequencing in the order Sordariales has thus far focussed mainly on families containing well-known model-organisms (i.e. Sordariaceae, Podosporaceae), and Chaetomiaceae, whose members produce a diversity of secondary metabolites. As a result, a large amount of whole genome sequencing data is present for these three families, while only a few genomes have been fully sequenced in others. We examined the evolution of different genomic properties amongst the three most densely sampled groups: the Chaetomiaceae, Podosporaceae, and Sordariaceae. Evolution of genomic properties was analysed using non-parametric tests for independent samples (Fig. 2) and ancestral state reconstruction (Fig. 3). Comparisons of the trait values for the genome properties amongst the three families showed that all properties were variable.

All three families were found to be significantly different in GC content ( $p < 0.05$ ; Fig. 2C). For the other traits there was no significant difference amongst trait values between Podosporaceae and Chaetomiaceae. The Chaetomiaceae displayed the highest overall GC content and intermediate numbers of genes. They contained the smallest average genome sizes, with intermediate levels of repeat content, RIP affected genome content and evolutionary rates. The Podosporaceae had the



**Fig. 2. Overview of genomic properties.** a. Next to the maximum likelihood phylogeny of Sordariales, we plotted (from left to right) genome size, gene number, evolutionary rate (substitutions per site), repeat content (%), RIP affected area (%), and GC content (%). Continuous trait values were plotted using a heatmap from white (lowest value), to black (highest value). The detailed values of all six properties for each taxon are given in Table S3. Families from top to bottom: Sordariaceae. Lasiosphaeriaceae. Naviculisporiaceae. Bombardiaceae. Lasiosphaeriaceae. Schizotheciaceae. Podosporaceae. Chaetomiaceae. The tree was rooted with Leotiomyces. b. Individual graphs showing the distribution of genomic properties in the tree largest families of the concatenation-based tree. Families from left to right in the individual graphs: C: Chaetomiaceae (n = 22). P: Podosporaceae (n = 17). S: Sordariaceae (n = 41). \*difference is significant at p < 0.05; \*\* difference is significant at p < 0.01.



**Fig. 3. Evolution of genomic properties across the phylogeny.** The continuous properties were reconstructed on the species phylogeny and their ancestral states were visualized on the concatenation-based tree rooted with *Phialemonium atrogisum*. Heatmap bars denote ancestral state values from small (light yellow) to large (red). Five ancestral state values are shown. Two for the most recent common ancestor of the families Sordariaceae, Podosporaceae and Chaetomiaceae, and three for the last inferred ancestor of each of the families. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

highest gene numbers amongst the three families, high evolutionary rates and intermediate GC content. The genomes of this family further displayed relatively low levels of RIP- and repeat content.

The *Sordariaceae* differed significantly from the other two families in having the largest genome sizes, highest percentage of repeats and highest levels of RIP in their genomes. This family displayed intermediate gene numbers, while the GC content and evolutionary rates were significantly lower than those of *Podosporaceae* and *Chaetomiaceae*. Overall, the *Sordariaceae* are overrepresented in the dataset, with many closely related *Neurospora* strains having been sequenced. These closely related strains, often belonging to one single species, all display very similar genomic characteristics. As a result, the trends described above are valid for the current dataset, but it is possible that the inferences of trends of genomic properties change as species outside *Neurospora* are added to the *Sordariaceae*, or as taxon sampling becomes more even over all families.

Uneven taxon sampling can influence trends of genomic properties over multiple genomes. Additionally, on the level of single genomes, estimation of genomic properties can be influenced by assembly quality.

Repeat content and RIP content in particular are highly dependent on genome assembly quality (Testa et al., 2016; van Wyk et al., 2019, 2021). As an example in our dataset, in *Neurospora crassa*, previous reports indicate RIP affected genome content to be around 15 % (van Wyk et al., 2019). Two of our three *N. crassa* genomes display similar levels of RIP and repeat levels (14.24 % repeat content in neucra-01 and 11.30 % repeat content for neucra-03), but the third (neucra-02) displays only 4.81 % repeat content. While BUSCO analysis shows high completeness of all three assemblies (Suppl. Table S1), neucra-02 is missing about 3–4 Mb in assembly size compared to neucra-01 and neucra-03. One hypothesis is that the genome size of neucra-02 has been streamlined to contain lower levels of repetitive elements and RIP. However, since repetitive elements represent a prevalent part of gaps in genome assemblies (Peona et al., 2018, 2021; Sotero-Caio et al., 2017; Tørresen et al., 2019; Weissensteiner and Suh, 2019), we are currently unable to determine whether this is an artefact of genome assembly or a genomic trait of neucra-02. Thus, future research should focus on sampling more taxa, as well as ensuring high quality data. This will reduce the chance of false trends and genomic assembly artefacts in genome analysis. In



particular, analysis of a wider variety of *N. crassa* genomes is needed to determine whether the lower levels of repetitive elements are a genomic trait of the neucra-02 strain or an assembly artefact.

