



HAL
open science

High quality genome of the basidiomycete yeast *Dioszegia hungarica* PDD-24b-2 isolated from cloud water

Domitille Jarrige, Sajeet Haridas, Claudine Bleykasten-Grosshans, Muriel Joly, Thierry Nadalig, Martine Sancelme, Stéphane Vuilleumier, Igor V Grigoriev, Pierre Amato, Françoise Bringel

► To cite this version:

Domitille Jarrige, Sajeet Haridas, Claudine Bleykasten-Grosshans, Muriel Joly, Thierry Nadalig, et al.. High quality genome of the basidiomycete yeast *Dioszegia hungarica* PDD-24b-2 isolated from cloud water. *G3*, 2022, 10.1093/g3journal/jkac282 . hal-03831438

HAL Id: hal-03831438

<https://hal.science/hal-03831438>

Submitted on 26 Oct 2022

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.

High quality genome of the basidiomycete yeast

Dioszegia hungarica PDD-24b-2 isolated from cloud water

Domitille Jarrige¹, Sajeet Haridas², Claudine Bleykasten-Grosshans¹, Muriel Joly³, Thierry Nadalig¹, Martine Sancelme³, Stéphane Vuilleumier¹, Igor V. Grigoriev^{2,4}, Pierre Amato³, Françoise Bringel^{1*}

Affiliations

¹Génétique Moléculaire, Génomique, Microbiologie (GMGM), Université de Strasbourg, UMR 7156 CNRS, Strasbourg, France

² U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA USA

³ Université Clermont Auvergne, Clermont Auvergne Institut National Polytechnique (INP), Centre National de la Recherche Scientifique (CNRS), Institut de Chimie de Clermont-Ferrand (ICCF), Clermont-Ferrand, France

⁴ Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA USA

*Author for Correspondence: Françoise Bringel, Génétique Moléculaire, Génomique, Microbiologie (GMGM), Université de Strasbourg, UMR 7156 CNRS, Strasbourg, France, E-mail: francoise.bringel@unistra.fr

Running Head: Genome of *Dioszegia hungarica* PDD-24b-2

Keywords: *Tremellaceae*, mitochondrial genome, *de novo* sequencing, fungi, airborne microorganisms, transposable elements, *Dioszegia hungarica* strain PDD-24b-2, aeromicrobiology, fungal spore, cold environment

23 **Abstract**

24 The genome of the basidiomycete yeast *Dioszegia hungarica* strain PDD-24b-2 isolated from cloud water at the
25 summit of puy de Dôme (France) was sequenced using a hybrid PacBio and Illumina sequencing strategy. The
26 obtained assembled genome of 20.98 Mb and a GC content of 57% is structured in 16 large-scale contigs
27 ranging from 90 kb to 5.56 Mb, and another 27.2 kb contig representing the complete circular mitochondrial
28 genome. In total, 8,234 proteins were predicted from the genome sequence. The mitochondrial genome shows
29 16.2% egu codon usage for arginine but has no canonical cognate tRNA to translate this codon. Detected
30 transposable element-related sequences account for about 0.63% of the assembled genome. A dataset of 2,068
31 hand-picked public environmental metagenomes, representing over 20 Tbp of raw reads, was probed for *D.*
32 *hungarica* related ITS sequences, and revealed worldwide distribution of this species, particularly in aerial
33 habitats. Growth experiments suggested a psychrophilic phenotype and the ability to disperse by producing
34 ballistospores. The high-quality assembled genome obtained for this *D. hungarica* strain will help investigate the
35 behavior and ecological functions of this species in the environment.

36 Introduction

37 There is increasing evidence that airborne microorganisms participate in chemical transformations and physical
38 processes in the atmosphere (Šantl-Temkiv *et al.* 2022). In particular, microorganisms found in clouds play a
39 central role in reactions of carbon-containing compounds at night, whereas during the day, photochemistry is
40 dominant (Vařtilingom *et al.* 2012, 2013). Both prokaryotic and eukaryotic microorganisms can be found in
41 clouds (Delort *et al.* 2017). Regarding eukaryotes, 1-3% of sequenced 18S rDNA amplicons belong to the class
42 Tremellomycetes of basidiomycete yeasts (Amato *et al.* 2017), which include the genus *Dioszegia* (Order:
43 Tremellales; Family: Tremellaceae/ Bulleribasidiaceae) (Liu *et al.* 2015). *Dioszegia hungarica* strain PDD-24b-2
44 was isolated from cloud water collected at the summit of the puy de Dôme, France (Vařtilingom *et al.* 2012)
45 (**Figure S1**). Strains identified as *Dioszegia* sp. are frequently isolated from cloud water sampled at this site (in
46 70% of studied samples; Vařtilingom *et al.*, 2012). This fungal taxon was repeatedly identified in various cold
47 environments, such as snow and glacial meltwater rivers (de García *et al.* 2007), and is also associated with
48 plants in Antarctica (Ferreira *et al.* 2019). The *D. hungarica* type strain CBS 4214^T was isolated from soil in
49 Kűlsó-tó, Hungary as described in (Takashima *et al.* 2001). Also found in warmer environments, it is part of the
50 core fungal community of the wheat phyllosphere (the aerial parts of plants) (Karlsson *et al.* 2017; Sapkota *et al.*
51 2017). *D. hungarica* was identified as one of the few ‘microbial hub taxa’ that, when influenced by plant host
52 and abiotic factors, act on the plant microbiome. For example, it directly inhibits the growth of specific bacterial
53 taxa on *Arabidopsis thaliana* seedlings, thus decreasing the phyllosphere bacterial community diversity (Aglar *et*
54 *al.* 2016). The atmospheric environment in which airborne microbes are found represents both a source
55 (immigration) and a sink (emigration) for the phyllosphere microbiome (Kinkel 1997). Examining the genome of
56 *D. hungarica* may provide valuable information to better understand the dynamics of fungal diversity, especially
57 at the plant/atmosphere interface, and its role in climate change-relevant ecosystems (e.g., clouds, cold
58 environments, phyllosphere).

59 *Dioszegia hungarica*, formerly classified as *Cryptococcus hungaricus* and *Bullera armeniaca*
60 (Takashima *et al.* 2001), is one of the 23 species of *Dioszegia* identified so far (Li *et al.* 2020). To date, genomes
61 of three other *Dioszegia* species have been sequenced: *D. aurantiaca* strain JCM 2956 and *D. crocea* strain JCM
62 2961, isolated from overwintered nettle stems of *Urtica* sp. and strawberry phyllosphere, respectively
63 (Takashima *et al.* 2019), and *D. cryoxerica* strain ANT03-071 (<https://mycocosm.jgi.doe.gov/Diocrl>), isolated
64 from moss in Antarctica (Connell *et al.* 2010). Previous analyses of the internal transcribed spacer and D1/D2
65 regions of the large subunit rRNA gene showed that *D. hungarica* is phylogenetically distant from these

66 genome-sequenced representatives of the genus (Trochine *et al.* 2017; Li *et al.* 2020). This makes the species
67 *D. hungarica* a good candidate to further investigate fungal genetic diversity. In this study, we describe the high-
68 quality assembled genome sequence of *D. hungarica* strain PDD-24b-2 obtained by a hybrid PacBio and
69 Illumina sequencing strategy. The assembled genome features 17 contigs, 16 large-scale linear contigs and a
70 smaller contig representing the complete circular mitochondrial genome.

