A haplotype-resolved chromosomal reference genome for the porcini mushroom *Boletus edulis*

Etienne Brejon Lamartinière ⬤ ,[1] Keaton Tremble,[2] Bryn T.M. Dentinger,[3,4] Kanchon K. Dasmahapatra,[5]
Joseph I. Hoffman[1,6,7,8,]*

[1]Faculty of Biology, Department of Evolutionary Population Genetics, Bielefeld University, 33501 Bielefeld, Germany
[2]Department of Biology, Duke University, Durham, NC 27707, USA
[3]School of Biological Sciences, The University of Utah, Salt Lake City, UT 84112, USA
[4]Natural History Museum of Utah, Salt Lake City, UT 84108, USA
[5]Department of Biology, University of York, Heslington, York YO10 5DD, UK
[6]Faculty of Biology, Center for Biotechnology (CeBiTec), Bielefeld University, 33615 Bielefeld, Germany
[7]Joint Institute for Individualisation in a Changing Environment, Bielefeld University and University of Münster, 33501 Bielefeld, Germany
[8]British Antarctic Survey, UKRI-NERC, Cambridge CB3 OET, UK

*Corresponding author: Faculty of Biology, Department of Evolutionary Population Genetics, Bielefeld University, 33501 Bielefeld, Germany; Faculty of Biology, Center for
Biotechnology (CeBiTec), Bielefeld University, 33615 Bielefeld, Germany; Joint Institute for Individualisation in a Changing Environment, Bielefeld University and University of
Münster, 33501 Bielefeld, Germany; British Antarctic Survey, UKRI-NERC, Cambridge CB3 OET, United Kingdom. Email: joseph.hoffman@uni-bielefeld.de

Haplotype-resolved chromosomal reference genomes are increasingly available for many fungi, offering insights into the evolution of
pathogenic and symbiotic lifestyles. However, these resources remain scarce for ectomycorrhizal fungi, which play crucial roles in forest
ecosystems. Here, we used a combination of chromatin conformation capture and PacBio sequencing to construct a haplotype-resolved
chromosomal genome assembly for *Boletus edulis*, a prized edible fungus and emerging model for ectomycorrhizal fungal research. Our
new reference assembly, "BolEdBiel_h2," derives from a *B. edulis* sporocarp sampled in Bielefeld, Germany. The genome assembly
spans 41.8 Mb, with a scaffold N50 of 4.1 Mb, and includes 11 chromosome-level scaffolds, achieving near telomere-to-telomere cover-
age across multiple chromosomes. We annotated a total of 15,406 genes, with a Benchmarking Universal Single-Copy Orthologs score
of 96.2%. Key genomic features such as mating loci, carbohydrate-active enzymes, and effector proteins, were identified. As a first ap-
plication of this new genomic resource, we mapped whole-genome resequencing data from 53 genets to investigate the population
structure and genetic diversity of the European lineage of *B. edulis*. We identified 2 distinct genetic clusters and found that high-latitude
populations from Iceland and Fennoscandia exhibited greater nucleotide diversity than populations from the United Kingdom and
Central Europe. Additionally, we discovered a 0.4-Mb inversion on chromosome 3 and identified several regions of locally elevated nu-
cleotide diversity, which may represent candidates for ecological adaptation. This genomic resource will facilitate a deeper understand-
ing of this ecologically and commercially important wild fungus.

Keywords: *Boletus edulis*; ectomycorrhizal fungi (EMF); reference genome assembly; Hi-C; population structure

Introduction

Ectomycorrhizal fungi (EMF) play a crucial role in the functioning
of forest ecosystems worldwide. They facilitate nutrient cycling
(Read and Perez-Moreno 2003), contribute to carbon sequestration
(Anthony *et al.* 2024), and enhance the growth, immunity, and
pathogen resistance of nearly 60% of all trees (Steidinger *et al.*
2019). However, despite their critical roles in forest ecosystems,
our understanding of the ecology and evolution of EMF remains
limited. This knowledge gap arises partly from the difficulty of
studying these predominantly subterranean organisms, which
typically cannot be cultured in the laboratory beyond the myce-
lium stage due to their symbiotic lifestyles.

Over the past 3 decades, molecular genetic approaches have
been pivotal in expanding our understanding of EMF, facilitating
both ecological and evolutionary research (Douhan *et al.* 2011).
Classical genetic markers such as randomly amplified

polymorphic DNAs, amplified fragment length polymorphisms,
and simple sequence repeats allowed researchers to distinguish
individual genotypes in natural settings, paving the way for stud-
ies of clonality and genetic diversity (Bonello *et al.* 1998; Amend
et al. 2009). More recently, advancements in high-throughput se-
quencing technologies have broadened the scope of EMF research,
providing detailed insights into population structure (Branco *et al.*
2016), gene flow (Tremble, Hoffman, *et al.* 2023), gene content
(Kohler *et al.* 2015), adaptation (Bazzicalupo *et al.* 2020), and mo-
lecular evolution (Looney *et al.* 2022).

A key requirement for modern population genomic studies is
the availability of high-quality, annotated, chromosome-level ref-
erence genomes. These resources enable gene discovery (Mariene
and Wasmuth 2025) and the characterization of patterns of
variation across the genome, including structural variants (SVs)
(Amarasinghe *et al.* 2020), runs of homozygosity (Brejon
Lamartinière *et al.* 2024), and recombination landscapes

(Weissensteiner *et al.* 2017). New sequencing approaches such as Hi-C (high-throughput chromosome conformation capture; Belton *et al.* 2012) have been instrumental in improving the contiguity of genome assemblies by allowing sequencing reads to be assembled into phased haplotypes. These techniques have already been used to generate chromosomal reference genomes for several cultured and domesticated fungi (Morin *et al.* 2012; Engel *et al.* 2013; Yu *et al.* 2022; Ma *et al.* 2023). However, to date, chromosomal reference genomes have only been published for 2 EMF, *Tricholoma matsutake* and *Suillus bovinus* (Kurokochi *et al.* 2023; Zhang *et al.* 2024).

Generating and analyzing long contiguous DNA sequences may represent the next frontier in genomic studies of EMF, as multiple recent studies of cultured and domesticated fungi have uncovered genomic structural variation both among and within dikaryons that may play a role in fungal evolution. For example, Sperschneider *et al.* (2023) used phased chromosomal assemblies of the arbuscular mycorrhizal fungus *Rhizophagus irregularis* to demonstrate that separate heterokaryon haplotypes are distinct functional and regulatory units that can independently modulate the expression of host plant genes. Similarly, Borgognone *et al.* (2018) showed that DNA methylation in the saprotrophic genus *Pleurotus* can be haplotype-specific and tends to be higher around transposable elements (TEs), where it reduces potentially maladaptive gene expression. As many EMF also have a heterokaryotic life stage, haplotype-resolved chromosomal EMF reference genomes are likely to become important resources for understanding the biology of this important group of fungi.

Boletus edulis Bull., known variously as the king Bolete, Penny Bun, cèpe de Bordeaux, Steinpilz or porcino, is one of the most charismatic and economically important EMF species worldwide. While the majority of commonly found EMF associate with just 1 or 2 host plant genera (Voller *et al.* 2024), *B. edulis* forms mutualistic associations with diverse plant genera, from the most dominant forest trees of the northern hemisphere (*Fagus*, *Quercus*, *Pinus*, *Picea*, *Betula*, *Castanea*, and *Pseudotsuga*) to alpine miniature shrubs (Treindl and Leuchtmann 2019). It also has a broad geographical distribution spanning Eurasia and North America. For these reasons, *B. edulis* has recently emerged as a promising model system for studying the ecology and evolution of EMF (Hoffman *et al.* 2020; Tremble *et al.* 2020; Tremble, Brejon Lamartinière, *et al.* 2023; Tremble, Hoffman, *et al.* 2023; Brejon Lamartinière *et al.* 2024).

