1 **Draft genome sequences of three monokaryotic isolates of the white-rot basidiomycete**

2 **fungus *Dichomitus squalens***

3 Sara Casado López[a], Mao Peng[a], Paul Daly[a], Bill Andreopoulos[b], Jasmyn Pangilinan[b], Anna

4 Lipzen[b], Robert Riley[b], Steven Ahrendt[b], Vivian Ng[b], Kerrie Barry[b], Chris Daum[b], Igor V.

5 Grigoriev[b], Kristiina S. Hildén[c], Miia R. Mäkelä[c#], Ronald P. de Vries[a,c#],

6

7 [a]Fungal Physiology, Westerdijk Fungal Biodiversity Centre & Fungal Molecular Physiology,

8 Utrecht University, Utrecht, The Netherlands

9 [b]US Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA

10 [c]Fungal Genetics and Biotechnology, Department of Microbiology, University of Helsinki,

11 Helsinki, Finland

12

13 Running title: *Dichomitus squalens* genome sequences

14

15 #Address correspondence to Ronald de Vries, r.devries@westerdijkinstitute.nl, Miia Mäkelä,

16 miia.r.makela@helsinki.fi

17

18

19

20

21

Abstract

Here we report the draft genome sequences of three isolates of the wood white-rotting basidiomycete fungus *Dichomitus squalens*. These monokaryons were sequenced to provide more information on intraspecies genomic diversity of this fungus, also in relation to the previously sequenced genome of *D. squalens* LYAD-421 SS1.

Main text

Dichomitus squalens is a wood-decaying white-rot fungus commonly found in Europe, Asia and North America (1). It is mainly found on softwoods (2, 3), and has an extensive repertoire of lignocellulose-degrading enzymes (4-6). Two of the sequenced monokaryons, CBS463.89 and CBS 464.89, are derived from the well-studied Polish dikaryon FBCC312 (CBS432.34) (4, 6-11), while OM18370.1 is derived from the Finnish dikaryon OM18370 (CBS139088).

Strains were maintained on 2% (wt/vol) malt extract (ME) 1.5% (wt/vol) agar plates, from which four plugs (ø 5 mm) were used to inoculate stationary 50 mL 2% (wt/vol) ME liquid cultures, which were incubated at 28°C for five days. Genomic DNA was extracted from homogenized mycelium with extraction buffer (2% CTAB [N-cetyl-N,N,N-trimethylammonium bromide], 100 mM Tris–HCl, 1.4 M NaCl, 20 mM EDTA, 0.2% β-mercaptoethanol) and purified with chloroform–isoamyl alcohol (24:1) (12). For RNA extraction, the isolates were pre-cultured on glycerol for 7 days (28°C) and transferred to solid-state cultures containing 2 g (dry weight) of Norway spruce wood sticks (2×0.2×0.2 cm) on top of 1% (wt/vol) water agar at 28°C for two

42   and four weeks (4). RNA extracts were layered over a 2 mL CsCl solution (5.7 M CsCl (Serva,

43   Germany); 25 mM sodium citrate, pH 7.0; 0.5% *N*-lauroylsarcosine (Sigma, USA); 0.1 M β-

44   mercaptoethanol (Sigma, USA)) in 13.2 mL polyallomer ultracentrifuge tubes (Beckman-

45   Coulter, Brea, CA, USA) and centrifuged at 33,000 rpm for 21 h at 4°C in a Optima L-90 K

46   ultracentrifuge, using the SW-41 Ti swinging bucket rotor (Beckman-Coulter, Brea, CA, USA).

47   After centrifugation, supernatant was removed by, the tube was inverted and all but the bottom 1

48   cm was sheared off. The RNA in the clear pellet was rinsed with 100 µL of DEPC-treated water

49   and then dissolved in 50 µL of DEPC-treated water and stored at -80°C (13). The genomes were

50   sequenced using Illumina platform and pairs of standard fragment (300 bp) and 4 Kbp long mate

51   pair (LMP) libraries. Fragment libraries were produced from 100 ng gDNA sheared to 300 bp

52   using the Covaris LE220 and size-selected using SPRI beads (Beckman Coulter). The fragments

53   were treated with end-repair, A-tailing, and ligation of Illumina compatible adapters (IDT, Inc)

54   using the Illumina library creation kit (KAPA biosystems). For LMP, 5 µg of DNA was sheared

55   using the Covaris g-TUBE™ (Covaris) and gel size selected for 4kb. The sheared DNA was

56   treated with end repair and ligated with biotinylated loxP adapters and circularized by a Cre

57   excision reaction (NEB). These were randomly sheared and treated as indicated for the fragment

58   library, and enriched using eight PCR cycles for the final library.