71 **Materials and Methods**

72 **Strain and growth conditions**

73 *Dioszegia hungarica* strain PDD-24b-2 was isolated from cloud water collected at the summit of puy de Dôme,
74 France on 17 January 2008 (Vařtilingom *et al.* 2012). R2A liquid medium was prepared as described previously
75 (Reasoner and Geldreich 1985). Commercial dehydrated R2A agar (Oxoid, Hampshire, U. K.) was used as solid
76 medium. Yeast mold (YM) medium (pH 6.2) contained per liter 3 g yeast extract, 3 g malt extract, 5 g peptone
77 (pancreatic digest gelatin), 10 g D-glucose, and was supplemented with 20 g agar for solid medium. Liquid
78 cultures were grown at 17°C with agitation (Sanyo MIR 254 refrigerated incubator, MA, USA). The ability to
79 produce ballistospores was assessed on R2A solid medium, placing an inoculated Petri dish above a sterile one
80 as described previously (Ianiri *et al.* 2014).

81 **DNA extraction and PCR amplification**

82 Total DNA was extracted from a 4-day aerobic culture (OD at 600 nm of 0.97) in 200 mL R2A medium
83 incubated at 17°C, using the MasterPure™ complete DNA and RNA purification kit as described by the
84 manufacturer (Lucigen, WI, USA). The 18S rRNA gene was PCR-amplified from total DNA (25 ng) using
85 primers Dios20F (5'-GTGCGTCTGATTCTTGACTCC-3') and Dios11R (5'-
86 CCCGACCGTCCCTATTAATCA-3') and DreamTaq DNA polymerase, as recommended by the manufacturer
87 (Thermo Fisher Scientific Baltics, Vilnius, Lithuania). The PCR program (Biometra TOne thermocycler,
88 Analytik Jena, Jena, Germany) involved DNA denaturation at 95°C for 5 min, 30 cycles of 45 s at 93°C, 20 s at
89 56°C and 1 min at 72°C, and a final 10 min extension at 72°C. The amplified 1,080 bp PCR fragment was
90 sequenced by the Sanger method (Microsynth France, Vaulx-en-Velin, France).

91 **Genome sequencing, assembly and automatic annotation**

92 Illumina library preparation (Nextera XT kit), PacBio library preparation (SMRTbell express template prep kit
93 2.0) and high throughput sequencing of *D. hungarica* PDD-24b-2 were performed by GenoScreen (Lille,

94 France). Libraries were sequenced using the MiSeq Illumina platform and the PacBio platform (SMRT cell
95 Pacbio Sequel). Illumina and PacBio reads were quality checked with FastQC v0.11.9 (Andrews, Braham
96 Bioinformatics).

97 Illumina adapter sequences were removed with CutAdapt v2.10 (Martin 2011) and paired-end reads cleaned with
98 Prinseq v0.20.4 (Schmieder and Edwards 2011): the first 15 nucleotides of each read were cut, nucleotides with
99 a Phred score under 30 were cut from the read 3' end, reads shorter than 60 nucleotides were discarded, reads
100 with a mean Phred score under 30 were discarded, as well as those containing undetermined nucleotides. Only
101 paired reads were conserved. After these processing steps 8,383,275 read pairs were obtained.

102 PacBio subreads were assembled with Flye v2.8.2 (Kolmogorov *et al.* 2019). The cleaned Illumina read pairs
103 were used to correct the PacBio assembly using BOWTIE2 v2.4.1 (Langmead and Salzberg 2012; Langmead *et*
104 *al.* 2019) and Pilon v1.23 (Walker *et al.* 2014). Contigs were aligned to each other using BLASTn v2.10.1
105 (Camacho *et al.* 2009) to resolve alternative haplotypes. Telomeric repeats of the sequence T₂AG₃₋₅, akin to
106 those of *Cryptococcus neoformans* (Edman 1992), were searched and visualized in this assembly using IGV v
107 2.12.0 (Robinson *et al.* 2011). Completeness of the *D. hungarica* PDD-24b-2 genome assembly was assessed
108 with BUSCO v5.2.2 (Manni *et al.* 2021) against tremellomycetes_odb10, and compared to the 3 previously
109 released genomes for the *Dioszegia* genus.

110 The nuclear genome was deposited in the MycoCosm platform (Grigoriev *et al.* 2014) and automatically
111 annotated, as previously described (Kuo *et al.* 2014) using the JGI Annotation Pipeline. The mitochondrial
112 genome annotation pipeline combined *ab initio* predictions, homology-based predictions with a curated
113 mitochondrial protein set, and HMM-based predictions, as described in (Haridas *et al.* 2018). The KOG
114 classification scheme was used to evaluate the number of genes associated with predicted processes with detailed
115 gene ID available at <https://mycocosm.jgi.doe.gov/cgi-bin/kogBrowser?type=KOG&db=Diohu1>). The KEGG
116 pathway database was used to identify metabolic pathway genes ([https://mycocosm.jgi.doe.gov/cgi-](https://mycocosm.jgi.doe.gov/cgi-bin/metapathways?db=Diohu1)
117 [bin/metapathways?db=Diohu1](https://mycocosm.jgi.doe.gov/cgi-bin/metapathways?db=Diohu1)).

118 **Identification of transposable elements**

119 Putative transposable elements (TEs) were searched in the *D. hungarica* PDD-24b-2 genome sequence using two
120 *de novo* approaches: the RepeatModeler v2.0.3 pipeline (Flynn *et al.* 2020) with its LTR pipeline extensions, and
121 the Extensive *de novo* TE Annotator (EDTA) pipeline (Bell *et al.* 2022). TEs were also identified by protein
122 homology with transposon sequences from other fungi (*Saitozyma podzolica*, *Cryptococcus neoformans*,

123 *Cryptococcus gattii*, *Rhodotorula toruloides*, *Candida glabrata*) using BLAST+tblastx v2.11.0 (Camacho *et al.*
124 2009). Detected sequences were manually curated using CD Search (Marchler-Bauer and Bryant 2004;
125 Marchler-Bauer *et al.* 2017) to predict conserved protein domains. Target Site Duplications (TSD) were
126 identified by manually checking for direct repeats in sequences adjacent to identified TEs, and confirmed by
127 surveying several copies. Detected putative TEs were classified into Orders and Superfamilies, as previously
128 described (Wicker *et al.* 2007). Unique candidate TEs for each family, with typical TE domains (e.g.
129 transposase, reverse transcriptase, RNase H, integrase, aspartic protease, *gag* domains) were then used to build a
130 putative TE library (**File S1**) to screen the genome sequence of *D. hungarica* PDD-24b-2 using
131 RepeatMasker v4.1.2-p1 and estimate putative TE copy number, including full-length and truncated copies. The
132 RepeatMasker.out file was parsed with the tool “One code to find them all” (Bailly-Bechet *et al.* 2014) to
133 assemble detected TE fragments. More information on the TE-mining process can be found in the following
134 GitHub repository: https://github.com/JarrigeD/Dioszegia_hungarica_sequencing

135 **Phylogenetic analysis**

136 The Internal Transcribed Spacer (ITS) region (ITS1, 5.8S ribosomal RNA gene, ITS2 and large subunit
137 ribosomal RNA gene partial sequence) phylogenetic tree of the *Dioszegia* genus was constructed entirely with
138 MEGA11 (Tamura *et al.* 2021). An alignment of the ITS region was generated using MUSCLE (Edgar 2004) on
139 a total of 462 positions for *D. hungarica* PDD-24b-2, reference strains of the 23 *Dioszegia* species described to
140 date, and 2 *Hannaella* type strains used as outgroup (**Table S2**). The MEGA11 “find best fit substitution model”
141 tool was used to choose the substitution model for tree building. The General Time Reversible model with rate
142 heterogeneity across sites (GTR+G) (Tavaré 1986; Yang 1996) had the lowest Bayesian information criterion
143 score and corrected Akaike information criterion score and was used to calculate the matrix of pairwise
144 distances. A discrete Gamma distribution was used to model evolutionary rate differences among sites
145 (5 categories (+G, parameter = 0.1890)). A total of 500 replicate trees were built with the Maximum Likelihood
146 method to calculate bootstrap support values, and the best tree topology with the highest log likelihood (-
147 2379.73) was selected.

