

Figure S1 Total assembly size and predicted gene number in fungi.

The *Rhizophagus irregularis* genomes (our RIR17, and two previously assembled genomes, Lin14 and Tis13) and 768 genomes registered on GenBank database. The fungal assembly statistics were obtained from the registered information on GenBank (ftp://ftp.ncbi.nlm.nih.gov/genomes/ASSEMBLY_REPORTS/All/).

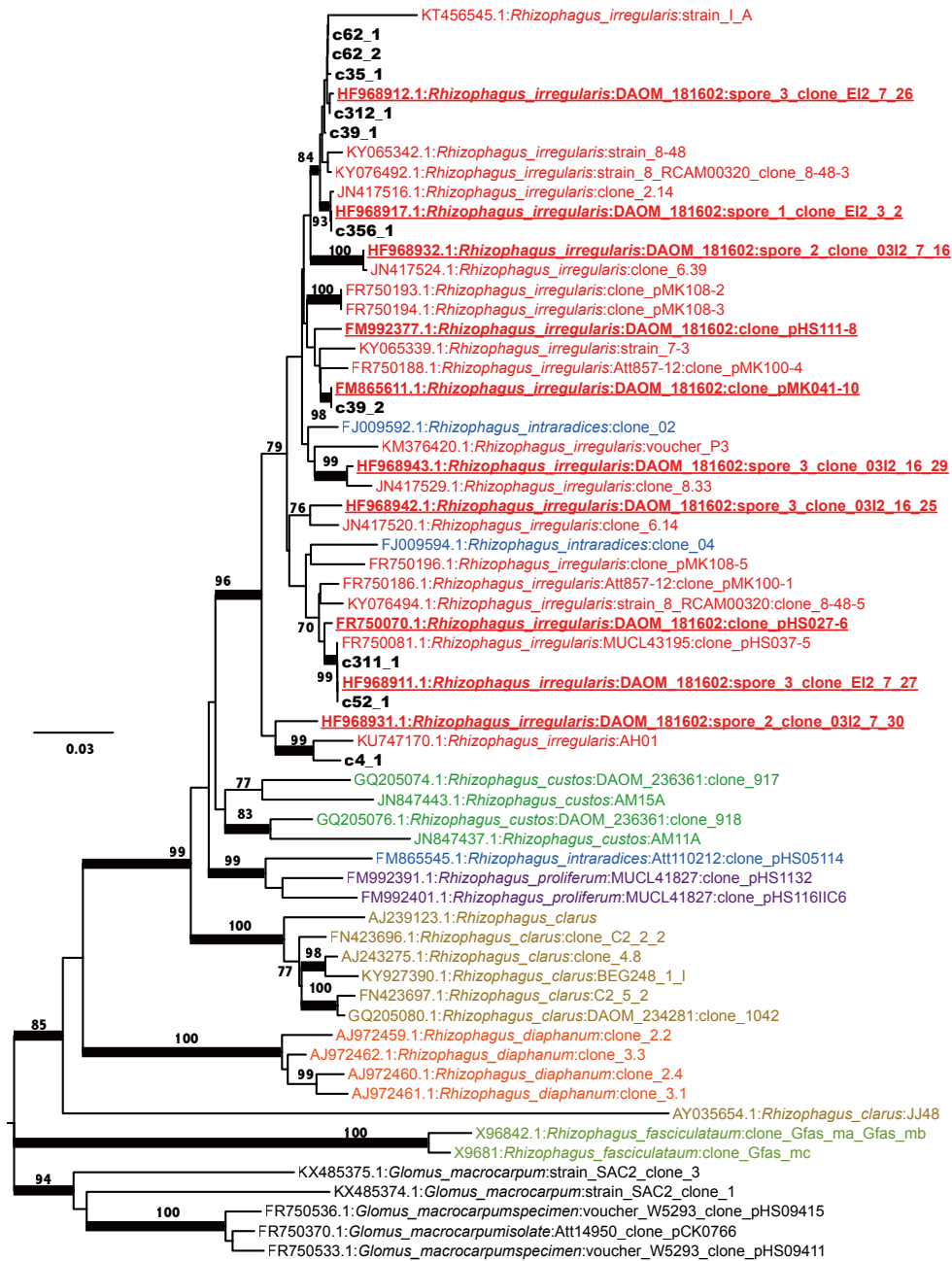


Figure S2. NJ tree based on 586 positions of 48S rDNA.