59   For transcriptomes, which were used for genome annotations, stranded cDNA libraries were

60   generated using the Illumina Truseq Stranded RNA LT kit. mRNA was purified using magnetic

61   beads containing poly-T oligos, fragmented and reversed transcribed using random hexamers and

62   SSII (Invitrogen) followed by second strand synthesis, and then treated with end-pair, A-tailing,

63   adapter ligation, and eight PCR cycles.

64 The prepared libraries were quantified using KAPA Biosystem's next-generation sequencing

65 library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified

66 libraries were then multiplexed with other libraries, and the pool of libraries was then prepared

67 for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster

68 kit, v4, and Illumina's cBot instrument to generate a clustered flow cell for sequencing.

69 Sequencing of the flow cell was performed on the Illumina HiSeq2500 sequencer using HiSeq

70 TruSeq SBS sequencing kits, v4, following a 2x150bp (2x100bp for LMP) indexed run recipe

71 (14).

72 Illumina fastq files were QC filtered for artifact/process contamination. DNA reads were

73 assembled AllPathsLG v. R49403 (15). For CBS463.89 lacking LMP, the initial assemblies of

74 fragment data with Velvet v.1.2.07 (16) were used to create *in silico* long mate-pair libraries with

75 insert 3000 +/- 300 bp. RNA reads were assembled using Rnnotator v.3.4.0 (17). All three

76 genomes were annotated using the JGI Annotation pipeline v.1.9, which combines several ab

77 initio, homology-based, and transcriptome-based gene predictors as well as tools and databases

78 for functional annotation (18, 19).

79 All four genomes are highly similar in genome size and characteristics (Table 1). The

80 improvement in sequencing methodology is reflected in the lower contig and gap number of the

81 three new genomes compared to the older genome (LYAD-421 SS1). This data is highly useful

82 to evaluate intraspecies genome variation in *D. squalens*.

83

84 **Data availability**

85 Genome assemblies and annotations are available via MycoCosm (jgi.doe.gov/fungi; (18)). The

86 data is deposited at DDBJ/EMBL/GenBank under BioProject/GenBank Accessions:

98   **References**

99   1.   Andrews SR, Gill LS. 1943. Western red rot in immature Ponderosa pine in the
100   Southwest. J Forest 41:565–573.

101   2.   Blanchette RA, Otjen L, Carlson MC. 1987. Lignin distribution in cell walls of birch
102   wood decayed by white rot basidiomycetes. Phytopathol 77:684-690.

103   3.   Renvall P, Renvall T, Niemelä T. 1991. Basidiomycetes at the timberline in Lapland 2.
104   An annotated checklist of the polypores of northeastern Finland. Karstenia 31:13-28.

105   4.   Daly P, Casado López S, Peng M, Lancefield CS, Purvine SO, Kim YM, Zink EM,
106   Dohnalkova A, Singan VR, Lipzen A, Dilworth D, Wang M, Ng V, Robinson E, Orr G,
107   Baker SE, Bruijnincx PCA, Hilden KS, Grigoriev IV, Mäkelä MR, de Vries RP. 2018.

108   *Dichomitus squalens* partially tailors its molecular responses to the composition of solid

109   wood. Environ Microbiol 20:4141-4156.

110 5.   Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martinez AT,

111   Otillar R, Spatafora JW, Yadav JS, Aerts A, Benoit I, Boyd A, Carlson A, Copeland A,

112   Coutinho PM, de Vries RP, Ferreira P, Findley K, Foster B, Gaskell J, Glotzer D,

113   Gorecki P, Heitman J, Hesse C, Hori C, Igarashi K, Jurgens JA, Kallen N, Kersten P,

114   Kohler A, Kües U, Kumar TK, Kuo A, LaButti K, Larrondo LF, Lindquist E, Ling A,

115   Lombard V, Lucas S, Lundell T, Martin R, McLaughlin DJ, Morgenstern I, Morin E,

116   Murat C, Nagy LG, Nolan M, Ohm RA, Patyshakuliyeva A, et al. 2012. The Paleozoic

117   origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science

118   336:1715-1719.

119 6.   Rytioja J, Hilden K, Di Falco M, Zhou MM, Aguilar-Pontes MV, Sietiö O-M, Tsang A,

120   de Vries RP, Mäkelä MR. 2017. The molecular response of the white-rot fungus

121   *Dichomitus squalens* to wood and non-woody biomass as examined by transcriptome and

122   exoproteome analyses. Environ Microbiol 19:1237-1250.