148 **Geographical distribution**

149 A total of 2,068 whole genome shotgun (WGS) raw read metagenomic datasets were hand-picked to maximize
150 geographic and environmental variety, and retrieved from the Sequence Read Archive (SRA) using sra-tools.
151 The presence of *D. hungarica* sequences was tested using sra-tools blastn_vdb megablast, with strain PDD-24-
152 2b-2 ITS region as query and a minimum percentage identity threshold of 97% . The resulting BLAST hits were

153 filtered to target members of the genus *Dioszegia* (≥ 45 nt with $\geq 99\%$ identity to the PDD-24b-2 5.8S rRNA
154 gene sequence) and of the species *D. hungarica* (≥ 15 nt to ITS1 or ITS2 sequences, E-value $\leq 10e^{-10}$). These
155 thresholds were defined using alignments of ITS regions of *D. hungarica* PDD-24b-2 to those of fungal type
156 strains in the NCBI ITS_RefSeq_Fungi database. Maps of WGS datasets with *D. hungarica* or *Dioszegia* sp. hits
157 were plotted in Python v3.10.2 using Matplotlib v3.5.1 and GeoPandas v0.10.2. Details of metadata and dataset
158 accessions, homolog search scripts, filtering parameters and mapping processes are available at
159 https://github.com/JarrigeD/Dioszegia_hungarica_sequencing.

160 The GlobalFungi database (accessible at <https://globalfungi.com/>) (17,000 ITS amplicons environmental
161 samples) (Větrovský *et al.* 2020), the MarDB database (accessible at <https://mmp.sfb.uit.no/blast/>) (14,600
162 marine microbial genomes) (Klemetsen *et al.* 2018; Priyam *et al.* 2019) and TARA Ocean Gene Atlas databases
163 (Villar *et al.* 2018) (accessible at <https://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas>) EUK_SMAGs (713
164 eukaryotic plankton Metagenome Assembled Genomes: MAGs) (Delmont *et al.* 2022), MATOUv1_metaG
165 (116.8 million eukaryotic expressed genes + 530 Arctic Ocean MAGs) (Carradec *et al.* 2018) and OM-
166 RGC_v2_metaG (370 marine metagenomes + 530 MAGs) (Salazar *et al.* 2019) were also searched for close
167 homologs of the ITS region of *Dioszegia hungarica* PDD-24b-2.

168

169 **Results and discussion**

170 **Cell morphology and growth characteristics**

171 Single ovoid cells of approximately 4 μm in length, dividing by polar budding, were observed by optical
172 microscopy (**Figure 1A**). *D. hungarica* strain PDD-24b-2 grows in R2A and YM media with the characteristic
173 deep orange color typical of this genus (Inacio *et al.* 2005), which becomes more pronounced at higher cell
174 density (**Figure 1B, 1C**). Growth in YM was faster than in R2A (**Figure 1D**), with incubation temperature
175 strongly affecting growth (**Figure 1D**). The shortest doubling times were observed at 17°C in both YM (297 ± 5
176 min) and R2A (381 ± 33 min) media. Incubation at 4°C and 25°C resulted in 3-fold and 2-fold longer doubling
177 times, respectively. No growth was observed at 30°C and 37°C after 2 weeks under all tested conditions (data
178 not shown). This is in line with the temperatures of low-altitude clouds at the French site from which the strain
179 was isolated, i.e. 5°C mean and 17°C maximal temperature, respectively (Vaitilingom *et al.* 2010). In addition,
180 strain PDD-24b-2 was able to produce ballistospores, a launched spore type specific to basidiomycetes, at 17°C
181 on R2A solid medium after 6 days of culture. Ballistospores are able to launch from an inoculated plate to a

182 neighbouring sterile one, on which colonies will grow following incubation, forming a “mirror” image of the
183 inoculated plate (**Figure 1E**). Ballistosporic basidiospores have been proposed to act as giant cloud condensation
184 nuclei that could increase precipitation by coalescing smaller droplets (Hassett *et al.* 2015). Unlike the strain
185 studied in this work, however, one of the *D. hungarica* strains isolated from terrestrial habitats was unable to
186 produce ballistospores (Takashima *et al.* 2001), suggesting that this trait is not conserved within *D. hungarica*.

187 **Genome sequencing, assembly and completeness**

188 The genome of *D. hungarica* PDD-24b-2 was sequenced by a hybrid strategy using a combination of PacBio
189 (average coverage of 97x, median subread size of 3,584 bp and 474,621 subreads in total) and Illumina (average
190 coverage of 101x; 9,901,968 read pairs of 151 bp) sequencing, yielding a high-quality assembly. The 28 contigs
191 assembled from PacBio subreads were corrected with the Illumina pair-end reads. BLASTn alignment analysis
192 identified two identical contigs which were merged, and 9 small contigs nearly identical to larger ones (between
193 99.96% and 100% identity) which could represent alternative haplotypes and were thus discarded from the final
194 assembly. This yielded a final genome assembly of 18 contigs, with 17 linear contigs corresponding to the
195 nuclear genome. One contig was circular, as evidenced by more than 500 Illumina reads bridging its ends (data
196 not shown), and corresponded to the mitochondrial genome. Its size of 27,226 bp was in close agreement with
197 that estimated for the *D. hungarica* strain CBS 4214^T from average contour-length on electron micrographs
198 twenty years ago (27.3 kbp, Gácsér *et al.* 2002).

199 After assembly, the beginning 5’ third and the remaining 3’ parts of the 18S rRNA gene were located at the
200 termini of two nuclear contigs, i.e. the smallest contig of 2 kbp (contig20) and the contig of 1.11 Mbp (contig11,
201 which also contained the remainder of the rRNA-encoding region). To confirm the linkage between contig20 and
202 contig11, PCR amplification of the 18S rRNA gene was performed using primers targeting contig20 and
203 contig11. The full 18S rRNA gene sequence including the 19 nt gap initially left out of the assembly was
204 sequenced. Accordingly, contig11 was merged with the smaller contig20 to restore a complete 18S rRNA gene
205 within the reunited rRNA region composed of 5S rRNA, 18S rRNA, 5.8S rRNA and 25S rRNA genes. Genome
206 regions with rRNA genes are notoriously difficult to resolve in eukaryotic genomic assemblies, as rRNA genes
207 can be found in tens to thousands tandem copies (Nelson *et al.* 2019). For instance, *Cryptococcus neoformans*,
208 another basidiomycetous yeast, contains around 55 tandem repeats of a single rRNA gene region (Loftus *et al.*
209 2005; Ganley and Kobayashi 2007). The rRNA gene copy number is usually estimated by relative read coverage
210 (Lofgren *et al.* 2019). For strain PDD-24b-2, Illumina read depth coverage of contig20 was 35 times higher than
211 for the rest of the genome (**Figure S2**). As expected, a similar increased coverage was also observed for the

212 terminal part of contig11 in which the rRNA gene cluster is located. This suggests that ribosomal RNA genes are
213 present in about 35 copies in *D. hungarica* PDD-24b-2, although the precise number of tandem repeats remains
214 unknown. To estimate the length of the whole region containing copies of the rRNA gene cluster, we multiplied
215 the rRNA unit length (10.29 kbp) by its relative coverage of 35. The estimated length of the complete contig11
216 would thus be approximately 1.46 Mbp (**Figure 2**).