The first global-scale population genomic study of *B. edulis* found evidence for six distinct lineages that diverged from one another between 1.6 and 2.7 million years ago (Tremble, Hoffman, *et al.* 2023). Reference genomes have already been generated for all of these lineages (Tremble, Brejon Lamartinière, *et al.* 2023), but they vary considerably in contiguity and completeness, and none of them are haplotype-resolved or assembled to the chromosomal level. The most contiguous of these reference genomes, which comprises a total of 38 scaffolds with an N50 of 2.5Mbp, is available for the Alaska lineage, while the least contiguous reference genome, which comprises 488 scaffolds with an N50 of 0.17 Mbp, is available for the European lineage. Moreover, these reference genomes were generated using haplotype-unaware methods, which limits the detection of SVs, especially in highly repetitive genomes (Ebert *et al.* 2021).

Within the *B. edulis* complex, the European lineage exhibits the widest geographical and ecological distribution, ranging from Mediterranean grasslands to the Scandinavian tundra. This lineage is also associated with the greatest diversity of host species (MyCoPortal, http://www.mycoportal.org/portal/index.php), which

is reflected by its expanded symbiosis-related gene repertoire (Tremble, Brejon Lamartinière, *et al.* 2023). Developing a chromosomal reference genome for this lineage would allow for a more in-depth exploration of population structure, local adaptation, and host specialization, not only across Eurasia, but also on a broader scale. This is particularly important in the context of climate change, as many forests of the northern hemisphere are being strongly impacted by warmer and drier conditions (Gazol *et al.* 2017). A deeper understanding of forest adaptation is essential and this must include EMF, which play a vital role in helping trees cope with climate-induced stress.

In this study, we combined highly accurate long-read Pacbio HiFi sequencing with Hi-C to generate a haplotype-resolved chromosomal genome assembly for the European *B. edulis* lineage. We predicted key genomic features including mating loci, carbohydrate-active enzymes (CAZymes), TEs, and overall gene content. Additionally, we investigated structural and gene copy variation within the reference individual and explored longer-term patterns of synteny with *S. bovinus* (Zhang *et al.* 2024). To further contextualize the reference genome, we mapped short-read data from 53 European samples to explore patterns of population genetic structure and diversity.

Materials and methods

Sporocarp tissue sampling

A total of 15 g of tissue from the inner cap flesh of a *B. edulis* sporocarp was collected from a *Fagus* woodland in Bielefeld, Germany. It was sequentially frozen at 4°C for 30 min and then at −20°C for 30 min, before being stored at −80°C. DNA isolation, library preparation, sequencing, and genome component prediction were performed by Biomarker Technologies GmbH as described below. In addition, tissue from the same individual was dried and archived in the Senckenberg Museum of Natural History, Görlitz, under the reference GLM-F139661.

Genomic DNA isolation and sequencing

DNA was extracted from the sporocarp tissue using a QIAGEN Genomic-tip 20G kit. To generate a highly contiguous assembly, long-read sequencing was conducted as follows. Libraries were constructed according to PacBio standard protocol and sequenced on a PacBio Sequel II platform by BMKgene. The resulting raw circular consensus sequences were quality-filtered using smrtlink v12 with the parameters –min-passes 5 –min-rq 0.9, assembled into scaffolds using Hifiasm v0.12 (Cheng *et al.* 2021), and corrected using Pilon v1.17 (Walker *et al.* 2014).

Hi-C libraries were constructed using the Illumina mate-pair kit and sequenced on an Illumina NovaSeq X plus platform. The resulting paired-end Hi-C reads were filtered using HiC-Pro v2.10.0 (Servant *et al.* 2015), separated into haplotypes through integration with long reads using Hifiasm v0.12, and aligned to the long-read assembly using bwa v0.7.10 (Li and Durbin 2009). The contigs were then clustered, ordered, and oriented using Lachesis (Burton *et al.* 2013). To visualize chromatin interactions and genome contiguity, we generated a chromatin contact heat map using the command-line version of Juicer v1.6 (Durand *et al.* 2016).

Preliminary genome annotation

A repeat database was constructed using a combination of LTR_FINDER v1.05 (Xu and Wang 2007), MITE-Hunter (Han and Wessler 2010), RepeatScout v1.0.5 (Price *et al.* 2005), and PILER-DF v2.4 (Edgar and Myers 2005). The database was sorted

with PASTEClassifier (Hoede *et al.* 2014) and merged with the Repbase database (Bao *et al.* 2015). RepeatMasker v4.0.6 (Smit et al. (2013-2015)) was then used to predict repeat elements in the fungal genome based on this combined database. Gene prediction was performed through both de novo and homology-based approaches. The de novo prediction utilized Genscan (Burge and Karlin 1997), Augustus v2.4 (Stanke *et al.* 2004), GlimmerHMM v3.0.4 (Majoros *et al.* 2004), GeneID v1.4 (Alioto *et al.* 2018), and SNAP (version 2006-07-28) (Johnson *et al.* 2008). For homology-based prediction, GeMoMa v1.3.1 (Keilwagen *et al.* 2019) was used to identify homologous protein-coding genes. The results from both approaches were integrated using EvidenceModeler v1.1.1 (Haas *et al.* 2008). Noncoding RNA was predicted using tRNAscan-SE -version 2.0 (Chan and Lowe 2019) for tRNA and Infernal 1.1 (Nawrocki and Eddy 2013) for other noncoding RNAs. Pseudogenes were identified by aligning homologous genes from the predicted protein list and the Swiss-Prot database (Bairoch and Apweiler 2000) to the genome using GenBlastA v1.0.4 (She *et al.* 2009), followed by the detection of early termination and frameshift mutations with GeneWise (Birney *et al.* 2004). Gene clusters were identified using antiSMASH v6.0.0 (Medema *et al.* 2011). CAZymes encoding genes were annotated using hmmer -v3.4 (Finn *et al.* 2011) based on the CAZy database (Cantarel *et al.* 2009). Loci encoding effector proteins were predicted using EffectorP v3.0 (Sperschneider *et al.* 2016). Approximate centromere locations were identified visually based on Hi-C interactions. AT content at these locations was inspected in R using 1-kb sliding windows.

Haplotype comparisons and evolutionary context

To investigate the quality of the alternative haplotypes, we identified the positions of conserved Benchmarking Universal Single-Copy Orthologs (BUSCO) genes from the basidiomycota_obd10 database using BUSCO v5.7.1 (Simão *et al.* 2015) separately for each haplotype. We also ran gene and TE prediction on each haplotype using Funnanotate v1.8.17 (Palmer and Stajich (2023)) and EDTA v2.2.0 (Ou *et al.* 2019), respectively. To evaluate mapping coverage, we aligned sequencing reads from 53 European *B. edulis* samples to both haplotypes using the mem2 algorithm from bwa (https://github.com/bwa-mem2/bwa-mem2) with the default parameters. These sequencing data included previously published Illumina MiSeq, HiSeq, and NovaSeq data from 49 samples (Tremble, Brejon Lamartinière, *et al.* 2023) as well as 4 additional samples from Bielefeld, Germany, that were 150 bp PE sequenced to 30× coverage on a BGI T7 platform. To compute the average depth of coverage per position, we used a 500-kbp sliding window with a 50-kbp step, combining the VCFtools command –site-mean-depth (Danecek *et al.* 2011) and the GenomicRanges Grange function in R (Lawrence *et al.* 2013).

After characterizing and comparing the 2 haplotypes, as described in the results, we selected haplotype 2 for all subsequent analyses. The mating locus MATa identified by (Tremble, Brejon Lamartinière, *et al.* 2023) and the annotated STE3 genes from the MATb locus in the *B. edulis* bed1 genome accessed from JGI MycoCosm (Miyauchi *et al.* 2020) were located in the reference genome using BLAST v2.2.31 (Altschul *et al.* 1990). We also checked for the presence of telomeric sequences and located them in the genome using the Telomere Identification Toolkit TIDK (Brown *et al.* 2023). Finally, we analyzed patterns of synteny between the *B. edulis* reference genome and the chromosomal assembly of *S. bovinus* (Zhang *et al.* 2024) through the identification and mapping of putatively identical Basidiomycota BUSCO genes.

