bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 2 # Parallel Concerted Evolution of Ribosomal Protein Genes in Fungi and Its Adaptive Significance

3

4 Alison Mullis[1#], Zhaolian Lu[1#], Yu Zhan[1#], Tzi-Yuan Wang[2], Judith Rodriguez[3], Ahmad
5 Rajeh[1,3], Ajay Chatrath[1], Zhenguo Lin[1*]

6

7 [1] Department of Biology, Saint Louis University, St. Louis, MO, USA 63103

8 [2] Biodiversity Research Center, Academia Sinica, Nankang, Taipei, Taiwan

9 [3] Program of Bioinformatics and Computational Biology, Saint Louis University, St.
10 Louis, MO USA 63103

11

12

13 [#] These authors contributed equally to this work

14 * To whom correspondence should be addressed

15 Zhenguo Lin, Email: zhenguo.lin@slu.edu

16 Phone: 314-977-9816

17

18 Keywords: ribosomal proteins, gene duplication, gene conversion, concerted evolution, fermentation

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

19 **ABSTRACT**

20 Ribosomal proteins (RPs) genes encode structure components of ribosomes, the cellular
21 machinery for protein synthesis. A single functional copy has been maintained in most of 78-80
22 RP families in animals due to evolutionary constraints imposed by gene dosage balance. Some
23 fungal species have maintained duplicate copies in most RP families. How the RP genes were
24 duplicated and maintained in these fungal species, and their functional significance remains
25 unresolved. To address these questions, we identified all RP genes from 295 fungi and inferred
26 the timing and nature of gene duplication for all RP families. We found that massive duplications
27 of RP genes have independently occurred by different mechanisms in three distantly related
28 lineages. The RP duplicates in two of them, budding yeast and Mucoromycota, were mainly
29 created by whole genome duplication (WGD) events. However, in fission yeasts, duplicate RP
30 genes were likely generated by retroposition, which is unexpected considering their dosage
31 sensitivity. The sequences of most RP paralogs in each species have been homogenized by
32 repeated gene conversion, demonstrating parallel concerted evolution, which might have
33 facilitated the retention of their duplicates. Transcriptomic data suggest that the duplication and
34 retention of RP genes increased RP transcription abundance. Physiological data indicate that
35 increased ribosome biogenesis allowed these organisms to rapidly consuming sugars through
36 fermentation while maintaining high growth rates, providing selective advantages to these
37 species in sugar-rich environments.

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

38 **INTRODUCTION**

39 Gene duplication has served as a driving force for the evolution of new phenotypic traits and
40 contributed to adaptation of organisms to their specific niches (Ohno 1970; Sidow 1996).
41 Duplicate genes are mainly generated by chromosome or whole genome duplication (WGD),
42 unequal crossing-over, and retroposition (Zhang 2013). Similar to other types of mutations, only
43 a small portion of duplicate genes can be eventually fixed in a population, and the survivors are
44 usually advantageous to the organisms (Zhang 2003; Kondrashov and Kondrashov 2006).
45 Highly diverse retention patterns of duplicate genes have been observed among gene families
46 (Hahn, et al. 2005). For instance, tens to hundreds of odorant receptors (ORs) genes can be
47 found in metazoan genomes (Sanchez-Gracia, et al. 2009). In contrast, many genes have been
48 maintained as a single copy since the divergence of eukaryotes, such as the DNA repair genes
49 RAD51, MSH2, and MLH1 (Lin, et al. 2006; Lin, et al. 2007; Zeng, et al. 2014).

50 Another notable example is the gene families encoding for cytosolic ribosomal proteins
51 (RPs), which are the structural components of ribosomes. Ribosomes carry out one of the most
52 fundamental processes of living systems by translating genetic information from mRNA into
53 proteins. In eukaryotes, each ribosome consists of two subunits, the small and large subunit,
54 which consist of 78-80 different RPs and four types of ribosomal RNAs (rRNA) (Wool 1979;
55 Wimberly, et al. 2000). RP genes are highly conserved in all domains of life (Korobeinikova, et
56 al. 2012). Each RP was found to have unique amino acid sequences with very limited to none
57 similarities between each other. For most animals studied, only a single functional copy of RP
58 gene is maintained in each family, although many processed pseudogenes may be found
59 (Dudov and Perry 1984; Kuzumaki, et al. 1987; Kenmochi, et al. 1998). As structural
60 components of the highly expressed macromolecular complex, the evolutionary constraints on
61 duplicate RP genes was believed to be imposed by gene dosage balance (Birchler and Veitia
62 2012). In plants, attributing to polyploidization or WGD events, multiple gene copies are usually
63 present in each RP families in polyploid plants (Vision, et al. 2000; Barakat, et al. 2001). This is
64 probably because all RP genes were duplicated simultaneously by WGD, allowing maintenance
65 of balanced dosage among RPs (Birchler and Veitia 2012).