217 For 7 contigs, T₂AG₃₋₅ telomeric repeats were detected at one of the ends only (**Figure 2**), suggesting incomplete
218 resolution of the nuclear genome. Nevertheless, the statistics and characteristics of the obtained final genome
219 assembly of *D. hungarica* strain PDD-24b-2, with 16 large-scale contigs and a complete mitochondrial genome
220 contig, compare favorably with previously reported genomes for the genus *Dioszegia* (**Table 1**). Specifically,
221 and with an L50 value of 4 and a N50 length of 2.17 Mbp, the genome assembly of *D. hungarica* PDD-24b-2
222 contains no gaps, unlike the three previously sequenced genomes of *Dioszegia* strains (**Table 1**). In particular,
223 the assembly of the *D. cryoxerica* ANT 3-071^T genome was strongly fragmented, with 111 scaffolds under
224 2 kbp, a L50 of 96 and a N50 of 0.12 Mbp. In the case of *D. aurantiaca* JCM 2956^T and *D. crocea* JCM 2961^T,
225 some large and a few small (<2 kbp) scaffolds were reported.

226 The GC content of *D. hungarica* is about 57%, similar to that reported for *D. cryoxerica*, and higher than about
227 53% for *D. aurantiaca* and *D. crocea*. A detailed comparative assembly assessment of the 4 genomes using
228 BUSCO v5.2.2 (Manni *et al.* 2021) was performed using the tremellomycetes_odb10 database. The estimated
229 occurrence of complete genes was similar for the 4 compared *Dioszegia* genomes, yet at about 89.5% instead of
230 the expected 100%. This suggests a lineage specific bias for the *Dioszegia* genus of the reference
231 tremellomycetes_odb10 database. A notable difference between *Dioszegia* genomes is the high percentage of
232 duplicated genes reported for *D. cryoxerica* (53.4%), possibly reflecting unresolved haplotypes in the diploid
233 assembly of its genome (<https://mycocosm.jgi.doe.gov/Dioer1>). This would also be consistent with the twice
234 larger length of its assembly (39.5 Mbp) compared to the 3 other reported genomes including *D. hungarica*.

235 **Genome annotation for protein-coding genes and predicted metabolic pathways**

236 The obtained number of 8,219 predicted protein-coding genes is close to that reported for *D. aurantiaca*
237 JCM 2956^T and *D. crocea* JCM 2961^T (**Table 1**), and also to the total number of unique protein-coding genes of
238 *D. cryoxerica* ANT 3-071^T. The KOG classification scheme was used to evaluate the number of genes involved
239 in cellular processes and signaling (1,301), information storage and processing (1,067), metabolism (1,460), and
240 genes with unknown functions (1,056) (detailed gene ID available at [https://mycocosm.jgi.doe.gov/cgi-](https://mycocosm.jgi.doe.gov/cgi-bin/kogBrowser?type=KOG&db=Diohu1)
241 [bin/kogBrowser?type=KOG&db=Diohu1](https://mycocosm.jgi.doe.gov/cgi-bin/kogBrowser?type=KOG&db=Diohu1)). The largest gene families include transporters from the Major

242 Facilitator Superfamily (138) and sugar transporters (66), protein kinases (102), and clusters of genes with WD
243 domain (100) and helicase-domain (75). Secondary metabolism is represented by three NRPS-like gene clusters
244 and a single PKS-like gene cluster.

245 Gene predictions were analyzed in the light of experimentally characterized metabolic traits in *D. hungarica*
246 (Takashima *et al.* 2001). Genes for glycolysis/gluconeogenesis (39 genes), the TCA cycle (19 genes), starch
247 utilization and production (61 genes), and nitrite utilization (1 nitrite reductase-encoding gene) (KOG
248 classification within the MycoCosm platform) reflect the previously reported utilization by *D. hungarica* of
249 glucose, succinic and citric acid, starch, and nitrite, respectively. Conversely, no genes were predicted for
250 methanol or nitrate utilization, or for thiamine biosynthesis, confirming the reported inability of *D. hungarica* to
251 use methanol or nitrate, and its thiamine auxotrophy. Identified genes for carotenoid biosynthesis (36 putative
252 genes, KEGG annotations, JGI Annotation Pipeline) are in line with previous reports of carotenoids in *Dioszegia*
253 strains (Madhour *et al.* 2005; Amaretti *et al.* 2014; Villarreal *et al.* 2016), and also with the bright orange color
254 culture observed for *D. hungarica* PDD-24b-2 (**Figure 1**). Carotenoids prevent oxidative stress (Madhour *et al.*
255 2005) and act as photo-protectants (Moliné *et al.* 2009) and cryo-protectants (Dieser *et al.* 2010), and may thus
256 favor survival under the harsh conditions of clouds (Šantl-Temkiv *et al.* 2022). In this context, strain PDD-24b-2
257 also encodes a putative antifreeze protein (protein ID: 32937, with a predicted ice-binding protein domain
258 (InterPro entry: IPR021884) and a predicted secretion signal. Secreted antifreeze proteins impair ice crystal
259 formation and protect cell integrity under cold conditions (Hashim *et al.* 2013), suggesting a role of this protein
260 in cold protection of *D. hungarica* in the cloud environment that remains to be experimentally validated .

261 **Transposable elements**

262 A total of 311 putative sequences related to transposable elements (TEs) were detected, and classified in
263 sixteen TE families (**Table S1; File S1**). TEs are dominated by Class I elements representing twelve families. Of
264 those, seven families of *Copia* and one family of *Gypsy* Long Terminal Repeat (LTR) TE were found. Class I
265 non-LTR elements putative families were distributed in three LINE families and one DIRS family. Four families
266 of Class II Terminal Inverted Repeat (TIR) elements were also detected. Only one family encodes a transposase
267 gene carrying a cl24015 domain attributed to MULE TE DDE transposases (Babu *et al.* 2006). The 10 bp long
268 TSD supports an assignation to the *Mutator* Superfamily. Four families of non-autonomous Miniature Inverted-
269 Repeat Transposable Elements (MITE) were also detected. One of them is related to the aforementioned *Mutator*
270 element (same TIR and 10 bp long TSD). The others may be related to the hAT superfamily, according to their
271 TSD length of 8 bp. However, we could not detect the corresponding autonomous copies encoding the

272 transposases to confirm their annotation. In total, putative transposon-related sequences (around 130 kbp)
273 represent 0.63% of the *D. hungarica* PDD-24b-2 genome, among the lowest so far for basidiomycete fungi
274 (Castanera *et al.* 2017). However, reported TE contents are highly variable (ranging between 0.1 and 42 %),
275 possibly also reflecting in part differences in sequence assembly and TE annotation protocols (Castanera *et*
276 *al.* 2017).

277 **Circular mitochondrial genome**

278 This study provides the first complete and circular mitochondrial genome for *D. hungarica*. Organization of the
279 mitochondrial genome of strain PDD-24b-2 differs from that of other *D. hungarica* strains basing on previously
280 reported physical maps (Gácsér *et al.* 2002). This is not unexpected as mitochondrial genome maps differed
281 between *D. hungarica* strains.

282 The mitochondrial genome of strain PDD-24b-2 is smaller (27 kbp) than those of *D. changbaiensis* (35 kbp;
283 Tan and Wang, 2021) and *D. cryoxerica* ANT 03-071^T (36 kbp; L. B. Connell, personal communication) but of
284 similar GC content (40-42%). The PDD-24b-2 mitochondrial genome contains all 15 known core protein-coding
285 genes of mitochondria in Basidiomycetes, 23 tRNAs and 2 rRNAs (**Table 2**).