3.3.2. Variation of genomic properties stems primarily from within-family evolution

Analysis of standard Pearson’s correlations amongst genomic properties revealed that half of the correlations were different before (i.e., standard Pearson’s correlations) and after (i.e., phylogenetically independent contrasts) accounting for phylogeny (Fig. 4). After accounting for phylogeny, negative correlations remained between GC content and genome size ( $p < 0.01$ ), GC- and repeat- content ( $p < 0.05$ ), and GC- and RIP affected genome content ( $p < 0.05$ ). RIP affected genome content was furthermore positively correlated with repeat content ( $p < 0.01$ ) and genome size ( $p < 0.01$ ). Lastly, bimodality of GC content was positively correlated with genome size ( $p < 0.05$ ), and negatively correlated with evolutionary rate ( $p < 0.05$ ) and RIP affected genome content ( $p < 0.01$ ). All correlations had  $r$  values below 0.6, indicating relatively weak correlation between the traits (Fig. 4). The other correlations found before accounting for phylogeny became insignificant at the  $p < 0.05$  level, indicating that phylogenetic signal, rather than trait correlation, could be the cause of the significant correlation values. We quantified phylogenetic signal to determine how trait variation was correlated with the phylogenetic relatedness of species (Suppl. Table S5). All four indices (Blomberg’s  $K$ , Pagel’s  $\lambda$ , Abouheif’s  $C_{mean}$ , and Moran’s  $I$ ) indicated significant phylogenetic signal in all traits. Pagel’s  $\lambda$  ranged from 0.299 for RIP affected genome content to 1.01 for evolutionary rate. Here, lower values indicate less phylogenetic signal in the trait and a value of  $\lambda = 1$  indicates that there is strong phylogenetic signal and the variation observed in the trait does not deviate from expectations under a Brownian motion model of evolution. The results indicate that genomic trait values are phylogenetically linked in the Sordariales, but the fraction of trait variation explained by the phylogeny is variable. Local Moran’s  $I$  tests further showed that phylogenetic signal in traits is concentrated within various families, indicating that variation in genomic properties stems from within-family evolution of

properties. The local autocorrelation tests showed that the signal is strongest in the *Chaetomiaceae* family, followed by the *Sordariaceae*. The *Podosporaceae* showed less signals of local autocorrelation for most of the traits, but significant positive autocorrelation was found for evolutionary rate for most species.

The genome-scale phylogeny was then used to infer the ancestral character states and reconstruct the evolution for each property. Values for genomic traits at the last common ancestors to *Sordariaceae*, *Podosporaceae*, and *Chaetomiaceae* were inferred to be quite similar (Fig. 3). For example, the inferred state values of GC content for the *Sordariaceae* last common ancestor and the last common ancestor of both *Podosporaceae* and *Chaetomiaceae* are 52.17 % and 53.18 % respectively, while there is a significant difference in average GC content amongst the three groups (Fig. 2c). The same trend is also observed for the other traits. This pattern further supports the hypothesis -similar to the phylogenetic signal- that the differences in genomic properties amongst *Chaetomiaceae*, *Podosporaceae*, and *Sordariaceae* stem from evolution after divergence of the three families.

3.3.3. Potential causes of divergence of genomic properties

In *Chaetomiaceae*, *Podosporaceae* and *Sordariaceae*, several traits covary. Phylogenetic signal and ancestral state reconstruction analysis indicated that variation in genomic properties has stemmed from within-family evolution of properties. For each of the three families, we examined potential causes of divergence of genomic properties. The majority of *Chaetomiaceae* showed a trend of high GC content, small genome size and predominantly low levels of repeats. One possible explanation for these trends in *Chaetomiaceae* could be the thermophilic lifestyle of at least part of the species in the dataset. Our current dataset contains ten species with known optimal growth temperatures, of which seven have been classified as thermophilic (van den Brink et al., 2015). Earlier research into prokaryotic thermophiles indicated that their genomes are GC rich and streamlined, most likely caused by GC-rich codons encoding for more thermostable amino acids (Wu et al., 2012; Hu et al., 2022) and optimization of energy utilization in stressful environments by reduction of genome size (Giovannoni et al., 2014). All

		Genome size	Number of genes	Evolutionary rate	GC content	Repeat content	RIP affected genome content
Phylogenetically independent contrasts	Genome size		0.291**	-0.582**	-0.566**	0.578**	0.577**
	Number of genes	0.105		-0.068	0.386**	-0.289**	-0.386**
	Evolutionary rate	-0.03	-0.092		0.248*	-0.327**	-0.339**
	GC content	-0.306**	0.092	-0.125		-0.529**	-0.637**
	Repeat content	0.439**	0.046	-0.152	-0.244*		0.940**
	RIP affected genome content	0.379**	-0.028	-0.093	-0.318*	0.592**	