European population structure and genetic diversity

As a first application of this newly generated resource, we analyzed patterns of population genetic structure and genome-wide diversity within the European *B. edulis* lineage. After mapping short-read data from 53 genets to haplotype 2 of the reference genome following the workflow described above, variants were called using a 3-step approach implemented in GATK v4.4.0 (McKenna *et al.* 2010). First, variants were called with HaplotypeCaller, then, the files were aggregated into a single database using GenomicsDBImport, and finally, an all-site unfiltered multi-sample VCF file was generated with GenotypeGVCF. This file was then filtered for single-nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) ≥ 5%, mapping quality ≥ 15, missing genotypes ≤ 20%, a minimum depth of coverage of 5, and a maximum depth of coverage of 100 using VCFtools. We then computed a principal component analysis (PCA) of the data using pcadapt (Luu *et al.* 2017) on the filtered VCF. In addition, we calculated nucleotide diversity (π) in windows of 10 kb with a step size of 1 kb across the genome. This analysis was conducted separately for individuals from different geographical regions with a sample size greater than 5 using the VCFtools –window-pi and –window-pi-step commands. The all-site VCF included invariant loci, which were pruned as described previously, while only excluding the MAF filter.

Results and discussion
## Genome assembly quality

The PacBio-generated scaffolds achieved a 100% Hi-C anchoring rate for both haplotypes, resulting in 2 chromosome-level assemblies consisting of 11 pseudo-chromosomes each ("BolEdBiel_h1" and "BolEdBiel_h2," respectively, Table 1, Fig. 1a, and Supplementary Fig. 1 and 2; see Supplementary Tables 1 and 2 for details). The high quality of these assemblies is evident from their excellent chromosome contiguity, low numbers of gaps per chromosome (Table 2), and high BUSCO scores, which exceeded 96.6% for both haplotypes, with the score being slightly higher for haplotype 2. Genome sizes for the 2 haplotypes were 53 and 41 Mbp, respectively, with both exhibiting a GC content of 53.7%. However, the scaffold N50 was slightly lower for haplotype 1 (3.8 Mbp compared with 4.1 Mbp for haplotype 2). This discrepancy is partly due to the fact that over 15% of the total length of haplotype 1 comprised unanchored scaffolds. When mapping short-read sequencing data from 53 European individuals to this haplotype, coverage for all but one unplaced scaffold was very low (mean = 2.2 vs 25.9, respectively, Fig. 1b), reflecting their high content of repetitive DNA.

We found a strong positive association between chromosome length and gene number ($z = 30.39$, $P < 0.001$, Fig. 1c). However, this relationship did not hold when comparing homologous chromosomes. In most cases, haplotype 1 was longer than haplotype 2 (Fig. 1a and d), yet there were no significant differences between homologous chromosomes in terms of gene count (Fig. 1d) and the abundance of TEs (Fig. 1e). Based on these results and the slightly better quality metrics for haplotype 2, including higher scaffold N50 and BUSCO scores, along with the low coverage of haplotype 1-specific contigs, we selected BolEdBiel_h2 as the preferred reference genome for the European lineage of *B. edulis*. Nonetheless, both haplotypes have excellent quality metrics and, given that there is no linkage across chromosomes, the choice of haplotype is somewhat arbitrary. Therefore, both assemblies could potentially be used for

Table 1. Quality metrics for the *B. edulis* haplotype assemblies.

Haplotype	Length (bp)	Chromosomes	Scaffolds	Contigs	Contig N50 (bp)	Scaffold N50 (bp)	n BUSCOs (%)	%GC content	Total length in transposable elements	Total length in repetitive DNA
BolEdBiel_h1	53,135,892	11	311	321	3,363,843	3,798,680	C: 1,704 (96.6%) S: 1,630 (92.4%) D: 74 (4.2%) F: 19 (1.1%) M: 41 (2.3%)	53.7	9,434,133	3,330,036
BolEdBiel_h2	41,815,337	11	73	90	2,952,305	4,104,162	C: 1,709 (96.8%) S: 1,643 (93.1%) D: 66 (3.7%) F: 17 (1%) M: 38 (2.2%)	53.7	8,893,489	3,536,600

C, complete; S, complete and single copy; D, complete and duplicated; F, fragmented; M, missing.
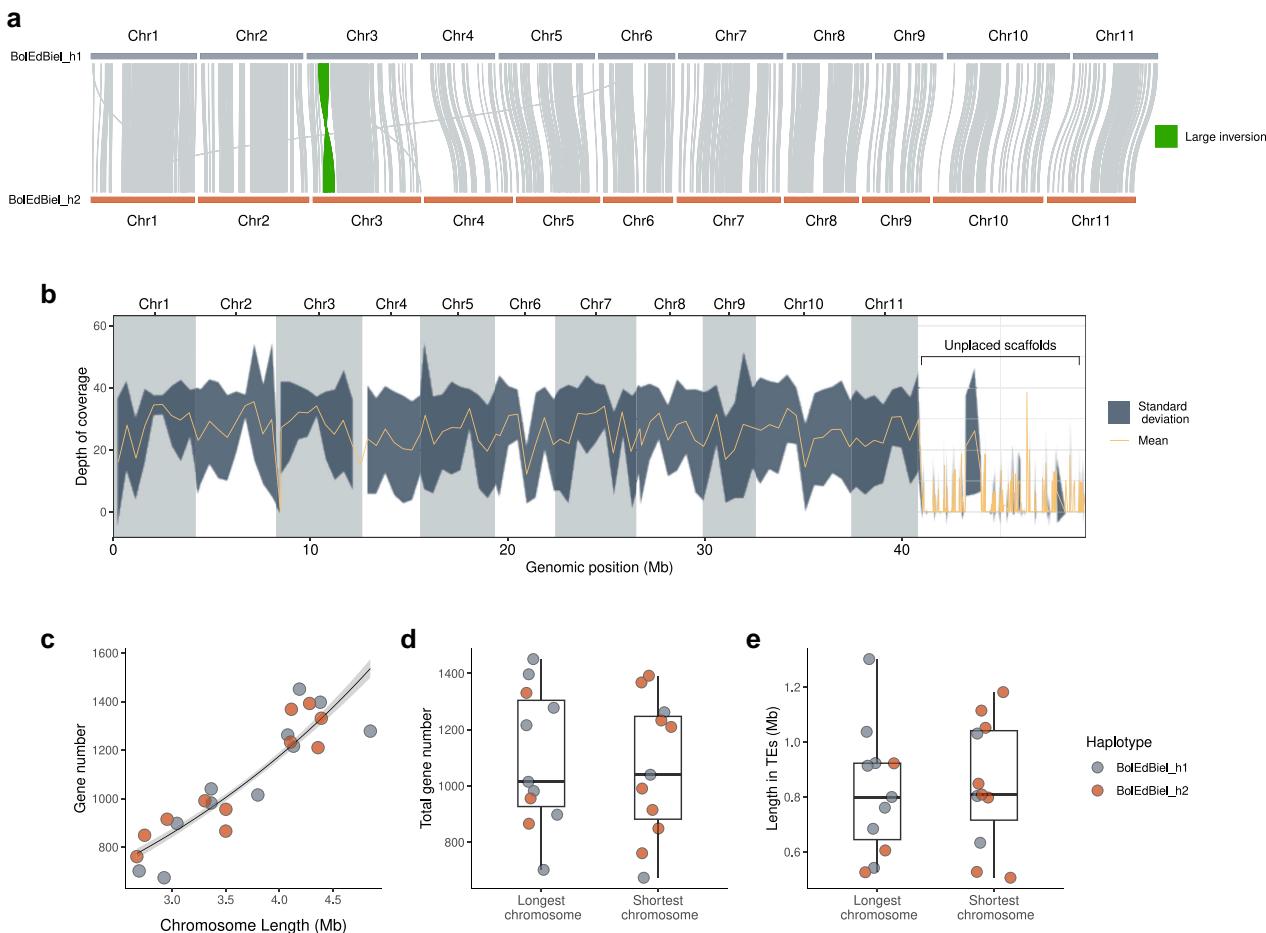
Fig. 1. Comparison of the 2 *B. edulis* haplotype assemblies. a) Patterns of synteny between the 2 haplotypes of each pair of homologous chromosomes. To facilitate visualization, we reversed and complemented the alignments of chromosomes 5, 6, 7, and 9 from haplotype 1 and highlighted the inversion on chromosome 3 in green. b) Depth of mapping coverage of whole-genome resequencing data from 53 genets from the European *B. edulis* linage to haplotype 1. The orange line represents the mean mapping coverage, with the dark gray shaded region representing the 95% confidence interval. The chromosomes are represented by alternating gray and white blocks. c) The relationship between chromosome length and gene number. The black line represents the fit of a Poisson regression with the gray shaded region representing the SE. d) and e) show comparisons of gene number counts and TE content between the longest and shortest haplotypes, respectively. In a), c), d), and e), haplotypes 1 and 2 are color-coded in gray and orange, respectively.