123 7.   Casado López S, Peng M, Issak TY, Daly P, de Vries RP, Mäkelä MR. 2018. Induction

124   of plant cell wall degrading CAZyme encoding genes by lignocellulose-derived

125   monosaccharides and cellobiose in the white-rot fungus *Dichomitus squalens*. Appl

126   Environ Microbiol 84:e00403-18.

127 8.   Casado López S, Theelen B, Manserra S, Issak TY, Rytioja J, Mäkelä MR, de Vries RP.

128   2017. Functional diversity in *Dichomitus squalens* monokaryons. IMA Fungus 8:17-25.

129 9. Mäkelä MR, Sietiö O-M, de Vries RP, Timonen S, Hildén K. 2014. Oxalate-metabolising genes of the white-rot fungus *Dichomitus squalens* are differentially induced on wood and at high proton concentration. PLoS One 9:e87959.

132 10. Rytioja J, Hilden K, Hatakka A, Mäkelä MR. 2014. Transcriptional analysis of selected cellulose-acting enzymes encoding genes of the white-rot fungus *Dichomitus squalens* on spruce wood and microcrystalline cellulose. Fungal Genet Biol 72:91-98.

135 11. Rytioja J, Hilden K, Mäkinen S, Vehmaanperä J, Hatakka A, Mäkelä MR. 2015. Saccharification of lignocelluloses by carbohydrate active enzymes of the white rot fungus *Dichomitus squalens*. PLoS One 10:e0145166.

138 12. Hildén K, Martínez AT, Hatakka A, Lundell T. 2005. The two manganese peroxidases Pr-MnP2 and Pr-MnP3 of *Phlebia radiata*, a lignin-degrading basidiomycete, are phylogenetically and structurally divergent. Fungal Genet Biol 42:403-419.

141 13. Patyshakuliyeva A, Mäkelä MR, Sietiö O-M, de Vries RP, Hildén KS. 2014. An improved and reproducible protocol for the extraction of high quality fungal RNA from plant biomass substrates. Fungal Genet Biol 72:201-206.

144 14. Walker C, Ryu S, Na H, Zane M, LaButti K, Lipzen A, Haridas S, Barry K, Grigoriev IV, Quarterman J, Slininger P, Dien B, Trinh CT. 2018. Draft genome assemblies of five robust *Yarrowia lipolytica* strains exhibiting high lipid production, pentose sugar utilization, and sugar alcohol secretion from undetoxified lignocellulosic biomass hydrolysates. Microbiol Resour Announc 7:e01040-18.

149 15. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of

152 mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A

153 108:1513-1518.

154 16. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de

155 Bruijn graphs. Genome Res 18:821-9.

156 17. Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T, Sherlock G, Snyder M, Wang

157 Z. 2010. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded

158 RNA-Seq reads. BMC Genomics 11:663.

159 18. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao

160 X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. 2014. MycoCosm

161 portal: gearing up for 1000 fungal genomes. Nucleic Acids Res 42:D699-704.

162 19. Haridas S, Salamov A, Grigoriev IV. 2018. Fungal Genome Annotation. Methods Mol

163 Biol. 1775:171-184.

164

165

166

167   **Table 1.** Genome statistics of the three *D. squalens* genomes CBS463.89, CBS464.89 and

168   OM18370.1 compared to the previously sequenced genome of LYAD-421 SS1 (5).

	CBS463.89	CBS464.89	OM18370.1	LYAD-421 SS1 (5)
Genome assembly size (Mbp)	36.87	39.60	39.32	42.75
Read coverage depth	145x	118.7x	100.8x	50.63x
# of reads sequenced	42.6M	32.7M	38.3M	7.7M
# of contigs	1373	1147	1126	2852
# of scaffolds	1259	467	439	542
Scaffold N50	134	44	39	16
Scaffold L50 (Mbp)	0.08	0.22	0.27	0.64
# of gaps	114	680	687	1155
% of scaffold length in gaps	0.2%	2.5%	2.6%	7.7%
Gene length (avg/median, bp)	1,691/1,437	1,678/1,425	1,694/1,449	1,890/1,562
Transcript length (avg/median, bp)	1,370/1,140	1,358/1,128	1,365/1,150	1,484/1,213
Exon length (avg/median, bp)	259/158	259/158	256/157	254/152
Intron length (avg/median, bp)	76/61	77/60	78/61	86/61
Protein length (avg/median, aa)	387/314	382/311	388/319	419/345
# exons per gene (avg/median)	5.3/4	5.25/4	5.34/4	5.84/4
# of gene models	14,946	15,295	14,950	12,290
GC content, %	55.7	55.6	55.6	55.6

169