Fig. 4. Genomic trait correlations. Pairwise standard Pearson’s correlation coefficients were conducted before (i.e., standard Pearson’s correlations; upper diagonal) and after accounting for phylogeny (i.e., phylogenetically independent contrasts; lower diagonal). For each cell, the value corresponds to Pearson’s coefficient value. Presence or absence of star signs indicate significance, NS:  $P > 0.05$ ; \*:  $P \leq 0.05$ ; \*\*:  $P \leq 0.01$ . Orange cells denote instances where correlation trends are significant after accounting for phylogeny. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

known thermophilic species in our dataset showed the general pattern of high GC content and small genome sizes. This result further supports earlier findings that high temperature environments may lead to similar changes in fungi and prokaryotes (van Noort et al., 2013). Deviations from the overall genomic pattern were seen in some species, which could be related to lower optimal growth temperatures. Indeed, the non-thermophilic species *T. antarcticum* contained the highest level of repeats in the dataset, relatively low GC content, and large genome size. Since the optimal growth temperature is unknown for the majority of *Chaetomiaceae* species in the dataset, future research in the field of thermophilia and how it affects genome evolution in this group of fungi will be an interesting endeavour to pursue.

On average, the *Podosporaceae* clade showed the highest evolutionary rate. In other lineages, such as both animals and plants, evolutionary rate is negatively associated with generation time (Thomas et al., 2010; Welch et al., 2008; Yue et al., 2010). Assuming equal mutation rates per generation, species with rapid sexual reproduction replicate their genomes more frequently, leading to a higher number of mutations per time unit. Although generation time is unknown for many fungi, the coprophilic lifestyle of the *Podosporaceae* is believed to select for rapid sexual reproduction rather than prolonged mycelial maintenance (see e.g. Geydan et al., 2012). Thus, the high evolutionary rates in *Podosporaceae* may stem at least partially from the rapid generation time observed in this group of fungi.

The *Sordariaceae* contained the largest number of genomes in our dataset. They had the largest average genome size of the three main families, with low GC content and gene numbers, and a wide variety of repeat contents indicating widespread presence of RIP. RIP is known to have profound impact on genome evolution (Borkovich et al., 2004; Clutterbuck, 2011; Galagan and Selker, 2004; Gladyshev, 2017) and could thus explain the trends seen in genomic traits. For example, research across the Ascomycota showed that RIP affected genomes have low GC levels, and genome size and RIP mutation content are moderately correlated (van Wyk et al., 2021). Relatively low gene numbers can furthermore be caused by the apparent lack of functional gene duplication, with RIP slowing the creation of new genes (Galagan et al., 2003). High levels of RIP could thus explain the overall trend of low GC content and gene number as well as the large genome size in the *Sordariaceae*.

#### 4. Concluding remarks

Even though recent phylogenetic studies have substantially advanced our understanding of Sordariales evolution, the absence of whole genome data caused uncertainties to remain in the phylogeny (Ament-Velázquez et al., 2020; Marin-Felix and Miller, 2022). In this study, 99 Sordariales genomes were used to create a whole-genome phylogeny and research genome evolution in the order, of which 52 newly sequenced as part of JGI community science programs. High UFSB and PP support values, together with high interfamily gCF and sCF values indicate that the family relationships are well supported, despite uneven taxon sampling. The few low intrafamily gCF and sCF values indicate that intrafamily conflict remains and it is possible that both the inference of the species phylogeny and the trends of genomic properties within families will change as more taxa are added. Addition of new taxa is especially important for underrepresented groups in the dataset, as taxonomic conflicts often remain in underrepresented families (Marin-Felix and Miller, 2022). Despite these caveats, using genome-scale information to infer the phylogeny of the order enabled us to resolve several controversies surrounding the phylogenetic relationships of the families within the order and to test the robustness of the inference of several contentious branches. For example, our study robustly supported *Chaetomiaceae* as a sister group to *Podosporaceae*. Within the *Podosporaceae* our study resolved the ambiguities surrounding the sister relationship of the clades containing *P. australis* and *P. anserina*.

The well-supported phylogeny was next used as a framework for

comparative genomic analysis in the Sordariales. The three largest families in our dataset (*Chaetomiaceae*, *Podosporaceae*, and *Sordariaceae*) showed clear trends in most of the investigated genomic properties. Local autocorrelation and ancestral state reconstruction of the traits further revealed that the variation of the properties stems primarily from within-family evolution. We hypothesize that some of the trends in genomic properties seen in the families *Chaetomiaceae*, *Podosporaceae*, and *Sordariaceae* may be caused by ecological, life-history, and molecular traits. Unfortunately, many ecological traits are analysed only amongst a subset of species. It therefore remains difficult to determine whether the described trends are caused by ecological and/or life history traits or if they are a signal of phylogeny alone. Future addition of genomes of underrepresented clades, together with research into the relation amongst genomic properties and ecological traits, will provide more knowledge about the key-drivers of genome evolution in this group of fungi.