research in this species, particularly for pan-genomic or structural variation analyses.

The synteny analysis also uncovered a 0.4-Mbp inversion on chromosome 3 (Fig. 1a). We confirmed this inversion by evaluating PacBio sequence mapping rates using the integrative genomics viewer (Thorvaldsdóttir *et al.* 2012). Inversions can play important roles in evolution by limiting or suppressing local recombination, which helps to maintain genetic diversity and adaptive potential, and can also drive population divergence and speciation (Hoffmann and Rieseberg 2008). In extreme cases, inversions have been found to induce shifts from mutualism to pathogenesis (Somvanshi *et al.* 2012). Consequently, further research into this

SV might help to illuminate the genetic mechanisms underlying the host adaptability and ecological flexibility of *B. edulis*.

Genome content

Focusing on BolEdBiel_h2, we de novo predicted a total of 15,406 genes with an average length of 1.7 kb, which together account for ~53.4% of the total genome length (41.8 Mbp). Additionally, we identified 11,890 repetitive sequences, mostly comprising long terminal repeat retrotransposons (Supplementary Fig. 3), which represented 35.4% of the genome length in chromosomes. Ten telomeric repeats, (TTAGG)*n*, were identified in 8 of the 11 chromosomes. Near telomere-to-telomere assemblies were achieved for chromosomes 5 and 10, with the former being gapless (Fig. 2 and Table 2), which is a rarity in EMF. Telomeres, and chromosome ends in general, have been linked to several important biological processes. For example, in the fungal pathogen *Pyricularia*, telomere-adjacent regions are enriched for genes involved in host adaptation (Rahnama *et al.* 2021). Similar

patterns might be expected in EMF. In addition, the centromeres were identified based on their elevated interaction frequencies within the 3-dimensional chromatin architecture (Mizuguchi *et al.* 2014) which was visible in the Hi-C contact data (Supplementary Figs. 1 and 2). We found that, for most of the chromosomes, the centromere region exhibited high AT content relative to the rest of the chromosome (Supplementary Fig. 2). A similar pattern has been reported in other fungi, including the yeasts *Candida glabrata* and *Saccharomyces cerevisiae, both of which possess centromeric DNA elements with an AT content exceeding 85%* (Roy and Sanyal 2011).

We sought to identify the genomic locations of genes encoding carbohydrate-active enzymes (CAZymes), which have evolved in fungi to degrade carbohydrates such as lignin and cellulose. While typically associated with saprotrophism, CAZymes are also present to a lesser extent in EMF where they may support nutrient acquisition (Gong *et al.* 2023). We annotated 413 genes as encoding CAZymes, representing 2.7% of the total gene content. This is comparable to the 2.2% reported for *S. bovinus,* another EMF species (Zhang *et al.* 2024). These genes were unevenly distributed across chromosomes, with chromosome 9 carrying only 15 CAZymes compared with 52 on chromosome 1 (Fig. 2). We also identified genomic regions where multiple CAZymes clustered together. For example, at position 14,673,340 in the genome (located on chromosome 4), 4 CAZymes were found within a 54-kbp region (Fig. 2). These clusters may represent remnants of an ancestral presymbiotic lifestyle, as similar gene arrangements have been observed in wood-decaying fungi (Li *et al.* 2017). Further exploration of these CAZyme clusters might therefore provide insights into how EMF adapt to diverse environmental conditions.

Effector proteins are secreted by pathogenic fungi to manipulate host defenses during colonization (Li *et al.* 2024). They have also been shown to play a similar role during the early establishment of mycorrhizal symbioses, facilitating fungal colonization of host roots (Li *et al.* 2024). However, this process remains poorly

Table 2. Length and number of gaps per chromosome from each haplotype.