286 One major difference between the mitochondrial genome of *D. hungarica* and that of *D. changbaiensis*, the only
287 other *Dioszegia* annotated mitochondrial genome to date, is the presence in *D. hungarica* of an additional tRNA
288 gene, *trnR(ucu)* for arginine (**Table 2**). Although similar arginine codon usages are found in both strains, this is
289 not the case for the aga and agg codons that are exclusively found in one of the mitochondrial genome. It is
290 possible that this additional tRNA-Arg(ucu) in *D. hungarica* is used to translate the agg codon (Agris *et al.*
291 2007). On the other hand, in the absence of tRNA-Arg(ucu) the translation of the aga codon in *D. changbaiensis*
292 remains unexplained.

293 A noticeable similarity between the mitochondrial genome of *D. hungarica* and that of *D. changbaiensis* is a
294 high cgu codon usage for arginine (16.2% of arginine codons for *D. hungarica* and 20.8% for *D. changbaiensis*)
295 (**Table 2**). Thus, as this cgu codon cannot be canonically translated by either tRNA-Arg(ucg) or tRNA-Arg(ucu)
296 without post-transcriptional modifications (Phizicky and Hopper 2010), experiments are needed to identify yet
297 unknown modification processes and their roles in translation in *D. hungarica* and *D. changbaiensis*
298 mitochondria.

299 **Phylogenetic analysis and environmental distribution**

300 A phylogenetic tree based on the analysis of the Internal Transcribed Spacer (ITS) region was constructed for 24
301 strains of the genus *Dioszegia*, including strain PDD-24b-2, with 2 strains of the genus *Hannaella*, as outgroups,
302 using the Maximum Likelihood method (**Table S2**). In this tree, *D. hungarica* PDD-24b-2 and the *D. hungarica*
303 type strain are clustered together and distinct from genome-sequenced strains of other *Dioszegia* species
304 (**Figure 3**), in accordance with previous taxonomical studies (Trochine *et al.* 2017; Li *et al.* 2020).

305 Geographical distribution and potential habitat specificity of *D. hungarica* were investigated with a large set of
306 public metagenomes selected to represent a wide diversity of environments, using the ITS region of strain PDD-
307 24b-2 as a query (**Figure 4A**). *D. hungarica* was detected at diverse latitudes around the world (**Figure 4B**), and
308 mostly in aerial biomes. In contrast, representatives of the *Dioszegia* genus were found to be more diversely
309 distributed (**Figure 4C**). Strikingly, ITS sequences specific of *D. hungarica* were not detected in marine samples
310 in our dataset of selected metagenomes, nor in the Mar and TARA Ocean Gene Atlas databases (no fungal hits
311 with over 97% identity were found). This suggests that *D. hungarica* is scarce in open sea environments. On the
312 other hand, *D. hungarica* sequences were not detected in soil metagenomes either (**Figure 4C**). This was
313 surprising since the *Dioszegia hungarica* type strain was isolated from soil (Takashima *et al.* 2001). However,
314 when using the GlobalFungi database, which is a terrestrial soil-focused database, ITS sequences of *D. hungarica*
315 were detected in soil samples. Considering the significant differences in types of sequences between
316 metagenomes (WGS, short raw reads) and the GlobalFungi database (targeted amplification of longer ITS
317 sequences), the stringency of search parameters used in our analysis may contribute to explain this discrepancy,
318 especially in environments with low abundance of *D. hungarica* communities. Nevertheless, the low occurrence
319 of *D. hungarica* in oceans is somewhat paradoxical considering that strain PDD-24b-2 was isolated from a
320 cloud of oceanic origin (**Figure S1**). We thus hypothesize that *D. hungarica* was picked up during air mass
321 travel across France through the puy de Dôme sampling site. As such, the detection of *D. hungarica* in cloud
322 water could serve as an indicator of air mass contact with terrestrial surfaces in future studies where detailed
323 characterization of investigated cloud microbiomes is of interest.

324 In conclusion, the obtained high-quality assembled and annotated genome of the orange-pigmented
325 psychrotrophic yeast *D. hungarica* PDD-24b-2, a major representative of the cloud microbiome, now provides a
326 blueprint for future functional genomics analyses of this environmentally relevant fungus. This will help
327 characterize its mechanisms of resistance to UV radiation (Inacio *et al.* 2005) and of survival in cold

328 environments (Dalluge *et al.* 2019), contribute to develop yeast enzymatic processes at low temperatures (Vaz *et*
329 *al.* 2011), and help to identify and characterize the biotic factors that play a role in cloud chemistry.

330 **Data availability**

331 The *Dioszegia hungarica* PDD-24b-2 Whole Genome Shotgun project was deposited at DDBJ/ENA/GenBank
332 under accession number JAKWFO000000000. The genome version used in this report is JAKWFO010000000.
333 The raw Illumina and PacBio reads were deposited at the Sequence Read Archive under accessions numbers
334 SRR18177991 and SRR18177990 respectively. Details on genome assembly and gene model properties are
335 provided on the MycoCosm genome portal (<https://mycocosm.jgi.doe.gov/Diohu1>). Strain *Dioszegia hungarica*
336 PDD-24b-2 is available upon request to Dr. Pierre Amato or Dr. Françoise Bringel. Representative sequences of
337 putative *D. hungarica* PDD-24b-2 TE families are in File S1. Putative TEs detected in *D. hungarica* PDD-24b-2
338 are in Table S1. ITS sequences used to construct the phylogenetic tree of the *Dioszegia* genus are in Table S2.
339 Environmental samples in which *D. hungarica* and *Dioszegia* species were searched are provided in Table S3.
340 The air mass trajectory of the cloud from which *D. hungarica* strain PDD-24b-2 was isolated is shown in Figure
341 S1. A close-up of the Illumina read depth coverage of *D. hungarica* PDD-24b-2 rRNA gene region is provided
342 in Figure S2. The homolog search scripts, environmental metagenome dataset, as well as more information on
343 biogeographic analyses and TE mining are available at
344 https://github.com/JarrigeD/Dioszegia_hungarica_sequencing.

345 **Acknowledgements**

346 We thank Léa Eck for her help with DNA extraction and Amandine Moreno for physiological tests and
347 microscopy photographs, Dr. Laurie B Connell for giving access to the *D. cryoxerica* ANT 03-071^T genome
348 prior to publication, Dr. Joseph Schacherer for discussion of sequencing strategy and Prof. Hubert Becker for his
349 help with tRNA analysis. This article is dedicated to Dr. Anne-Marie Delort, pioneer of investigations on the role
350 of microorganisms in atmospheric chemistry, on the occasion of her retirement.

351 FB initiated the study. FB and DJ wrote the manuscript. FB, DJ and SV revised the manuscript. MJ and MS
352 cultivated the strain under supervision of PA. FB performed wet lab experiments. SH performed genome
353 annotation on the MycoCosm platform under supervision of IG. DJ performed the phylogenetic analysis and
354 geographic distribution study. DJ performed the transposable element search under the guidance of CBG. TN
355 contributed to discussion. All authors have read, edited and approved the final manuscript.

356 **Conflicts of interest**

357 None Declared.

358 **Funding**

359 This study and the postdoctoral grant to DJ were funded by the French National Agency, grant ANR-19-CE01-
360 0004-02, project METACLOUD. Genome annotation was performed by the U.S. Department of Energy Joint
361 Genome Institute, a DOE Office of Science User Facility, supported by the Office of Science of the U.S.
362 Department of Energy under contract no. DE-AC02-05CH11231.