#### CRediT authorship contribution statement

**Noah Hensen:** Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Lucas Bonometti:** Writing – review & editing, Data curation. **Ivar Westerberg:** Data curation, Writing – review & editing. **Ioana Onut Brännström:** Methodology, Writing – review & editing. **Sonia Guillou:** Software, Data curation. **Sandrine Cros-Aarteil:** Resources, Data curation. **Sara Calhoun:** Resources, Data curation. **Sajeet Haridas:** Resources, Data curation. **Alan Kuo:** Resources, Data curation. **Stephen Mondo:** Resources, Data curation. **Jasmyn Pangilinan:** Resources, Data curation. **Robert Riley:** Resources, Data curation. **Kurt LaButti:** Resources, Data curation. **Bill Andreopoulos:** Resources, Data curation. **Anna Lipzen:** Software, Data curation. **Cindy Chen:** Software, Data curation. **Mi Yan:** Resources, Data curation. **Chris Daum:** Resources, Data curation. **Vivian Ng:** Resources, Data curation. **Robin A. Ohm:** Resources, Data curation. **Francis Martin:** Resources, Data curation. **Philippe Silar:** Resources, Data curation, Writing – review & editing. **Donald O. Natvig:** Resources, Data curation, Writing – review & editing. **Christophe Lalanne:** Resources, Data curation. **Valérie Gautier:** Resources, Data curation. **Sandra Lorena Ament-Velázquez:** Resources, Data curation, Writing – review & editing. **Ása Krüys:** Resources, Data curation, Writing – review & editing. **Miriam I. Hutchinson:** Resources, Data curation. **Amy Jo Powell:** Resources, Data curation. **Kerrie Barry:** Resources, Data curation. **Andrew N. Miller:** Resources, Data curation, Writing – review & editing. **Igor V. Grigoriev:** Resources, Data curation, Writing – review & editing. **Robert Debuchy:** Resources, Data curation, Writing – review & editing. **Pierre Gladieux:** Resources, Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing. **Markus Hiltunen Thorén:** Conceptualization, Supervision, Methodology, Writing - review & editing. **Hanna Johannesson:** Resources, Conceptualization, Funding acquisition, Supervision, Writing – review & editing, Project administration.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

We thank Petter Madsen and Aaron Vogan for useful discussions and constructive feedback. Sabine Huhndorf is thanked for providing strains. Sandia National Laboratories is a multimission laboratory managed and operated by the National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security

Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the U.S. Government.

**Data and materials:** The computations/data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2018-05973. Assemblies were produced through JGI community sequencing programs, with the additional four outgroup genomes publicly available through National Center for Biotechnology Information.

**Funding sources:** This work was funded by JGI's Community Science Program grant #504394 and Marie Curie postdoctoral fellowship FP7-PEOPLE-2010-IOF-No.273086 to Pierre Gladieux. The work (proposal: 10.46936/10.25585/60000837, 10.46936/10.25585/60001060 and 10.46936/10.25585/60001199) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. We further acknowledge funding from The Bergianus Foundation of The Royal Swedish Academy of Sciences.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymprev.2023.107938>.