Chromosome	BolEdBiel _h1 length	BolEdBiel _h1 gaps	BolEdBiel _h2 length	BolEdBiel _h2 gaps
1	4,187,054	4	4,111,618	0
2	4,077,455	1	4,390,755	1
3	4,381,500	0	4,281,452	1
4	2,924,582	0	3,539,504	5
5	3,798,680	0	3,304,873	0
6	3,044,205	1	2,742,422	2
7	4,131,713	1	4,104,162	0
8	3,361,260	0	2,952,305	0
9	2,691,760	3	2,670,009	3
10	4,848,437	0	4,360,388	4
11	3,363,843	0	3,500,508	1
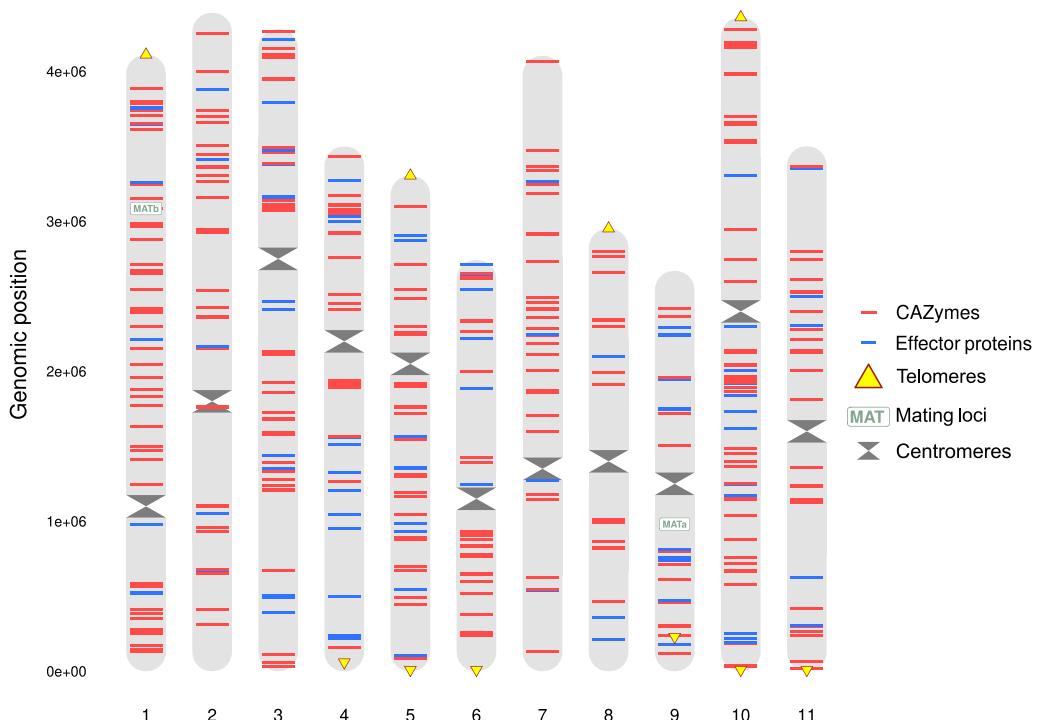
Fig. 2. The distribution of key features in the *B. edulis* reference genome. Shown are the genomic locations (in haplotype 2, BolEdBiel_h2) of mating loci, genes encoding CAZymes and effector proteins, centromeres, and telomeres.
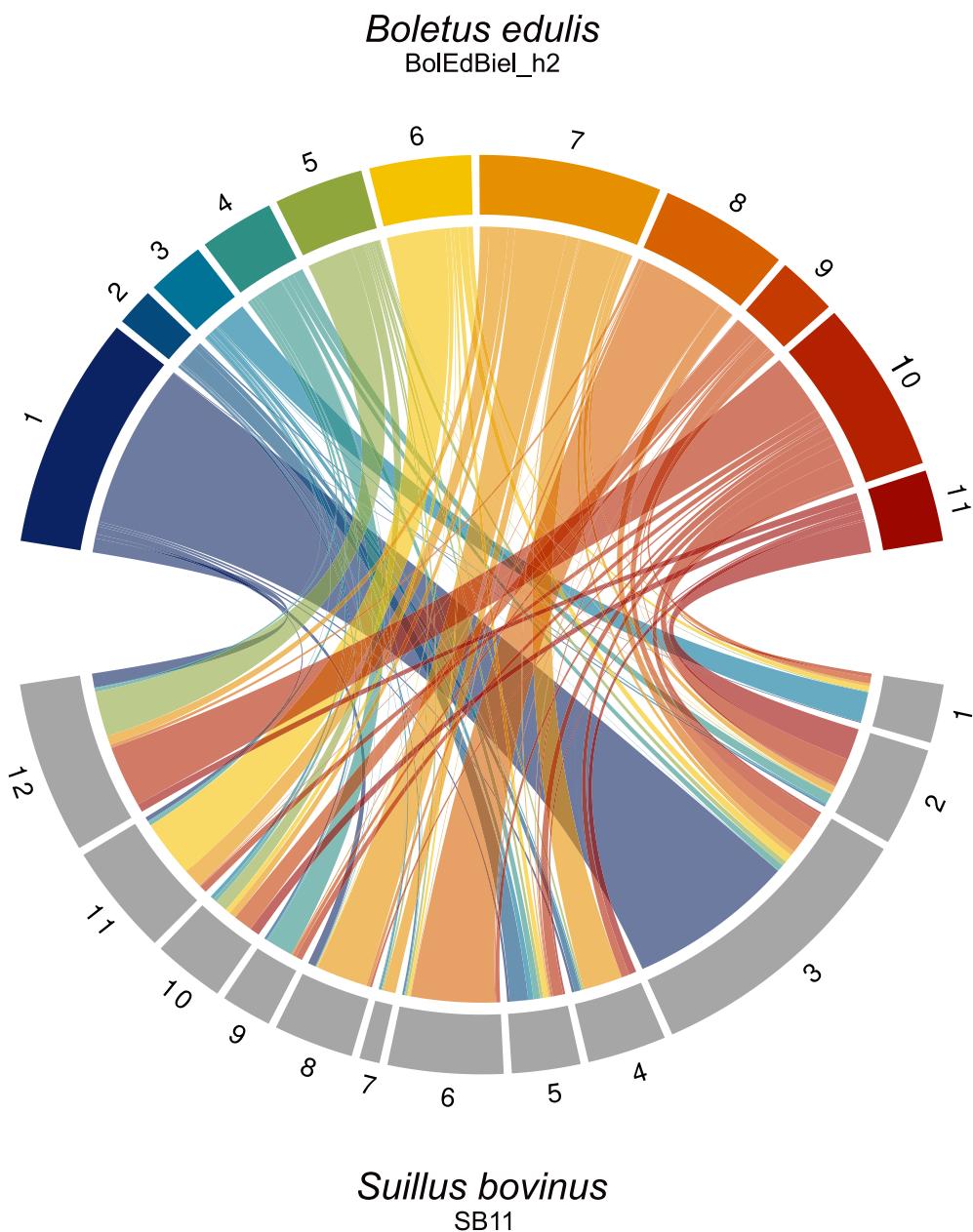
Fig. 3. Patterns of chromosomal synteny between *B. edulis* and *S. bovinus*. The Circos plot depicts alignments of BUSCO genes from haplotype 2 of *B. edulis* (top) to the reference genome of *S. bovinus* (below). The *B. edulis* chromosomes are shown in color, and the *S. bovinus* chromosomes are shown in gray.

understood, and only a small number of EMF effectors have been linked to plant gene expression pathways (Plett *et al.* 2011; Daguerre *et al.* 2020). Identifying these genes in the reference genome is therefore an important first step toward understanding their function and regulation in this species. We identified a total of 116 genes putatively encoding effector proteins. In pathogenic fungi, effector genes are often clustered in genomic regions experiencing rapid evolution, such as near telomeres (Zaccaron *et al.* 2022). However, the flanking regions of the 10 telomeric sequences identified here did not appear to be enriched for genes encoding effector proteins (Fig. 2).

Mating type (MAT) loci play a major role in determining sexual compatibility and reproduction in fungi (Coelho *et al.* 2017). In this tetrapolar species, we identified both mating loci: MATa on chromosome 9 and MATb on chromosome 1 (Fig. 2). Tetrapolar species are generally considered less prone to selfing than bipolar

species. In line with this, a recent study of *B. edulis* found no evidence of recent inbreeding (Brejon Lamartinière *et al.* 2024), despite the presence of locally elevated relatedness (Hoffman *et al.* 2020). This suggests that the mating loci may contribute toward the maintenance of high levels of heterozygosity and genetic diversity within *B. edulis* populations. Future studies could use pedigrees to model the involvement of MAT loci in mating outcomes and inbreeding/outbreeding dynamics.