363

Content	<i>D. hungarica</i> PDD-24b-2	<i>D. aurantiaca</i> JCM 2956 ^T	<i>D. crocea</i> JCM 2961 ^T	<i>D. cryoxerica</i> ANT 03-071 ^T
BioProject accession n°	PRJNA809585	PRJDB3721	PRJDB3718	PRJNA196046
Reference	this study	<i>Takashima et al., 2019</i>	<i>Takashima et al., 2019</i>	<i>(Connell L.B. personal communication)</i>
Sequencing and assembly^a statistics				
Sequencing read coverage depth (technology)	97x (PacBio Sequel) + 101x (Illumina MiSeq)	112x (Illumina HiSeq 2500)	176x (Illumina HiSeq 2500)	98.5x (Illumina HiSeq 2500)
Assembly size (Mbp)	20.96	19.34	20.60	39.52
Scaffolds/contigs	18/18	52/139	26/86	865/1,318
Longest scaffold (Mbp)	5.56	4.12	3.59	0.43
L50 scaffold value	4	5	10	96
N50 scaffold length (Mbp)	2.17	1.28	1.95	0.12
Scaffolds over 2 kbp	17 ^b	44	21	754
GC content (%)	57.2	53.6	53.2	56.9
Gaps (%)	0.00	0.82	0.59	1.3
Linear contigs	16 ^b	139	86	1,318
Mitochondrial genome (kbp)	27 (circular)	NA ^c	NA	36
Annotation statistics				
Gene models	8,219	8,106	8,753	15,948
Average transcript length (bp)	1,538	1,817	1,801	1,415
Average exon/intron length (bp)	247/67	258/59	259/61	264/61
Average exons per gene	6.23	5.90	5.80	5.36
Average protein length (aa)	513	507	500	429
Genes with GO annotations	3,925	NA	NA	6,951
BUSCO v5.2.2 assembly completeness assessment (%)^d				
Complete	89.3	89.0	89.3	90.0
Single	88.8	88.9	88.7	36.6
Duplicated	0.5	0.1	0.6	53.4
Fragmented	3.3	3.9	3.9	3.2
Missing	7.4	7.1	6.8	6.8

365 *Table 1: Genome assembly and predicted annotation data of D. hungarica PDD-24b-2 compared to the three*
366 *other genome-sequenced Dioszegia sp. ^aHaploid assemblies except for D. cryoxerica ANT 03-071^T which is*
367 *probably diploid. ^bScaffolds and linear contigs left after merging contig20 and contig11. ^cNA. Not Available.*
368 *^dBUSCO reference dataset: tremellomycetes_odb10 (2021-06-28).*

Content	<i>D. hungarica</i> PDD-24b-2	<i>D. changbaiensis</i> CGMCC AS 2.2309 ^T
Accession n°	JAKWFO000000000	MT755637
Reference	this study	Tan and Wang, 2021
General statistics		
Size (bp)	27,226	34,853
GC content (%)	40.6	41.9
Mitochondrial genome contents		
Protein coding genes	<i>atp6, atp8, atp9, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad5, nad6, rps3</i>	<i>atp6, atp8, atp9, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad5, nad6, rps3</i>
tRNA	<i>trnA(ugc), trnD(guc), trnE(uuc), trnF(gaa), trnG(ucc), trnH(gug), trnI(gau), trnK(uuu), trnL(uaa), trnL(uag), trnM(cau), trnM(cau), trnN(guu), trnP(ugg), trnQ(uug), trnR(ucg), trnR(ucu), trnS(gcu), trnS(uga), trnT(ugu), trnV(uac), trnW(cca), trnY(gua)</i>	<i>trnA(ugc), trnD(guc), trnE(uuc), trnF(gaa), trnG(ucc), trnH(gug), trnI(gau), trnK(uuu), trnL(uaa), trnL(uag), trnM(cau), trnM(cau), trnN(guu), trnP(ugg), trnQ(uug), trnR(ucg), trnS(gcu), trnS(uga), trnT(ugu), trnV(uac), trnW(cca), trnY(gua)</i>
rRNA	<i>rns, rnl</i>	<i>rns, rnl</i>
Arg codons in mitochondrial CDS	aga: 0, agg: 1, cga: 72, cgc: 0, cgg: 10, cgU: 16	aga: 1, agg: 0, cga: 66, cgc: 1, cgg: 8, cgU: 20

370 *Table 2: Comparison of the mitochondrial genomes of D. hungarica and D. changbaiensis. The additional tRNA*
371 *gene in D. hungarica is underlined.*

372

373 **Figure legends**

374 **Fig. 1. Morphological and growth characteristics of *D. hungarica* strain PDD-24b-2.** (A) Cellular
375 morphology observed with a Leica DM4000 B microscope at 1000x magnification after growth in YM broth at
376 17°C. (B) Colonies on R2A solid medium at 17°C. (C) Liquid cultures after growth at different temperatures for
377 4 days. Highest cell density was observed at 17°C. No growth was observed above 25°C or in sterile YM
378 medium. (D) Effect of culture incubation temperatures on growth rate (μ). For each medium (YM or R2A), the
379 mean growth rate was the mean of two biological replicates. (E) Ballistospore production on solid R2A
380 medium. The inoculated plate (left) was placed on top of the uninoculated one (right). After 6 days of
381 incubation, colonies also appeared on the bottom plate as a partial mirror image of the top inoculated plate.

382 **Fig. 2. Telomeric sequences distribution in assembled contigs of *D. hungarica* PDD-24b-2.** Detected
383 telomeric T₂AG₃₋₅ repeats are highlighted in orange (not drawn to scale). The complete ribosomal RNA gene
384 region (with an estimated 35 copies of the ribosomal RNA gene cluster) is highlighted in blue. Identification
385 labels for assembled contigs are given as in MycoCosm (<https://mycocosm.jgi.doe.gov/Diohu1>).

386 **Fig. 3. Phylogenetic analysis of the *Dioszegia* genus based on the internal transcribed spacer (ITS) region.**
387 The tree was obtained from a sequence alignment of 462 nt of the ITS region with the Maximum Likelihood
388 method and General Time Reversible model. Branch lengths (number of substitutions per site) are indicated in
389 black, bootstrap support values (percentage of replicate trees in which the associated taxa clustered together in
390 the bootstrap test 500 replicates) in red. Sequence accession numbers are indicated in brackets. Diamonds (◆)
391 indicate strains for which a draft genome is available. *Hannaella sinensis* and *Hannaella luteola* were used as
392 outgroups.

393 **Fig. 4. Geographical and environmental distribution of the *Dioszegia* genus and *D. hungarica* species.**
394 (A) Environmental metagenome exploration pipeline. The megablast hits filtering process is indicated with
395 unfiltered hits (in black), filtered hits representing *D. hungarica* (in orange) or *Dioszegia* sp. sequences (in
396 purple) mapped on *D. hungarica* PDD-24b-2 ITS region. (B) Geographical distribution of sequences assigned to
397 *D. hungarica* or *Dioszegia* sp. in environmental metagenomic datasets (see **Table S3** for detailed information).
398 (C) Environmental distribution of *D. hungarica* and *Dioszegia* sp. compared to that of the complete metagenomic
399 dataset.