## References

- Altschul, S., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25 (17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
- Ament-Velásquez, L.S., Johannesson, H., Giraud, T., Debuchy, R., Saupe, S.J., Debets, A. J.M., Bastiaans, E., Malagnac, F., Grognet, P., Peraza-Reyes, L., Gladieux, P., Krusys, A., Silar, P., Huhndorf, S.M., Miller, A.N., Vogan, A.A., 2020. The taxonomy of the model filamentous fungus *Podospora anserina*. *Mycospora* 75, 51–69. <https://doi.org/10.3897/mycokeys.75.55968>.
- Anselin, L., 1996. The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In: *Spatial Analytical Perspectives on GIS*. Routledge.
- Borkovich, K.A., Alex, L.A., Yarden, O., Freitag, M., Turner, G.E., Read, N.D., Seiler, S., Bell-Pedersen, D., Paietta, J., Plesofsky, N., Plamann, M., Goodrich-Tanrikulu, M., Schulte, U., Mannhaupt, G., Nargang, F.E., Radford, A., Selitrennikoff, C., Galagan, J. E., Dunlap, J.C., Pratt, R., 2004. Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. *Microbiol. Mol. Biol. Rev.* 68 (1), 1–108. <https://doi.org/10.1128/MMBR.68.1.1-108.2004>.
- Cai, L., Jeewon, R., Hyde, K.D., 2006. Phylogenetic investigations of *Sordariaceae* based on multiple gene sequences and morphology. *Mycol. Res.* 110 (2), 137–150. <https://doi.org/10.1016/j.mycres.2005.09.014>.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25 (15), 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
- Chadefaud, M., Emberger, L., 1963. *Traité De Botanique Systématique* 32 (4), 125–127.
- Challis, R., Richards, E., Rajan, J., Cochrane, G., Blaxter, M., 2020. BlobToolKit – interactive quality assessment of genome assemblies. *G3: Genes|Genomes|Genetics* 10 (4), 1361–1374. <https://doi.org/10.1534/g3.119.400908>.
- Charria-Girón, E., Surup, F., Marin-Felix, Y., 2022. Diversity of biologically active secondary metabolites in the ascomycete order Sordariales. *Mycol. Prog.* 21 (4), 43. <https://doi.org/10.1007/s11557-022-01775-3>.
- Clutterbuck, J.A., 2011. Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. *Fungal Genet. Biol.* 48 (3), 306–326. <https://doi.org/10.1016/j.fgb.2010.09.002>.
- IBM Corp, 2021. IBM SPSS Statistics for Windows, Version 28.0. IBM Corp.
- Davis, R.H., Perkins, D.D., 2002. *Neurospora*: a model of model microbes. *Nat. Rev. Genet.* 3 (5), 397–403. <https://doi.org/10.1038/nrg797>.
- Elzanowski, A., Ostell, J., 2019. *The Genetic Codes*. National Center for Biotechnology Information (NCBI).
- Felsenstein, J., 1985. Phylogenies and the comparative method. *Am. Nat.* 125 (1), 1–15.
- Fouché, S., Oggenfuss, U., Chanclud, E., Croll, D., 2022. A devil's bargain with transposable elements in plant pathogens. *Trends Genet.* 38 (3), 222–230. <https://doi.org/10.1016/j.tig.2021.08.005>.
- Galagan, J.E., Selker, E.U., 2004. RIP: the evolutionary cost of genome defense. *Trends Genet.* 20 (9), 417–423. <https://doi.org/10.1016/j.tig.2004.07.007>.
- Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., FitzHugh, W., Ma, L.-J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S., Nielsen, C.B., Butler, J., Endrizzi, M., Qui, D., Ianakiev, P., Birren, B., 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422 (6934), 859–868. <https://doi.org/10.1038/nature01554>.
- Geydan, T.D., Debets, A.J.M., Verkley, G.J.M., Van Diepingen, A.D., 2012. Correlated evolution of senescence and ephemeral substrate use in the Sordariomycetes: Correlated evolution senescence and ecology. *Mol. Ecol.* 21 (11), 2816–2828. <https://doi.org/10.1111/j.1365-294X.2012.05569.x>.
- Giovannoni, S.J., Cameron Thrash, J., Temperton, B., 2014. Implications of streamlining theory for microbial ecology. *ISME J.* 8 (8), 1553–1565. <https://doi.org/10.1038/ismej.2014.60>.
- Gladieux, P., De Bellis, F., Hann-Soden, C., Svedberg, J., Johannesson, H., Taylor, J.W., 2020. *Neurospora* from natural populations: population genomics insights into the life history of a model microbial eukaryote. In: Duthell, J.Y. (Ed.), *Statistical Population Genomics*. Springer, US, vol. 2090, pp. 313–336. 10.1007/978-1-0716-0199-0\_13.
- Gladyshev, E., 2017. Repeat-induced point mutation and other genome defense mechanisms in fungi. *Microbiol. Spect.* 5 (4), 5.4.02. <https://doi.org/10.1128/microbiolspec.FUNK-0042-2017>.
- Graña, F., Lespinet, O., Rimbault, B., Dequard-Chablat, M., Coppin, E., Picard, M., 2001. Genome quality control: RIP (repeat-induced point mutation) comes to *Podospora*: RIP (repeat-induced point mutation) in *Podospora*. *Mol. Microbiol.* 40 (3), 586–595. <https://doi.org/10.1046/j.1365-2958.2001.02367.x>.