Partial 18S, ITS1, 5.8S, ITS2 and partial 28S rDNAs were used. The ten rDNA paralogs from RIR17 and 58 *Rhizopagus* sequences from DDBJ were chosen as OTU. The 58 *Rhizopagus* sequences were selected from 329 OTUs on DDBJ (209 OTUs for DAOM-181602, 57 OTUs for other *R. irregularis* strains, and 63 OTUs for other *Rhizopagus* species), using CD-Hit clustering (-c 0.98 -n 5). Five *Glomus* sequences were used as out group OTUs. Red lined OTUs are sequences from *R. irregularis* DAOM-181602, and other red OTUs are data from other strains of *R. irregularis*. Nodes supported by over 80 bootstrap value are marked by a bold line. All *R. irregularis* OTUs made a single clade with *Rhizopagus intraradices* that is a morphologically non-distinct sister group of *R. irregularis*.

Table S1 Summary of the sequencing data

Sequence ID	DRA ID	Libraly construction and sequencing method	DNA/RNA	Read #	Read length	Total basepair (Mbp)
Rir_DNA_Pac_1	DRA004849	PacBio RSII, P6-C4 sequence chemistry	DNA	348,539	*7559	2,635
Rir_DNA_Pac_2	DRA004878	PacBio RSII, P6-C4 sequence chemistry	DNA	394,261	*8562	3,376
Rir_DNA_Pac_3	DRA004889	PacBio RSII, P6-C4 sequence chemistry	DNA	761,112	*7500	5,708
Rir_DNA_PE180	DRA004835	Hiseq PE126,TruSeq DNA prep kit	DNA	423,041,682	126	53,303
Rir_RNA_SS	DRA005204	Hiseq PE101,TruSeq RNA prep kit	RNA	16,122,964	101	1,628
Rir_RNA_rRNA	DRA006039	Miseq PE301,TruSeq RNA prep kit	RNA	20,904,594	301	6,292

*Average read length

Table S2 Blast hit data of eliminated contamination sequences

Query; Illumina-polished assembly, Database; Refseq genome, blast; default blastn search

Assembly ID	Hit ID	Assembly length	Identity %	Query cover %	Max bit score	Total score	E-value	Subject title
unitig_103	NW_016089427	4223	99	100	7548	25444	0	<i>Daucus carota</i> subsp. <i>sativus</i> cultivar DH1 unplaced genomic scaffold, ASM162521v1 DCARv2_B11, whole genome shotgun sequence
unitig_228	NC_030383	2192	93	100	3208	87603	0	<i>Daucus carota</i> subsp. <i>sativus</i> cultivar DH1 chromosome 3, ASM162521v1, whole genome shotgun sequence
unitig_85	NW_016092113	14490	99	99	10975	106600	0	<i>Daucus carota</i> subsp. <i>sativus</i> cultivar DH1 unplaced genomic scaffold, ASM162521v1 DCARv2_S316, whole genome shotgun sequence
unitig_109	NW_016089417	1602	99	100	2826	52864	0	<i>Daucus carota</i> subsp. <i>sativus</i> cultivar DH1 chromosome 2 genomic scaffold, ASM162521v1 2, whole genome shotgun sequence
unitig_88	NW_016092113	14867	98	100	11607	90386	0	<i>Daucus carota</i> subsp. <i>sativus</i> cultivar DH1 unplaced genomic scaffold, ASM162521v1 DCARv2_S316, whole genome shotgun sequence
unitig_100	NW_016089417	8259	94	100	12676	539700	0	<i>Daucus carota</i> subsp. <i>sativus</i> cultivar DH1 chromosome 2 genomic scaffold, ASM162521v1 2, whole genome shotgun sequence
unitig_83	NW_016092113	46670	98	100	11686	328700	0	<i>Daucus carota</i> subsp. <i>sativus</i> cultivar DH1 unplaced genomic scaffold, ASM162521v1 DCARv2_S316, whole genome shotgun sequence
unitig_84	NW_016092113	14454	98	100	11673	103600	0	<i>Daucus carota</i> subsp. <i>sativus</i> cultivar DH1 unplaced genomic scaffold, ASM162521v1 DCARv2_S316, whole genome shotgun sequence
unitig_101	NC_030388	3879	95	100	6058	146300	0	<i>Daucus carota</i> subsp. <i>sativus</i> cultivar DH1 chromosome 8, ASM162521v1, whole genome shotgun sequence

Table S3 Assembly statistics of *R. irregularis* genomes

Assembly ID	IDs in public database (database)	# assembles	Total length	% total length	# N's bases	N50 (bp)	Longest assembly	CEGMA completeness	
								#	%
RIR17	BDIQ01000000 (DDBJ)	210	149,750,837	97.24	0	2,308,146	5,727,599	244	98.4
Lin_14	JEMT00000000 (Genbank)	30,233	140,695,967	91.36	30,115	16,434	198,953	229	92.3
Tis_13	<i>R. irregularis</i> v1.0 (MycCosm)	28,371	91,083,792	59.15	270,422	4,186	57,883	-	98