66 Similar to polyploid plants, most of RP families in the budding yeast *Saccharomyces*
67 *cerevisiae* have duplicate copies due to the occurrence of a WGD (Wolfe and Shields 1997;
68 Kellis, et al. 2004). Many RP paralogous genes in *S. cerevisiae* generated by WGD, or RP
69 ohnologs, are more similar to each other than to their orthologous genes due to interlocus gene

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

70 conversion (Evangelisti and Conant 2010; Casola, et al. 2012). During a interlocus gene
71 conversion, one gene serves as a DNA donor that replaces the sequences of its paralogous
72 gene (Chen, et al. 2007). As a result, the sequences of paralogous genes have been
73 homogenized, resulted in the ancient duplicate events to appear much more recent, which was
74 called "concerted evolution" (Brown, et al. 1972). One of the best-known examples of concert
75 evolution is the genes encoding RNA component of ribosomes, the rRNA genes, in both
76 prokaryotes and eukaryotes (Arnheim, et al. 1980; Schlotterer and Tautz 1994; Blattner, et al.
77 1997).

78 According to the Ribosomal Protein Gene Database (RPG) (Nakao, et al. 2004), three of ten
79 fungal species listed have multiple gene copies in most RP families, including *S. cerevisiae*, a
80 fission yeast *Schizosaccharomyces pombe*, and a pin mold *Rhizopus oryza*e. The duplicate RP
81 genes in *R. oryzae* could be generated by a WGD event it is ancestor (Ma, et al. 2009),
82 although it has not been systematically examined. In addition, most RP families have more than
83 four gene copies, which cannot be explained by a single WGD. Unlike *S. cerevisiae* and *R.*
84 *oryzae*, no WGD has been detected during the evolution of *Sch. pombe* (Rhind, et al. 2011),
85 suggesting that each RP family might be duplicated independently by small scale duplication
86 (SSDs) events. This observation is unexpected because the duplicate genes encoding
87 macromolecules generated by SSDs are much less likely to survive because they are sensitive
88 to gene dosage balance (Li, et al. 1996; Conant and Wolfe 2008). It remains an unexplored
89 dimension about how RP genes have been duplicate and maintained in fungi, particularly in the
90 fission yeast *Sch. pombe*.

91 The expression of RP genes has been thoroughly linked to growth and proliferation,
92 reflecting their central role in the regulation of growth in yeast (Montagne, et al. 1999;
93 Jorgensen, et al. 2002; Brauer, et al. 2008). In rapid growth yeast cells, ~50% of RNA
94 polymerase II (Pol II) transcription initiation events are devoted to RP expression (Warner
95 1999). Therefore, the duplication and retention of RP genes might have more functional impacts
96 on these microorganisms than animals or plants. Like other types of mutations, the occurrence
97 of gene duplication is largely due to scholastic events, but retention of duplicate genes have
98 been mainly driven by natural selection (Panchy, et al. 2016). A better understanding of
99 evolutionary fates of RP duplicate genes could offer new insights into how gene duplication
100 produced adaptive solutions to microorganisms.

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

101 To better understand the evolutionary patterns of RP genes and their adaptive significance,
102 we conducted systematic identification and evolutionary analyses of all RP families in all fungal
103 species with well-annotated genomes. We searched for RP genes from 295 fungal species and
104 identified independent duplications of most RP families in three distantly related fungal lineages.
105 We inferred the timing and nature of gene duplication for each RP family in each fungal lineage.
106 We found that a vast majority of RP paralogous genes have experienced repeated gene
107 conversion events that have homogenized their sequences in each species. In aligning with
108 integrative analyses of genomic, transcriptomic data and physiological data, we proposed that
109 the massive duplication, retention and concerted evolution of RP genes have contributed to the
110 evolution of fermentative lifestyle in fungal species. This study offers a classic example
111 illustrating the mechanisms and adaptive significance of maintaining duplicate genes encoding
112 macromolecules.