- Gregory, T.R., 2005. Synergy between sequence and size in Large-scale genomics. *Nat. Rev. Genet.* 6 (9), 699–708. <https://doi.org/10.1038/nrg1674>.
- Griffiths, A.J.F., 1992. Fungal senescence. *Annu. Rev. Genet.* 26 (1), 351–372. <https://doi.org/10.1146/annurev.ge.26.1.20192.002031>.
- Hall, B., DeRego, T., Geib, S., 2014. GAG: the Genome Annotation Generator (version 1.0). <http://genomeannotation.github.io/GAG>.
- Hawksworth, D.L., Eriksson, O.E., 1986. The names of accepted orders of Ascomycetes (5 (1), 75–184) [Systema Ascomycetum].
- Heath, T.A., Hedtke, S.M., Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. *J. Syst. Evol.* 46 (3), 239.
- Hu, E.-Z., Lan, X.-R., Liu, Z.-L., Gao, J., Niu, D.-K., 2022. A positive correlation between GC content and growth temperature in prokaryotes. <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-022-08353-7>.
- Huang, S.-K., Hyde, K.D., Mapook, A., Maharachchikumbura, S.S.N., Bhat, J.D., McKenzie, E.H.C., Jeewon, R., Wen, T.-C., 2021. Taxonomic studies of some often over-looked Diaporthomycetidae and Sordariomycetidae. *Fungal Divers.* 111 (1), 443–572. <https://doi.org/10.1007/s13225-021-00488-4>.
- Huhndorf, S.M., Miller, A.N., Fernández, F.A., 2004. Molecular systematics of the sordariales: the order and the family lasiosphaeriaceae redefined. *Mycologia* 96 (2), 368–387. <https://doi.org/10.2307/3762068>.
- Hutchinson, M.I., Powell, A.J., Herrera, J., Natvig, D.O., 2019. New Perspectives on the Distribution and Roles of Thermophilic Fungi. In: Tiquia-Arashi, S.M., Grube, M. (Eds.), *Fungi in Extreme Environments: Ecological Role and Biotechnological Significance*. Springer International Publishing, pp. 59–80. [https://doi.org/10.1007/978-3-030-19030-9\\_4](https://doi.org/10.1007/978-3-030-19030-9_4).
- Hyde, K., 2020. Refined families of Sordariomycetes. *Mycosphere* 11 (1), 305–1059. <https://doi.org/10.5943/mycosphere/11/1/7>.
- Keck, F., Rimet, F., Bouchez, A., Franc, A., 2016. phyloSignal: an R package to measure, test, and explore the phylogenetic signal. *Ecol. Evol.* 6 (9), 2774–2780. <https://doi.org/10.1002/ece3.2051>.
- Kriventseva, E.V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F.A., Zdobnov, E.M., 2019. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* 47 (D1), D807–D811. <https://doi.org/10.1093/nar/gky1053>.
- Krusys, A., Huhndorf, S.M., Miller, A.N., 2015. Coprophilous contributions to the phylogeny of Lasiosphaeriaceae and allied taxa within Sordariales (Ascomycota, Fungi). *Fungal Divers.* 70 (1), 101–113. <https://doi.org/10.1007/s13225-014-0296-3>.
- Lanfear, R., 2018. Calculating and interpreting gene- and site-concordance factors in phylogenomics. [http://www.robertlanfear.com/blog/files/concordance\\_factors.html](http://www.robertlanfear.com/blog/files/concordance_factors.html).
- Letunic, I., Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49 (W1), W293–W296. <https://doi.org/10.1093/nar/gkab301>.
- Li, X.-Q., Du, D., 2014. Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. *PLoS One* 9 (2), e88339.
- Mallet, J., Besansky, N., Hahn, M.W., 2016. How reticulated are species? *Bioessays* 38 (2), 140–149. <https://doi.org/10.1002/bies.201500149>.
- Manni, M., Berkeley, M.R., Seppy, M., Simão, F.A., Zdobnov, E.M., 2021a. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38 (10), 4647–4654. <https://doi.org/10.1093/molbev/msab199>.
- Manni, M., Berkeley, M.R., Seppy, M., Zdobnov, E.M., 2021b. BUSCO: assessing genomic data quality and beyond. *Curr. Protoc.* 1 (12) <https://doi.org/10.1002/cpz1.323>.
- Marin-Felix, Y., Miller, A.N., 2022. Corrections to recent changes in the taxonomy of the Sordariales. *Mycol. Prog.* 21 (8), 69. <https://doi.org/10.1007/s11557-022-01814-z>.
- Marin-Felix, Y., Miller, A.N., Cano-Lira, J.F., Guarro, J., García, D., Stadler, M., Huhndorf, S.M., Stchigel, A.M., 2020. Re-evaluation of the Order Sordariales: Diplogelasinosporaceae, Naviculisporaceae, and delimitation of lasiosphaeriaceae s. str., and introduction of the new families Schizotheciaceae. *Microorganisms* 8 (9), E1430. <https://doi.org/10.3390/microorganisms8091430>.
- Mat Razali, N., Cheah, B.H., Nadarajah, K., 2019. Transposable elements adaptive role in genome plasticity, pathogenicity and evolution in fungal phytopathogens. *Int. J. Mol. Sci.* 20 (14), 3597. <https://doi.org/10.3390/ijms20143597>.