400

401

402 **Literature Cited**

- 403 Agler, M. T., J. Ruhe, S. Kroll, C. Morhenn, S.-T. Kim *et al.*, 2016 Microbial hub taxa link host and abiotic
404 factors to plant microbiome variation (M. K. Waldor, Ed.). *PLoS Biol.* 14: e1002352.
- 405 Agris, P. F., F. A. P. Vendeix, and W. D. Graham, 2007 tRNA's wobble decoding of the genome: 40 years of
406 modification. *J. Mol. Biol.* 366: 1–13.
- 407 Amaretti, A., M. Simone, A. Quartieri, F. Masino, S. Raimondi *et al.*, 2014 Isolation of carotenoid-producing
408 yeasts from an alpine glacier. *Chem. Eng. Trans.* 38: 217–222.
- 409 Amato, P., M. Joly, L. Besaury, A. Oudart, N. Taib *et al.*, 2017 Active microorganisms thrive among extremely
410 diverse communities in cloud water (G. Moreno-Hagelsieb, Ed.). *PLoS ONE* 12: e0182869.
- 411 Andrews, S. *FastQC*. Babraham Bioinformatics.
- 412 Babu, M. M., L. M. Iyer, S. Balaji, and L. Aravind, 2006 The natural history of the WRKY–GCM1 zinc fingers
413 and the relationship between transcription factors and transposons. *Nucleic Acids Res.* 34: 6505–6520.
- 414 Bailly-Bechet, M., A. Haudry, and E. Lerat, 2014 “One code to find them all”: a perl tool to conveniently parse
415 RepeatMasker output files. *Mob. DNA* 5: 13.
- 416 Bell, E. A., C. L. Butler, C. Oliveira, S. Marburger, L. Yant *et al.*, 2022 Transposable element annotation in
417 non-model species: The benefits of species-specific repeat libraries using semi-automated EDTA
418 and DeepTE *de novo* pipelines. *Mol. Ecol. Resour.* 22: 823–833.
- 419 Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and
420 applications. *BMC Bioinformatics* 10: 421.
- 421 Carradec, Q., E. Pelletier, C. Da Silva, A. Alberti, Y. Seeleuthner *et al.*, 2018 A global ocean atlas of eukaryotic
422 genes. *Nat. Commun.* 9: 373.
- 423 Castanera, R., A. Borgognone, A. G. Pisabarro, and L. Ramírez, 2017 Biology, dynamics, and applications of
424 transposable elements in basidiomycete fungi. *Appl. Microbiol. Biotechnol.* 101: 1337–1350.
- 425 Connell, L. B., R. Redman, R. Rodriguez, A. Barrett, M. Iszard *et al.*, 2010 *Dioszegia antarctica* sp. nov. and
426 *Dioszegia cryoxerica* sp. nov., psychrophilic basidiomycetous yeasts from polar desert soils in
427 Antarctica. *Int. J. Syst. Evol. Microbiol.* 60: 1466–1472.
- 428 Dalluge, J. J., E. C. Brown, and L. B. Connell, 2019 Toward a rapid method for the study of biodiversity in cold
429 environments: the characterization of psychrophilic yeasts by MALDI-TOF mass spectrometry.
430 *Extremophiles* 23: 461–466.

431 Delmont, T. O., M. Gaia, D. D. Hinsinger, P. Frémont, C. Vanni *et al.*, 2022 Functional repertoire convergence
432 of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genomics* 2: 100123.

433 Delort, A. M., M. Vařtilingom, M. Joly, P. Amato, N. Wirgot *et al.*, 2017 Clouds: a transient and stressing
434 habitat for microorganisms., pp. 215–245 in *Microbial Ecology of Extreme Environments*, edited by C.
435 Chénard and F. M. Lauro. Springer International Publishing, Cham.

436 Dieser, M., M. Greenwood, and C. M. Foreman, 2010 Carotenoid pigmentation in Antarctic heterotrophic
437 bacteria as a strategy to withstand environmental stresses. *Arct. Antarct. Alp. Res.* 42: 396–405.

438 Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic
439 Acids Res.* 32: 1792–1797.

440 Edman, J. C., 1992 Isolation of telomere-like sequences from *Cryptococcus neoformans* and their use in high-
441 efficiency transformation. *Mol. Cell. Biol.* 12: 7.

442 Ferreira, E. M. S., F. M. P. de Sousa, L. H. Rosa, and R. S. Pimenta, 2019 Taxonomy and richness of yeasts
443 associated with angiosperms, bryophytes, and meltwater biofilms collected in the Antarctic peninsula.
444 *Extremophiles* 23: 151–159.

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

447 Gácsér, A., Z. Hamari, I. Pfeiffer, J. Litter, F. Kevei *et al.*, 2002 Organization of mitochondrial DNA in the
448 basidiomycetous *Dioszegia hungarica* (*Cryptococcus hungaricus*) species. *FEMS Microbiol. Lett.* 212:
449 1–6.

450 Ganley, A. R. D., and T. Kobayashi, 2007 Highly efficient concerted evolution in the ribosomal DNA repeats:
451 total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.* 17: 184–
452 191.

453 de García, V., S. Brizzio, D. Libkind, P. Buzzini, and M. Van Broock, 2007 Biodiversity of cold-adapted yeasts
454 from glacial meltwater rivers in Patagonia, Argentina: yeasts from Patagonian glacial waters. *FEMS
455 Microbiol. Ecol.* 59: 331–341.

456 Grigoriev, I. V., R. Nikitin, S. Haridas, A. Kuo, R. Ohm *et al.*, 2014 MycoCosm portal: gearing up for 1000
457 fungal genomes. *Nucleic Acids Res.* 42: D699–D704.

458 Haridas, S., A. Salamov, and I. V. Grigoriev, 2018 Fungal Genome Annotation, pp. 171–184 in *Fungal
459 Genomics*, edited by R. P. de Vries, A. Tsang, and I. V. Grigoriev. *Methods in Molecular Biology*,
460 Springer New York, New York, NY.

461 Hashim, N. H. F., I. Bharudin, D. L. S. Nguong, S. Higa, F. D. A. Bakar *et al.*, 2013 Characterization of Afp1,
462 an antifreeze protein from the psychrophilic yeast *Glaciozyma antarctica* PI12. *Extremophiles* 17: 63–
463 73.

464 Hassett, M. O., M. W. F. Fischer, and N. P. Money, 2015 Mushrooms as rainmakers: how spores act as nuclei
465 for raindrops (A. S. Gladfelter, Ed.). *PLoS ONE* 10: e0140407.

466 Ianiri, G., R. Abhyankar, A. Kihara, and A. Idnurm, 2014 Phs1 and the synthesis of very long chain fatty acids
467 are required for ballistospore formation (J.-H. Yu, Ed.). *PLoS ONE* 9: e105147.

468 Inacio, J., L. Portugal, I. Spencer-Martins, and A. Fonseca, 2005 Phylloplane yeasts from Portugal: seven novel
469 anamorphic species in the Tremellales lineage of the Hymenomycetes (Basidiomycota) producing
470 orange-coloured colonies. *FEMS Yeast Res.* 5: 1167–1183.

471 Karlsson, I., H. Friberg, A.-K. Kolseth, C. Steinberg, and P. Persson, 2017 Organic farming increases richness of
472 fungal taxa in the wheat phyllosphere. *Mol. Ecol.* 26: 3424–3436.

473 Kinkel, L. L., 1997 Microbial population dynamics on leaves. *Annu. Rev. Phytopathol.* 35: 327–347.

474 Klemetsen, T., I. A. Raknes, J. Fu, A. Agafonov, S. V. Balasundaram *et al.*, 2018 The MAR databases:
475 development and implementation of databases specific for marine metagenomics. *Nucleic Acids Res.*
476 46: D692–D699.

477 Kolmogorov, M., J. Yuan, Y. Lin, and P. A. Pevzner, 2019 Assembly of long, error-prone reads using repeat
478 graphs. *Nat. Biotechnol.* 37: 540–546.

479 Kuo, A., B. Bushnell, and I. V. Grigoriev, 2014 Fungal genomics, pp. 1–52 in *Advances in Botanical Research*,
480 Elsevier.

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

482 Langmead, B., C. Wilks, V. Antonescu, and R. Charles, 2019 Scaling read aligners to hundreds of threads on
483 general-purpose processors (J. Hancock, Ed.). *Bioinformatics* 35: 421–432.

484 Li, A.-H., F.-X. Yuan, M. Groenewald, K. Bensch, A. M. Yurkov *et al.*, 2020 Diversity and phylogeny of
485 basidiomycetous yeasts from plant leaves and soil: proposal of two new orders, three new families,
486 eight new genera and one hundred and seven new species. *Stud. Mycol.* 96: 17–140.