Table S4 Predicted gene number and BUSCO completeness of *R. irregularis* DAOM-181602 genomes

	Gene model (Longest isoform)		Isoform model		Complete BUSCOs (S+D)		Complete and single-copy BUSCOs (S)		Complete and duplicated BUSCOs (D)		Fragmented BUSCOs		Missing BUSCOs	
	#	Average length (AA)	#	Average length (AA)	#	%	#	%	#	%	#	%	#	%
RIR17	41,572	377	43,675	378	273	94.1	243	83.8	30	10.3	11	3.8	6	2.1
Lin_14	27,301	347	29,911	351	256	88.3	186	64.1	70	24.1	7	2.4	27	9.3
Tis_13	30,282	270	-	-	257	88.6	251	86.6	6	2.1	29	10.0	4	1.4

Table S5 Gene model confidence based on RNA-seq expression support, homology evidence or protein motif evidence
(See "Table_S05.xlsx")

Table S6 Gene number joined for each orthologous group in orthoDB9
(See "Table_S06.xlsx")

Table S7 Blast search of genes involved in the degradation of plant cell wall

Query; Corresponding gene in *Aspergillus niger*, Database; all of constructed gene models from RIR17 (for blastp search), or RIR17 genomic sequences (for tblastn search)

Activity category	CAZy ID	Corresponding gene in <i>Aspergillus niger</i> CBS 513.88	Blastp (e-value < 1E-4)	tBlastn (e-value < 1E-4)
Polysaccharide lyases	PL4	CAK48272	No-hit	No-hit
		CAK46452	No-hit	No-hit
	PL1	CAK48551	No-hit	No-hit
		CAK48529	No-hit	No-hit
		CAK48388	No-hit	No-hit
		CAK47350	No-hit	No-hit
		CAK40523	No-hit	No-hit
Cellobiohydrolases	GH7	CAK37997	No-hit	No-hit
		AAF04492	No-hit	No-hit
	GH6	AAF04491	No-hit	No-hit
		CAK41068	No-hit	No-hit
Lytic polysaccharide mono-oxygenases	AA9	CAK39856	No-hit	No-hit
		CAK97324	No-hit	No-hit
		CAK97151	No-hit	No-hit
		CAK46515	No-hit	No-hit
		CAK45495	No-hit	No-hit
		CAK42466	No-hit	No-hit
		CAK41095	No-hit	No-hit
Proteins with cellulose-binding motif 1	CBM1	CAK38942	No-hit	No-hit
		CAK97324	No-hit	No-hit
		CAK46515	No-hit	No-hit
		CAK44069	No-hit	No-hit
		CAK43558	No-hit	No-hit
		CAK42930	No-hit	No-hit
		CAK42466	No-hit	No-hit
		CAK41068	No-hit	No-hit
		AAF04492	No-hit	No-hit

Table S8 Blast search of "the missing ascomycete core genes orthologs" by Tang et al. 2016
(See "Table_S08.xlsx")

Table S9 Blast hits of the index genes used for previous heterokaryosis studies.
(See "Table_S09.xlsx")

Table S10 Masked repetitive region of RIR17 and selected organisms
(See "Table_S10.xlsx")

Table S11 Number of detected isoforms having a transposase related domain

Pfam ID	Description	# isoform
PF10683	Hermes transposase DNA-binding domain	5
PF03108	MuDR family transposase	1
PF10551	MULE transposase domain	208
PF12784	PD-(D/E)XK nuclease family transposase	1
PF03221	Tc5 transposase DNA-binding domain	160
PF01498	Transposase	132
PF02992	Transposase family tnp2	294
PF13843	Transposase IS4	6
PF12017	Transposase protein	1
PF00872	Transposase, Mutator family	2
	Total	810

Table S12 Position and size of ribosomal DNAs in RIR17 assemblies

(See "Table_S12.xlsx")

Figure S13 Mapped read depth of coverage to RIR17 genes.