- Miller, A.N., Huhndorf, S.M., 2005. Multi-gene phylogenies indicate ascomal wall morphology is a better predictor of phylogenetic relationships than ascospore morphology in the Sordariales (Ascomycota, Fungi). *Mol. Phylogenet. Evol.* 35 (1), 60–75. <https://doi.org/10.1016/j.ympev.2005.01.007>.
- Minh, B.Q., Hahn, M.W., Lanfear, R., 2020a. New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol.* 37 (9), 2727–2733. <https://doi.org/10.1093/molbev/msaa106>.
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020b. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37 (5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffrers, K., Thuiller, W., 2012. How to measure and test phylogenetic signal: How to measure and test phylogenetic signal. *Methods Ecol. Evol.* 3 (4), 743–756. <https://doi.org/10.1111/j.2041-210X.2012.00196.x>.
- National Center for Biotechnology Information. (n.d.). Retrieved 7 October 2022, from <https://www.ncbi.nlm.nih.gov/>.
- Noumeur, S.R., Teponno, R.B., Helaly, S.E., Wang, X.-W., Harzallah, D., Houbraken, J., Crous, P.W., Stadler, M., 2020. Diketopiperazines from *Batnamyces globularicola*, gen. & sp. Nov. (Chaetomiaceae), a fungus associated with roots of the medicinal plant *Globularia alypum* in Algeria. *Mycol. Prog.* 19 (6), 589–603. <https://doi.org/10.1007/s11557-020-01581-9>.
- Oggenfuss, U., Badet, T., Wicker, T., Hartmann, F.E., Singh, N.K., Abraham, L., Karisto, P., Vonlanthen, T., Mundt, C., McDonald, B.A., Croll, D., 2021. A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. *Elife* 10, e69249.
- Paradis, E., Schliep, K., 2019. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35 (3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>.
- Patel, H., Rawat, S., 2021. Thermophilic fungi: Diversity, physiology, genetics, and applications. In: *New and Future Developments in Microbial Biotechnology and Bioengineering*. Elsevier, pp. 69–93. <https://doi.org/10.1016/B978-0-12-821005-5.00005-3>.
- Peona, V., Weissensteiner, M.H., Suh, A., 2018. How complete are “complete” genome assemblies?—An avian perspective. *Mol. Ecol. Resour.* 18 (6), 1188–1195. <https://doi.org/10.1111/1755-0998.12933>.
- Peona, V., Blom, M.P.K., Xu, L., Burri, R., Sullivan, S., Bunikis, I., Liachko, I., Haryoko, T., Jønsson, K.A., Zhou, Q., Irestedt, M., Suh, A., 2021. Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Mol. Ecol. Resour.* 21 (1), 263–286. <https://doi.org/10.1111/1755-0998.13252>.
- R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ranwez, V., Douzery, E.J.P., Cambon, C., Chantret, N., Delsuc, F., 2018. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol. Biol. Evol.* 35 (10), 2582–2584. <https://doi.org/10.1093/molbev/msy159>.
- Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things): phytools: R package. *Methods Ecol. Evol.* 3 (2), 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>.
- Roche, C.M., Loros, J.J., McCluskey, K., Glass, N.L., 2014. *Neurospora crassa*: looking back and looking forward at a model microbe. *Am. J. Bot.* 101 (12), 2022–2035. <https://doi.org/10.3732/ajb.1400377>.
- Rstudio Team, 2022. RStudio: Integrated Development for R. RStudio [RStudio]. PBS. <http://www.rstudio.com/>.
- Scheckhuber, C.Q., Osiewacz, H.D., 2008. *Podospora anserina*: a model organism to study mechanisms of healthy ageing. *Mol. Genet. Genomics* 280 (5), 365. <https://doi.org/10.1007/s00438-008-0378-6>.
- Schoch, C.L., Ciuffo, S., Domrachev, M., Hottot, C.L., Kannan, S., Khovanskaya, R., Leipe, D., McVeigh, R., O'Neill, K., Robertse, B., Sharma, S., Soussov, V., Sullivan, J.P., Sun, L., Turner, S., Karsch-Mizrachi, I., 2020. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database*, 2020, baaa062. [10.1093/database/baaa062](https://doi.org/10.1093/database/baaa062).
- Shen, X.-X., Steenwyk, J.L., LaBella, A.L., Opulente, D.A., Zhou, X., Kominek, J., Li, Y., Groenewald, M., Hittinger, C.T., Rokas, A., 2020. Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Sci. Adv.* 6 (45), eabd0079. <https://doi.org/10.1126/sciadv.abd0079>.