Patterns of synteny

To provide phylogenetic and evolutionary context, we investigated patterns of synteny between *B. edulis* and *S. bovinus* based on shared BUSCO genes. While both species belong to the order Boletales, *S. bovinus* is part of the family Suillaceae, which diverged from Boletaceae ~115 million years ago (Wu *et al.* 2022). Despite sharing an ectomycorrhizal lifestyle with *B. edulis*,
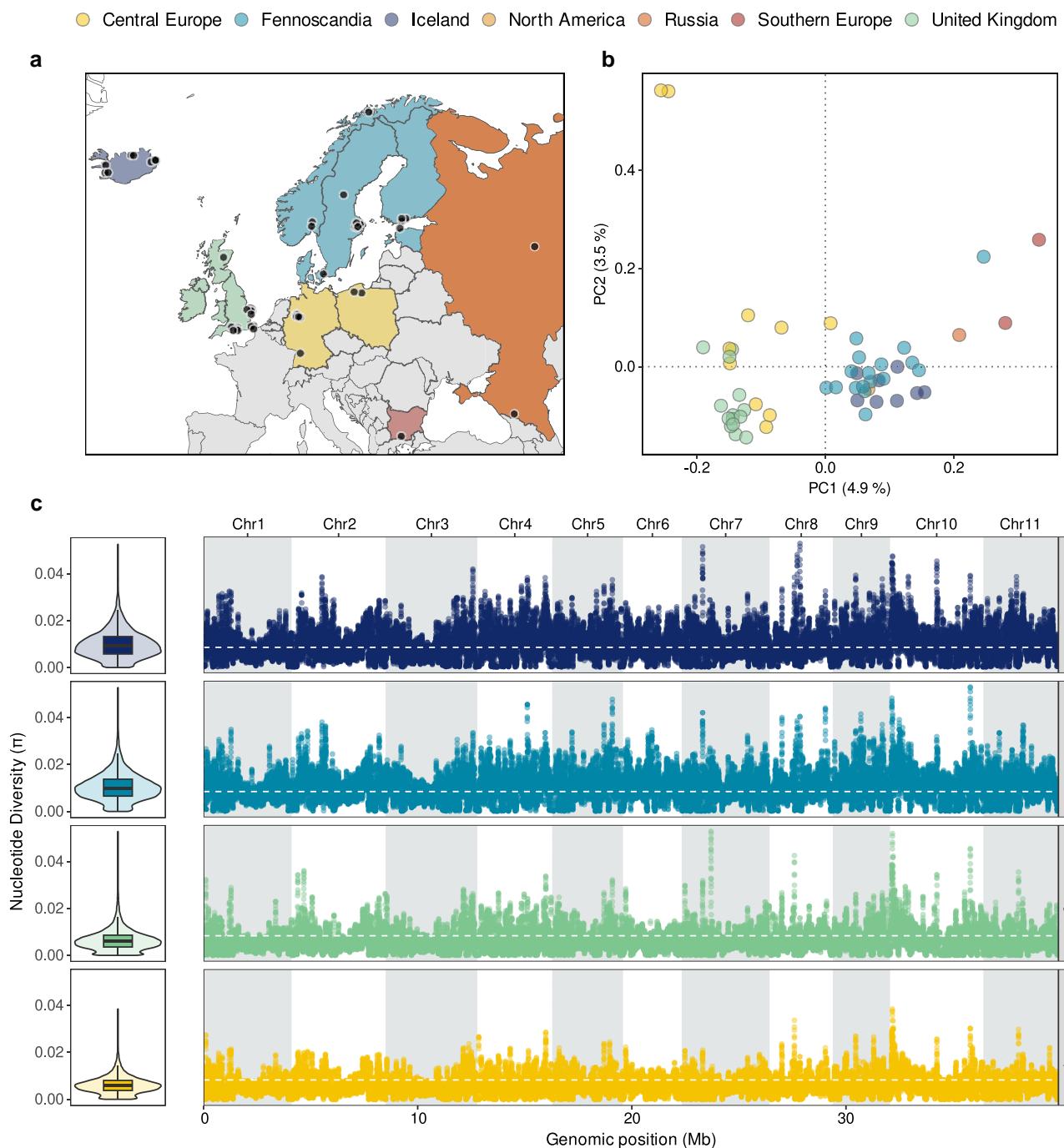
Fig. 4. Population structure and genetic diversity of the European lineage of *B. edulis*. a) Sampling locations of 53 genets, color-coded by geographical region. b) Scatter plots of individual variation in principal component (PC) scores derived from PCA of the genomic data. The amounts of variation explained by each PC are given as percentages on the axis labels, and the samples are color-coded as shown in a). c) Nucleotide diversity (π) calculated in 10-kb sliding windows across the genome for every population. The left panels show violin plots of the kernel densities of π together with standard Tukey boxplots (center line = median, bounds of the box = 25th and 75% percentiles, upper and lower whiskers = largest and smallest values but no further than 1.5 * inter-quartile range from the hinge). The right panels show π values for each window across the 11 chromosomes. The dashed white lines represent lineage-specific genome-wide average π values.

S. bovinus has a more restricted host range, primarily associating with Pinaceae (Zhang *et al.* 2024). Surprisingly, we observed high levels of synteny between these species (Fig. 3), despite their considerable phylogenetic distance, differences in chromosome numbers, and the structural and size variation present between homologous *B. edulis* haplotypes (Supplementary Table 1). Notably, large BUSCO gene segments on chromosomes 1 and 8 of *B. edulis* showed structural conservation with chromosomes 3

and 6 respectively of *S. bovinus* (Fig. 3). Generating additional chromosomal reference genomes for other EMF species would provide deeper insights into genome evolution through comparative chromosomal synteny analyses across multiple species.

Population structure and genetic diversity

We analyzed short-read data from 53 genets (Fig. 4a, Supplementary Table 3, Tremble, Hoffman, *et al.* 2023; Brejon

Lamartinière *et al.* 2024, and this manuscript) to genotype 937,328 SNPs for the assessment of population structure and genetic diversity within the European *B. edulis* lineage. Two main clusters were resolved in the PCA, with samples from the United Kingdom and Central Europe separating from samples from Iceland and Fennoscandia along the first axis (Fig. 4b). Notably, a single genet sampled in North America, which was assigned to the European lineage by Tremble, Hoffman, *et al.* (2023), clustered in the PCA alongside the Fennoscandian and Icelandic samples (Fig. 4b). Although we cannot be certain, this suggests that this individual may have been introduced to America from a tree plantation originating in northern Europe.

Genetic diversity was not evenly distributed among the genetic clusters (Fig. 4b). Populations at higher latitudes, specifically those in Iceland and Fennoscandia, exhibited higher genome-wide nucleotide diversity (π) than populations from the United Kingdom and Central Europe. This observation aligns with our previous finding of a negative (but nonsignificant) association between latitude and genomic inbreeding within the European *B. edulis* lineage (Brejon Lamartinière *et al.* 2024). Several factors may contribute to this pattern, including greater habitat availability and continuity in the north, where much of Iceland and Fennoscandia remains covered by forests or tundra. Additionally, differences in the amount of gene flow with the neighboring "AK" lineage, which spans Alaska and Siberia, may also play a role (Tremble, Brejon Lamartinière *et al.* 2023). Above and beyond this pattern, we also identified regions of elevated π within the *B. edulis* genome. Some of these regions were shared among the clusters, such as a prominent peak near the beginning of chromosome 10, while others were not universally shared, appearing only on specific chromosomes (Fig. 4c). These regions could represent areas of structural variation, they may be under balancing selection, or they might experience locally elevated recombination. This highlights the potential of the European lineage of *B. edulis* as a model for studying these diverse evolutionary processes.

Conclusion

Population genomic studies of EMF hold great potential for advancing our understanding of these ecologically important organisms. However, the contiguity of reference genomes continues to be a limiting factor. This study presents one of the first haplotype-resolved chromosomal reference assemblies for a dikaryotic EMF, achieving near telomere-to-telomere coverage across multiple chromosomes. Using chromatin conformation capture, we successfully anchored PacBio long reads to chromosomes for each haplotype. This approach revealed structural variation within homologous chromosomes of the reference individual and identified key areas for future research. Specifically, we identified mating loci, CAZymes and effector proteins, discovered a 0.4-Mb inversion on chromosome 3, investigated the population genetic structure of the European lineage, and discovered several genomic regions with locally elevated nucleotide diversity. We anticipate that this new resource will facilitate future discoveries in this enigmatic and ecologically important fungus.

Data availability statement

The code, reference assembly, the Hi-C and Pacbio raw reads generated in this study, and the annotation are available at Zenodo (DOI: 10.5281/zenodo.14311982). The genome assembly and the newly generated short reads are deposited at NCBI under the BioProject PRJNA1187522. All sequences used in this study are publicly available, and their corresponding accession numbers are listed in Supplementary Table 3.

Supplemental material available at G3 online.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

E.B.L. and J.I.H. conceived the study and drafted the manuscript. J.I.H. acquired funding and supervised the PhD student (E.B.L.) with additional scientific input from K.K.D. and B.T.M.D. E.B.L. and K.T. analyzed the data. All of the authors commented upon and approved the final manuscript.

Literature cited

Alioto T, Blanco E, Parra G, Guigó R. 2018. Using geneid to identify genes. Curr Protoc Bioinformatics. 64(1):e56. doi:10.1002/cpbi.56.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410. doi:10.1016/S0022-2836(05)80360-2.

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 21(1):30. doi:10.1186/s13059-020-1935-5.

Amend A, Keeley S, Garbelotto M. 2009. Forest age correlates with fine-scale spatial structure of matsutake mycorrhizas. Mycol Res. 113(5):541–551. doi:10.1016/j.mycres.2009.01.005.

Anthony MA, Tedersoo L, De Vos B, Croisé L, Meesenburg H, Wagner M, Andreae H, Jacob F, Lech P, Kowalska A, *et al.* 2024. Fungal community composition predicts forest carbon storage at a continental scale. Nat Commun. 15(1):2385. doi:10.1038/s41467-024-46792-w.

Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28(1):45. doi:10.1093/nar/28.1.45.

Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 6(1):11. doi:10.1186/s13100-015-0041-9.

Bazzicalupo AL, Ruytinx J, Ke Y-H, Coninx L, Colpaert JV, Nguyen NH, Vilgalys R, Branco S. 2020. Fungal heavy metal adaptation through single nucleotide polymorphisms and copy-number variation. Mol Ecol. 29(21):4157–4169. doi:10.1111/mec.15618.

Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi–C: a comprehensive technique to capture the

conformation of genomes. Methods. 58(3):268–276. doi:10.1016/j.ymeth.2012.05.001.

Birney E, Clamp M, Durbin R. 2004. GeneWise and genomewise. Genome Res. 14(5):988. doi:10.1101/gr.1865504.

Bonello P, Bruns TD, Gardes M. 1998. Genetic structure of a natural population of the ectomycorrhizal fungus Suillus pungens. New Phytol. 138(3):533–542. doi:10.1046/J.1469-8137.1998.00122.X.

Borgognone A, Castanera R, Morselli M, López-Varas L, Rubbi L, Pisabarro AG, Pellegrini M, Ramírez L. 2018. Transposon-associated epigenetic silencing during Pleurotus ostreatus life cycle. DNA Res. 25(5):451–464. doi:10.1093/dnares/dsy016.

Branco S, Bi K, Liao H-L, Gladieux P, Badouin H, Ellison CE, Nguyen NH, Vilgalys R, Peay KG, Taylor JW, *et al.* 2016. Continental-level population differentiation and environmental adaptation in the mushroom Suillus brevipes. Mol Ecol. 26(7):2063. doi:10.1111/mec.13892.

Brejon Lamartinière E, Tremble K, Dentinger BTM, Dasmahapatra KK, Hoffman JI. 2024. Runs of homozygosity reveal contrasting histories of inbreeding across global lineages of the edible porcini mushroom, Boletus edulis. Mol Ecol. n/a(16):e17470. 10.1111/mec.17470.

Brown M, González De la Rosa PM, Mark B. 2023. A Telomere Identification Toolkit. Zenodo. https://doi.org/10.5281/zenodo.10091385.

Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA1. J Mol Biol. 268(1):78–94. doi:10.1006/jmbi.1997.0951.

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol. 31(12):1119. doi:10.1038/nbt.2727.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 37(Database):D233–D238. doi:10.1093/nar/gkn663.

Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol. 1962:1. doi:10.1007/978-1-5339-9173-0_1.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 18(2):170–175. doi:10.1038/s41592-020-01056-5.

Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. 2017. Fungal sex: the Basidiomycota. Microbiol Spectr. 5(3):5.3.12. doi:10.1128/microbiolspec.FUNK-0046-2016.

Daguerre Y, Basso V, Hartmann-Wittulski S, Schellenberger R, Meyer L, Bailly J, Kohler A, Plett JM, Martin F, Veneault-Fourrey C. 2020. The mutualism effector MiSSP7 of Laccaria bicolor alters the interactions between the poplar JAZ6 protein and its associated proteins. Sci Rep. 10(1):20362. doi:10.1038/s41598-020-76832-6.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011. The variant call format and VCFtools. Bioinformatics. 27(15):2156–2158. doi:10.1093/bioinformatics/btr330.

Douhan GW, Vincenot L, Gryta H, Selosse MA. 2011. Population genetics of ectomycorrhizal fungi: from current knowledge to emerging directions. Fungal Biol. 115(7):569–597. doi:10.1016/j.funbio.2011.03.005.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3(1):95–98. 10.1016/j.cels.2016.07.002.

Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, Sulovari A, Ebler J, Zhou W, Serra Mari R, *et al.* 2021. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science. 372(6537):eabf7117. doi:10.1126/science.abf7117.

Edgar RC, Myers EW. 2005. PILER: identification and classification of genomic repeats. Bioinformatics. 21(Suppl 1):i152–i158. doi:10.1093/bioinformatics/bti1003.

Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, *et al.* 2013. The reference genome sequence of Saccharomyces cerevisiae: then and now. G3 (Bethesda) (Bethesda, Md.). 4(3):389. doi:10.1534/g3.113.008995.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39(Web Server issue):W29. doi:10.1093/nar/gkr367.

Gazol A, Camarero JJ, Anderegg WRL, Vicente-Serrano SM. 2017. Impacts of droughts on the growth resilience of Northern Hemisphere forests. Glob Ecol Biogeogr. 26(2):166–176. doi:10.1111/geb.12526.

Gong Y, Lebreton A, Zhang F, Martin F. 2023. Role of carbohydrate-active enzymes in mycorrhizal symbioses. Essays Biochem. 67(3):471–478. doi:10.1042/EBC20220127.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 9(1):R7. doi:10.1186/gb-2008-9-1-r7.

Han Y, Wessler SR. 2010. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. 38(22):e199. doi:10.1093/nar/gkq862.

Hoede C, Arnoux S, Moisset M, Chaumier T, Inizan O, Jamilloux V, Quesneville H. 2014. PASTEC: an automatic transposable element classification tool. PLoS One. 9(5):e91929. doi:10.1371/journal.pone.0091929.

Hoffman JI, Nagel R, Litzke V, Wells DA, Amos W. 2020. Genetic analysis of Boletus edulis suggests that intra-specific competition may reduce local genetic diversity as a woodland ages. R Soc Open Sci. 7(7):200419. doi:10.1098/RSOS.200419.

Hoffmann AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annu Rev Ecol Evol Syst. 39(1):21–42. doi:10.1146/annurev.ecolsys.39.110707.173532.

Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, Bakker PIWD. 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 24(24):2938. doi:10.1093/bioinformatics/btn564.

Keilwagen J, Hartung F, Grau J. 2019. Gemoma: homology-based gene prediction utilizing intron position conservation and RNA-seq data. Methods Mol Biol. 1962:161–177. doi:10.1007/978-1-5339-9173-0_9.

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A, *et al.* 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nat Genet. 47(4):410–415. doi:10.1038/ng.3223.

Kurokochi H, Tajima N, Sato MP, Yoshitake K, Asakawa S, Isobe S, Shirasawa K. 2023. Telomere-to-telomere genome assembly of matsutake (Tricholoma matsutake). DNA Res. 30(3):dsad006. doi:10.1093/dnares/dsad006.

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. PLoS Comput Biol. 9(8):e1003118. doi:10.1371/journal.pcbi.1003118.

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

Li W-C, Huang C-H, Chen C-L, Chuang Y-C, Tung S-Y, Wang T-F. 2017. Trichoderma reesei complete genome sequence, repeat-induced point mutation, and partitioning of CAZyme gene clusters. Biotechnol Biofuels. 10(1):170. doi:10.1186/s13068-017-0825-x.

Li G, Newman M, Yu H, Rashidzade M, Martínez-Soto D, Caicedo A, Allen KS, Ma L-J. 2024. Fungal effectors: past, present, and future. Curr Opin Microbiol. 81:102526. doi:10.1016/j.mib.2024.102526.

Looney B, Miyauchi S, Morin E, Drula E, Courty PE, Kohler A, Kuo A, LaButti K, Pangilinan J, Lipzen A, *et al.* 2022. Evolutionary transition to the ectomycorrhizal habit in the genomes of a hyperdiverse lineage of mushroom-forming fungi. New Phytol. 233(5):2294–2309. doi:10.1111/nph.17892.

Luu K, Bazin E, Blum MGB. 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Resour. 17(1):67–77. doi:10.1111/1755-0998.12592.

Ma L, Yang C, Xiao D, Liu X, Jiang X, Lin H, Ying Z, Lin Y. 2023. Chromosome-level assembly of Dictyophora rubrovolvata genome using third-generation DNA sequencing and Hi-C analysis. G3 (Bethesda). 13(8):jkad102. doi:10.1093/g3journal/jkad102.

Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. Bioinformatics. 20(16):2878–2879. doi:10.1093/bioinformatics/bth315.

Mariene GM, Wasmuth JD. 2025. Genome assembly variation and its implications for gene discovery in nematode. Int J Parasitol. 55(5):239–252. doi:10.1016/j.ijpara.2025.01.004.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al.* 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303. doi:10.1101/gr.107524.110.

Medema MH, Blin K, Cimermancic P, Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39(Web Server issue):W339–W346. doi:10.1093/nar/gkr466.

Miyauchi S, Kiss E, Kuo A, Drula E, Kohler A, Sánchez-García M, Morin E, Andreopoulos B, Barry KW, Bonito G, *et al.* 2020. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. Nat Commun. 11(1):5125. doi:10.1038/S41467-020-18795-W.

Mizuguchi T, Fudenberg G, Mehta S, Belton J-M, Taneja N, Folco HD, Fitzgerald P, Dekker J, Mirny L, Barrowman J, *et al.* 2014. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. *pombe*. Nature. 516(7531):432–435. doi:10.1038/nature13833.

Morin E, Kohler A, Baker AR, Foulongne-Oriol M, Lombard V, Nagye LG, Ohm RA, Patyshakuliyeva A, Brun A, Aerts AL, *et al.* 2012. Genome sequence of the button mushroom agaricus bisporus reveals mechanisms governing adaptation to a humic-rich ecological niche. Proc Natl Acad Sci U S A. 109(43):17501–17506. doi:10.1073/pnas.1206847109.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 29(22):2933–2935. doi:10.1093/bioinformatics/btt509.

Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, *et al.* 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 20(1):275. doi:10.1186/s13059-019-1905-y.

Palmer JM, Stajich JE. 2023. Funannotate. [accessed 2024 Oct 18]. https://github.com/nextgenusfs/funannotate

Plett JM, Kemppainen M, Kale SD, Kohler A, Legué V, Brun A, Tyler BM, Pardo AG, Martin F. 2011. A secreted effector protein of Laccaria bicolor is required for symbiosis development. Curr Biol. 21(14):1197–1203. doi:10.1016/j.cub.2011.05.033.

Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. Bioinformatics. 21(Suppl 1):i351–i358. doi:10.1093/bioinformatics/bti1018.

Rahnama M, Wang B, Dostart J, Novikova O, Yackzan D, Yackzan A, Bruss H, Baker M, Jacob H, Zhang X, *et al.* 2021. Telomere roles in fungal genome evolution and adaptation. Front Genet. 12:676751. doi:10.3389/fgene.2021.676751.

Read DJ, Perez-Moreno J. 2003. Mycorrhizas and nutrient cycling in ecosystems—a journey towards relevance? New Phytol. 157(3):475–492. doi:10.1046/j.1469-8137.2003.00704.x.

Roy B, Sanyal K. 2011. Diversity in requirement of genetic and epigenetic factors for centromere function in fungi. Eukaryot Cell. 10(11):1384–1395. doi:10.1128/ec.05165-11.

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 16(1):259. doi:10.1186/s13059-015-0831-x.

She R, Chu JS-C, Wang K, Pei J, Chen N. 2009. genBlastA: enabling BLAST to identify homologous gene sequences. Genome Res. 19(1):143–149. doi:10.1101/gr.082081.108.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31(19):3210–3212. doi:10.1093/bioinformatics/btv351.

Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. [accessed 2024 Oct 18]. http://www.repeatmasker.org

Somvanshi VS, Sloup RE, Crawford JM, Martin AR, Heidt AJ, Kim K, Clardy J, Ciche TA. 2012. A single promoter inversion switches Photorhabdus between pathogenic and mutualistic states. Science. 337(6090):88–93. doi:10.1126/science.1216641.

Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, Manners JM, Taylor JM. 2016. Effectorp: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210(2):743–761. doi:10.1111/nph.13794.

Sperschneider J, Yildirir G, Rizzi YS, Malar C M, Mayrand Nicol A, Sorwar E, Villeneuve-Laroche M, Chen ECH, Iwasaki W, Brauer EK, *et al.* 2023. Arbuscular mycorrhizal fungi heterokaryons have two nuclear populations with distinct roles in host–plant interactions. Nat Microbiol. 8(11):2142–2153. doi:10.1038/s41564-023-01535-8.

Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32(Web Server issue):W309–W312. doi:10.1093/nar/gkh379.

Steidinger BS, Crowther TW, Liang J, Van Nuland ME, Werner GDA, Reich PB, Nabuurs GJ, de-Miguel S, Zhou M, Picard N, *et al.* 2019. Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. Nature. 569(7756):404–408. doi:10.1038/s41586-019-1128-0.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2012. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 14(2):178. doi:10.1093/bib/bbs017.

Treindl AD, Leuchtmann A. 2019. A king amongst dwarfs: Boletus edulis forms ectomycorrhiza with dwarf willow in the Swiss Alps. Alp Bot. 129(2):185–189. doi:10.1007/s00035-019-00218-2.

Tremble K, Brejon Lamartinière E, Bradshaw AJ, Arzú RF, Hoffman J, Dentinger BTM. 2023. From populations to pan-genomes: investigating the role of ecology and genomic architecture in

maintaining species boundaries in the porcini mushroom, *Boletus edulis*. bioRxiv 556370. https://doi.org/10.1101/2023.09.05.556370.

Tremble K, Hoffman JI, Dentinger BTM. 2023. Contrasting continental patterns of adaptive population divergence in the holarctic ectomycorrhizal fungus Boletus edulis. New Phytol. 237(1):295–309. doi:10.1111/nph.18521.

Tremble K, Suz LM, Dentinger BTM. 2020. Lost in translation: population genomics and long-read sequencing reveals relaxation of concerted evolution of the ribosomal DNA cistron. Mol Phylogenet Evol. 148:106804. doi:10.1016/j.ympev.2020.106804.

Voller F, Ardanuy A, Taylor AFS, Johnson D. 2024. Maintenance of host specialisation gradients in ectomycorrhizal symbionts. New Phytol. 242(4):1426–1435. doi:10.1111/nph.19395.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al*. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9(11):e112963. doi:10.1371/journal.pone.0112963.

Weissensteiner MH, Pang AWC, Bunikis I, Höijer I, Vinnere-Petterson O, Suh A, Wolf JBW. 2017. Combination of short-read, long-read, and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. Genome Res. 27(5):697–708. doi:10.1101/gr.215095.116.

Wu G, Miyauchi S, Morin E, Kuo A, Drula E, Varga T, Kohler A, Feng B, Cao Y, Lipzen A, *et al*. 2022. Evolutionary innovations through gain and loss of genes in the ectomycorrhizal Boletales. New Phytol. 233(3):1383–1400. doi:10.1111/nph.17858.

Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 35(Web Server issue):W265–W268. doi:10.1093/nar/gkm286.

Yu H, Zhang L, Shang X, Peng B, Li Y, Xiao S, Tan Q, Fu Y. 2022. Chromosomal genome and population genetic analyses to reveal genetic architecture, breeding history and genes related to cadmium accumulation in Lentinula edodes. BMC Genomics. 23(1):120. doi:10.1186/s12864-022-08325-x.

Zaccaron AZ, Chen L-H, Samaras A, Stergiopoulos I. 2022. A chromosome-scale genome assembly of the tomato pathogen Cladosporium fulvum reveals a compartmentalized genome architecture and the presence of a dispensable chromosome. Microb Genom. 8(4):000819. doi:10.1099/mgen.0.000819.

Zhang J, An M, Chen Y, Wang S, Liang J. 2024. Chromosome-level assembly and comparative genomic analysis of Suillus bovinus provides insights into the mechanism of mycorrhizal symbiosis. J Fungi. 10(3):211. doi:10.3390/jof10030211.

Editor: M. Sachs