487 Liu, X.-Z., Q.-M. Wang, M. Göker, M. Groenewald, A. V. Kachalkin *et al.*, 2015 Towards an integrated
488 phylogenetic classification of the *Tremellomycetes*. *Stud. Mycol.* 82: 1–21.

489 Lofgren, L. A., J. K. Uehling, S. Branco, T. D. Bruns, F. Martin *et al.*, 2019 Genome-based estimates of fungal
490 rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Mol. Ecol.* 28: 721–
491 730.

492 Loftus, B. J., E. Fung, P. Roncaglia, D. Rowley, P. Amedeo *et al.*, 2005 The genome of the basidiomycetous
493 yeast and human pathogen *Cryptococcus neoformans*. *Science* 307: 1321–1324.

494 Madhour, A., H. Anke, A. Mucci, P. Davoli, and R. W. S. Weber, 2005 Biosynthesis of the xanthophyll
495 plectanixanthin as a stress response in the red yeast *Dioszegia* (Tremellales, Heterobasidiomycetes,
496 Fungi). *Phytochemistry* 66: 2617–2626.

497 Manni, M., M. R. Berkeley, M. Seppey, F. A. Simão, and E. M. Zdobnov, 2021 BUSCO update: novel and
498 streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic,
499 prokaryotic, and viral genomes (J. Kelley, Ed.). *Mol. Biol. Evol.* 38: 4647–4654.

500 Martin, M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*
501 17: 10–12.

502 Moliné, M., D. Libkind, M. del Carmen Diéguez, and M. van Broock, 2009 Photoprotective role of carotenoids
503 in yeasts: response to UV-B of pigmented and naturally-occurring albino strains. *J. Photochem.*
504 *Photobiol.* B 95: 156–161.

505 Nelson, J. O., G. J. Watase, N. Warsinger-Pepe, and Y. M. Yamashita, 2019 Mechanisms of rDNA copy number
506 maintenance. *Trends Genet.* 35: 734–742.

507 Phizicky, E. M., and A. K. Hopper, 2010 tRNA biology charges to the front. *Genes Dev.* 24: 1832–1860.

508 Priyam, A., B. J. Woodcroft, V. Rai, I. Moghul, A. Munagala *et al.*, 2019 Sequenceserver: A modern graphical
509 user interface for custom BLAST databases. *Mol. Biol. Evol.* 36: 2922–2924.

510 Reasoner, D. J., and E. E. Geldreich, 1985 A new medium for the enumeration and subculture of bacteria from
511 potable water. *Appl. Environ. Microbiol.* 49: 1–7.

512 Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander *et al.*, 2011 Integrative genomics
513 viewer. *Nat. Biotechnol.* 29: 24–26.

514 Salazar, G., L. Paoli, A. Alberti, J. Huerta-Cepas, H.-J. Ruscheweyh *et al.*, 2019 Gene expression changes and
515 community turnover differentially shape the global ocean metatranscriptome. *Cell* 179: 1068–1083.

516 Šantl-Temkiv, T., P. Amato, E. O. Casamayor, P. K. H. Lee, and S. B. Pointing, 2022 Microbial ecology of the
517 atmosphere. *FEMS Microbiol. Rev.* 46: fuac009.

518 Sapkota, R., L. N. Jørgensen, and M. Nicolaisen, 2017 Spatiotemporal variation and networks in the mycobiome
519 of the wheat canopy. *Front. Plant Sci.* 8: 1357.

520 Schmieder, R., and R. Edwards, 2011 Quality control and preprocessing of metagenomic datasets.
521 *Bioinformatics* 27: 863–864.

522 Stein, A. F., R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen *et al.*, 2015 NOAA’s HYSPLIT
523 Atmospheric Transport and Dispersion Modeling System. *Bull. Am. Meteorol. Soc.* 96: 2059–2077.

524 Takashima, M., T. Deak, and T. Nakase, 2001 Emendation of *Dioszegia* with redescription of *Dioszegia*
525 *hungarica* and two new combinations, *Dioszegia aurantiaca* and *Dioszegia crocea*. *J. Gen. Appl.*
526 *Microbiol.* 47: 75–84.

527 Takashima, M., R. Manabe, and M. Ohkuma, 2019 Draft genome sequences of basidiomycetous epiphytic
528 phylloplane yeast type strains *Dioszegia crocea* JCM 2961 and *Dioszegia aurantiaca* JCM 2956 (V.
529 Bruno, Ed.). *Microbiol. Resour. Announc.* 8: e01727-18.

530 Tamura, K., G. Stecher, and S. Kumar, 2021 MEGA11: Molecular Evolutionary Genetics Analysis version 11
531 (F. U. Battistuzzi, Ed.). *Mol. Biol. Evol.* 38: 3022–3027.

532 Tan, M., and Q. Wang, 2021 Characterization of the complete mitochondrial genome of *Dioszegia*
533 *changbaiensis* (Tremellales: *Bulleribasidiaceae*) with phylogenetic implications. *Mitochondrial DNA*
534 *Part B Resour.* 6: 3315–3317.

535 Tavaré, S., 1986 Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life*
536 *Sci.* 17:.

537 Trochine, A., B. Turchetti, A. B. M. Vaz, L. Brandao, L. H. Rosa *et al.*, 2017 Description of *Dioszegia*
538 *patagonica* sp. nov., a novel carotenogenic yeast isolated from cold environments. *Int. J. Syst. Evol.*
539 *Microbiol.* 67: 4332–4339.

540 Vařtilingom, M., P. Amato, M. Sancelme, P. Laj, M. Leriche *et al.*, 2010 Contribution of microbial activity to
541 carbon chemistry in clouds. *Appl. Environ. Microbiol.* 76: 23–29.

542 Vařtilingom, M., E. Attard, N. Gaiani, M. Sancelme, L. Deguillaume *et al.*, 2012 Long-term features of cloud
543 microbiology at the puy de Dôme (France). *Atmos. Environ.* 56: 88–100.

544 Vařtilingom, M., L. Deguillaume, V. Vinatier, M. Sancelme, P. Amato *et al.*, 2013 Potential impact of microbial
545 activity on the oxidant capacity and organic carbon budget in clouds. *Proc. Natl. Acad. Sci.* 110: 559–
546 564.

547 Vaz, A. B. M., L. H. Rosa, M. L. A. Vieira, V. de Garcia, L. R. Brandão *et al.*, 2011 The diversity, extracellular
548 enzymatic activities and photoprotective compounds of yeasts isolated in Antarctica. *Braz. J. Microbiol.*
549 42: 937–947.

550 Větrovský, T., D. Morais, P. Kohout, C. Lepinay, C. Algora *et al.*, 2020 GlobalFungi, a global database of
551 fungal occurrences from high-throughput-sequencing metabarcoding studies. *Sci. Data* 7: 228.

552 Villar, E., T. Vannier, C. Vernet, M. Lescot, M. Cuenca *et al.*, 2018 The Ocean Gene Atlas: exploring the
553 biogeography of plankton genes online. *Nucleic Acids Res.* 46: W289–W295.

554 Villarreal, P., M. Carrasco, S. Barahona, J. Alcaíno, V. Cifuentes *et al.*, 2016 Tolerance to ultraviolet radiation
555 of psychrotolerant yeasts and analysis of their carotenoid, mycosporine, and ergosterol content. *Curr.*
556 *Microbiol.* 72: 94–101.

557 Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel *et al.*, 2014 Pilon: an integrated tool for comprehensive
558 microbial variant detection and genome assembly improvement (J. Wang, Ed.). *PLoS ONE* 9: e112963.

559 Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy *et al.*, 2007 A unified classification system for
560 eukaryotic transposable elements. *Nat. Rev. Genet.* 8: 973–982.

561 Yang, Z., 1996 Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* 11: 367–
562 372.

563