- Silar, P., 2020. *Podospora anserina*, 1st ed. HAL science ouverte.
- Smit, A., Hubley, R., 2008. RepeatModeler Open-1.0 (Open 1.0). <http://www.repeatmasker.org/>.
- Smit, A., Hubley, R., Green, P., 2013. RepeatMasker Open-4.0.
- Sotero-Caio, C.G., Platt, R.N., Suh, A., Ray, D.A., 2017. Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* 9 (1), 161–177. <https://doi.org/10.1093/gbe/evw264>.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., Morgenstern, B., 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucl. Acids Res.* 34(Web Server issue), W435–W439. [10.1093/nar/gkl200](https://doi.org/10.1093/nar/gkl200).
- Testa, A.C., Oliver, R.P., Hane, J.K., 2016. OcculterCut: a comprehensive Survey of AT-Genome Biol. *Evol.* 8 (6), 2044–2064. <https://doi.org/10.1093/gbe/evw121>.
- Thomas, J.A., Welch, J.J., Lanfear, R., Bromham, L., 2010. A generation time effect on the rate of molecular evolution in invertebrates. *Mol. Biol. Evol.* 27 (5), 1173–1180. <https://doi.org/10.1093/molbev/msq009>.
- Tørresen, O.K., Star, B., Mier, P., Andrade-Navarro, M.A., Bateman, A., Jarnot, P., Gruca, A., Grynberg, M., Kajava, A.V., Promponas, V.J., Anisimova, M., Jakobsen, K. S., Linke, D., 2019. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Res.* 47 (21), 10994–11006. <https://doi.org/10.1093/nar/gkz841>.
- van den Brink, J., Facun, K., de Vries, M., Stielow, J.B., 2015. Thermophilic growth and enzymatic thermostability are polyphyletic traits within *Chaetomiaceae*. *Fungal Biol.* 119 (12), 1255–1266. <https://doi.org/10.1016/j.funbio.2015.09.011>.
- van Noort, V., Bradatsch, B., Arumugam, M., Amlacher, S., Bange, G., Creevey, C., Falk, S., Mende, D.R., Sinning, I., Hurt, E., Bork, P., 2013. Consistent mutational paths predict eukaryotic thermostability. *BMC Evol. Biol.* 13 (1), 7. <https://doi.org/10.1186/1471-2148-13-7>.
- van Wyk, S., Harrison, C.H., Wingfield, B.D., De Vos, L., van der Merwe, N.A., Steenkamp, E.T., 2019. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. *PeerJ* 7, e7447.
- van Wyk, S., Wingfield, B.D., De Vos, L., van der Merwe, N.A., Steenkamp, E.T., 2021. Genome-wide analyses of repeat-induced point mutations in the ascomycota. *Front. Microbiol.* 11, 622368. <https://doi.org/10.3389/fmicb.2020.622368>.
- Vogan, A.A., Miller, A.N., Silar, P., 2021. (2803) Proposal to change the conserved type of *Podospora*, nom. Cons. (Ascomycota). *TAXON* 70 (2), 429–430. <https://doi.org/10.1002/tax.12478>.
- Wang, X.W., Bai, F.Y., Bensch, K., Meijer, M., Sun, B.D., Han, Y.F., Crous, P.W., Samson, R.A., Yang, F.Y., Houbraken, J., 2019. Phylogenetic re-evaluation of *Thielavia* with the introduction of a new family *Podosporaceae*. *Stud. Mycol.* 93, 155–252. <https://doi.org/10.1016/j.simyco.2019.08.002>.
- Weissensteiner, M.H., Suh, A., 2019. Repetitive DNA: The Dark Matter of Avian Genomics. In: Kraus, R.H.S. (Ed.), *Avian Genomics in Ecology and Evolution*. Springer International Publishing, pp. 93–150. [https://doi.org/10.1007/978-3-030-16477-5\\_5](https://doi.org/10.1007/978-3-030-16477-5_5).
- Welch, J.J., Bininda-Emonds, O.R., Bromham, L., 2008. Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol. Biol.* 8 (1), 53. <https://doi.org/10.1186/1471-2148-8-53>.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed. Springer International Publishing : Imprint: Springer. [10.1007/978-3-319-24277-4](https://doi.org/10.1007/978-3-319-24277-4).
- Wu, H., Zhang, Z., Hu, S., Yu, J., 2012. On the molecular mechanism of GC content variation among eubacterial genomes. *Biol. Direct* 7 (1), 2. <https://doi.org/10.1186/1745-6150-7-2>.
- Yue, J.-X., Li, J., Wang, D., Araki, H., Tian, D., Yang, S., 2010. Genome-wide investigation reveals high evolutionary rates in annual model plants. *BMC Plant Biol.* 10 (1), 242. <https://doi.org/10.1186/1471-2229-10-242>.
- Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinf.* 19 (6), 153. <https://doi.org/10.1186/s12859-018-2129-y>.