113 ## RESULTS

114 ### Massive duplications of RP genes found in three distantly related fungal lineages

115 To determine the prevalence of gene duplications in RP families in fungi, we first searched
116 for RP homologous genes for all fungal species with NCBI Reference Sequence (RefSeq)
117 protein data (Supplementary table 1). As of March 2019, 285 fungal species were annotated
118 with RefSeq protein data, covering five of the seven fungal phyla. We conducted BLASTP
119 searches against the 285 RefSeq protein datasets using amino acid sequences of RP genes
120 from both *S. cerevisiae* and *Sch. pombe* as queries (see Methods and Materials). Based on
121 BLASTP search results, we calculated the gene copy numbers in each RP family for every
122 examined species, and the total number of RP families with duplicate copies (Supplementary
123 table 1).

124 We considered a species with massive RP duplications if more than 50% (≥ 40) of RP
125 families have duplicate copies. Among the 285 fungi examined, only ten species meet the
126 criterion of massive RP duplication. The ten species distributes in three distantly related fungal
127 lineages: three in the class of Saccharomycetes (budding yeast), four in the class of
128 Schizosaccharomycetes (fission yeast), and three in the phylum of Mucoromycota
129 (Supplementary table 1). Although multiple hits of most CRP families were found in a budding
130 yeast *Candida viswanathii*, because the assembly type of its genome is diploid, the two hits

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

131 found in most RP families are alleles instead of paralogous gene, it was not considered as
132 massive RP duplication.

133 Because protein annotations of a genome could be incomplete or inaccurate, manual
134 curation is required for a more accurate survey of RP repertoire. It is necessary to carry out a
135 second-round identification of RP genes with manual curation, focusing on the three fungal
136 lineages. We selected 24 species from the three fungal lineages, including the ten species with
137 massive RP duplication. To provide a more even distribution of taxonomic groups in each
138 lineage, we included ten other species whose genomic data are available in NCBI Whole
139 Genome Shotgun (WGS), Yeast Gene Order Browser (YGOB) and JGI (Byrne and Wolfe 2005;
140 Maguire, et al. 2013). In total, our second-round search examined 34 fungal species, which
141 includes 23 Saccharomycetes species, five Taphrinomycotina species (including the four fission
142 yeasts), and six Mucoromycota species (fig. 1, Supplementary table 2). The phylogenetic
143 relationships of the 34 species were inferred using the amino acid sequences of the largest
144 subunit of RNA Pol II proteins (Supplementary fig. 1). Including the 285 species analyzed in our
145 first round analysis, we have examined a total number of 295 fungal genomes in the two rounds
146 of RP gene searches, representing the largest scale of RP gene survey in fungi to our
147 knowledge.

148 To manually curate RP repertoire in a genome, we performed both BLASTP and TBLASTN
149 searches for each of the 34 fungal species. By comparing BLASTP and TBLASTN search
150 results, we identified discrepancies in the number of RP genes and aligned regions. We found
151 that many TBLASTN hits were not present in BLASTP searches, indicating the presence of
152 unannotated RP genes. Thus, we have manually predicted 259 novel RP genes from 32 of the
153 34 species. We also revised the annotations of open reading frame (ORF) for 95 RP genes. In
154 total, we identified a total number of 3950 RP genes from the 34 fungal species (Supplementary
155 tables 2 and 3).

156 We constructed maximum likelihood (ML) phylogenetic trees for each RP family (see
157 Materials and Methods). Similar tree topologies were observed among RP families with
158 duplicate copies (Supplementary File 1). For instances, two copies of RPL6 genes are present
159 in all ten post-WGD budding yeasts and all four fission yeast species (fig. 2A). More copies of
160 RPL6 genes are present in Mucoromycota species *Lobosporangium transversale*, *Phycomyces*
161 *blakesleeanus*, *Rhizopus microspores*, and *Rhizopus delemar* (former name *R. oryzae*), which
162 have 2, 4, 3, and 5 copies of RPL6 genes, respectively. According to the ML tree (fig. 2A), the