*Calculated by mapback of DNA short reads (Rir_DNA_PE180) to selected ten BUSCO genes and consensus rDNA sequences from ten rDNA paralogs.

	isoform ID	Mean read depth of coverage*	rDNAs/single -copy BUSCOs
rDNAs	rDNA_Consensus_IUPAC	1733.3	
	g15690.t1	159.3	
	g16450.t1	160.3	
	g24696.t1	157.4	
	g13527.t1	149.0	
	g1455.t1	162.1	
BUSCOs	g11464.t1	153.4	
	g13372.t1	152.5	
	g31539.t1	155.3	
	g6363.t1	146.1	
	g11459.t1	144.2	
	Average of ten BUSCOs	154.0	11.3

Table S14 Different positions among rDNA paralogs in RIR17
(See "Table_S14.xlsx")

Table S15 Mean Intragenomic differences among ten rDNA paralogs

SD = standard deviation

Region	Aligned positions	# difference		# indel		Percent identity (%)	
		Average	SD	Average	SD	Average	SD
18SrDNA	1811	1.58	1.09	0.47	0.50	99.91	0.06
ITS1	110	7.13	7.85	2.76	2.54	93.45	7.20
5.8SrDNA	158	5.29	5.14	3.00	3.14	96.65	3.25
ITS2	220	20.51	14.76	8.80	7.72	90.28	6.84
28SrDNA	3548	73.47	48.49	12.29	12.68	97.93	1.37
Whole 48SrDNA cluster	5847	106.13	69.60	27.31	22.84	98.18	1.19

Table S16 Detected number of tRNA by tRNAscan-SE

	tRNAs decoding Standard 20 AA	Selenocysteine tRNAs (TCA)	Possible suppressor tRNAs (CTA,TTA)	tRNAs with undetermined/unknown isotypes	Predicted pseudogenes	tRNAs with introns
RIR17	110	0	1	17	12	37
<i>Rhizopus oryzae</i>	237	0	0	0	2	12
<i>Phycomyces blakesleeanus</i>	308	0	1	0	0	12
<i>Mortierella elongata</i>	253	0	0	2	3	12
<i>Abisporus varbisporus</i>	157	0	0	0	3	91
<i>Saccharomyces cerevisiae</i>	286	1	0	2	6	59

Table 17 Difference among the ten paralogs from RIR17 and cloning-based partial rDNAs from DAOM-181602

Cloning-based sequence ID	Spore ID	# different sites from each rDNA paralogs in RIR17									
		c39_2	c62_2	c39_1	c312_1	c35_1	c311_1	c52_1	c356_1	c62_1	c4_1
HF968915	1	0	21	0	26	25	39	39	24	21	26
HF968920	1	1	22	1	25	24	38	38	25	22	25
HF968987	1	19	2	19	7	6	22	22	36	2	42
HF968919	1	20	1	20	6	5	23	23	37	1	41
HF968922	1	21	0	21	5	4	24	24	36	0	42
HF968986	1	21	2	21	7	6	26	26	36	2	42
HF968917	1	24	36	24	41	40	53	53	0	36	6
HF968921	1	24	36	24	41	40	53	53	0	36	6
HF968984	1	27	43	27	46	45	56	56	7	43	1
HF968916	1	40	21	40	22	21	3	3	52	21	54
HF968985	1	31	18	31	21	20	8	8	47	18	47
HF968918	1	38	19	38	18	17	7	7	52	19	54
HF968989	2	24	36	24	41	40	53	53	0	36	6
HF968992	2	24	36	24	41	40	53	53	0	36	6
HF968927	2	24	36	24	41	40	53	53	0	36	6
HF968923	2	25	4	25	1	0	24	24	40	4	44
HF968990	2	25	4	25	1	0	24	24	40	4	44
HF968991	2	25	41	25	44	43	54	54	5	41	1
HF968988	2	29	8	29	3	4	28	28	44	8	48
HF968924	2	32	46	32	51	50	45	45	10	46	12
HF968925	2	16	9	16	10	9	29	29	36	9	36
HF968926	2	10	21	10	26	25	43	43	15	21	21
HF968914	3	2	19	2	24	23	39	39	22	19	24
HF968910	3	21	0	21	5	4	24	24	36	0	42
HF968912	3	21	0	21	5	4	24	24	36	0	42
HF968913	3	21	0	21	5	4	24	24	36	0	42
HF968982	3	26	42	26	45	44	55	55	6	42	0
HF968911	3	39	24	39	25	24	0	0	53	24	55
HF968983	3	39	24	39	25	24	0	0	53	24	55
HF968981	3	40	25	40	26	25	1	1	54	25	56

Table 18 Heterozygous SNPs detected by the mapping back of "Rir_DNA_PE180" to repeat-masked RIR17
(See "Table_S18.xlsx")

Table 19 SNPs detected by mapping of the previously assigned data from single-spore DAOM-181602
(See "Table_S19.xlsx")

Table S20 Commonly observed SNPs between our "Rir_DNA_PE180" and the single-spore sequencing data
(See "Table_S20.xlsx")