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

163 RPL6 genes are more closely related to their paralogous genes in each species, instead of their
164 orthologous genes. Similar patterns are present in the RPS19 genes, which encode a ribosomal
165 small subunit protein (fig. 2B). These tree topologies suggest that the RP genes have been
166 independently duplicated in each species after their divergence from each other. However, at
167 least in the post-WGD budding yeasts, it has been documented that RPL6 and RPS19 were
168 generated by the WGD occurred prior to the divergence of *S. cerevisiae* and *Vanderwaltozyma*
169 *polyspora* (Conant and Wolfe 2006). Therefore, the phylogenetic trees do not accurately depict
170 the evolutionary history of RPL6 and RPS19 families in budding yeasts. It has been shown that
171 RPL6 and RPS19 genes have experienced gene conversion during the evolution of *S.*
172 *cerevisiae*, which explains the discrepancy between the tree topology and their duplication
173 history in the budding yeast (Evangelisti and Conant 2010; Casola, et al. 2012). However, it is
174 not known whether it is the same case in the fission yeast and Mucoromycota species, which
175 requires accurate timing of duplication events in the two lineages.

176 Based on these cases, we cannot infer the evolutionary history of RP genes in fungi solely
177 based on the topology of phylogenetic trees due to the possibility of gene conversion. It is
178 necessary to carry out additional analyses to determine when gene duplication events have
179 occurred. Because only a small number of species have experienced massive RP duplication in
180 each of the three fungal lineages (fig. 1 and Supplementary table 2), the most parsimonious
181 scenario is that the expansion of RP genes in each fungal lineage occurred independently. In
182 our subsequent analyses, we separately inferred the timing and nature of gene duplications for
183 each RP family and determined whether they have experienced gene conversion after gene
184 duplication in each lineage.

185 **Duplication and concerted evolution of RP genes in the budding yeasts**

186 We manually identified all RP genes for the 23 Saccharomycetes species (budding yeasts).
187 A total number of 59 RP families have duplicate copies in most WGD species (fig. 1 and
188 Supplementary table 2). 55 of them are ohnologs generated by an ancestral WGD (Conant and
189 Wolfe 2006). The other four RP families, including RPP1, RPP2, RPL9, and RPS22, have
190 duplicates in most budding yeasts, including these non-WGD species, suggesting that they
191 have been duplicated before the divergence of all budding yeasts. Therefore, most post-WGD
192 budding yeasts have a significant increase in RP gene number. 135-137 RP genes are present
193 in the four species in the *Saccharomyces sensu stricto* group, *S. cerevisiae*, S. *mikatae*, *S.*
194 *uvarum,* and *S. eubayanus*. The two *Naumovozyma* species have 134 and 137 RP genes,

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

195 respectively. The numbers are relatively smaller (106 and 104) in the other two early-diverging
196 post-WGD species, *Tetrapisispora phaffii* and *Vanderwaltozyma polyspora*. In contrast, the
197 human opportunistic pathogen *Candida glabrata* (*Nakaseomyces glabrata*) and its closely
198 related species, *Nakaseomyces bacillisporus,* have only 85 and 97 RP genes respectively,
199 suggesting most RP duplicate genes have been lost in these species.

200 Because the timings of these RP duplications in budding yeast have already been
201 determined, it is possible to infer which RP paralogs have experienced gene conversion by
202 comparing their gene tree with their true duplication history. We can also use the tree topology
203 to infer when concerted evolution had terminated, which is the time when the paralogous genes
204 started to accumulate mutations independently in each paralogous gene. To simplify this
205 process, we constructed phylogenetic trees for each duplicate RP family using five
206 representative WGD species with different divergence times, including *S. cerevisiae*, *S.*
207 *mikatae*, *S. eubayanus*, *N. castellii* and *T. phaffii* (fig. 3 and Supplementary File 2). We found
208 that at least 52 RP duplicate pairs in *S. cerevisiae* have experienced gene conversion, including
209 50 pairs generated by WGD and two pairs (RPL9 and RPS22) generated by the ancient
210 duplication events (fig. 3). Therefore, 88% (52 out of 59) of RP paralogous genes in *S.*
211 *cerevisiae* have experienced gene conversion, which is more than that of previously identified
212 (16 and 29) (Evangelisti and Conant 2010; Casola, et al. 2012), suggesting that concerted
213 evolution of RP genes in the budding yeasts is more prevalent than previously recognized.

214 Based on gene tree topologies, we inferred when converted evolution of RP genes had
215 terminated during the evolution of budding yeasts. In 21 RP families, two copies of each RP
216 family gene in *S. cerevisiae* form a species-specific clade in gene tree (fig. 3A and
217 Supplementary File 2), suggesting that the concerted evolution is still ongoing in *S. cerevisiae* or
218 has recently terminated after its divergence from *S. mikatae*. In seven RP families, termination
219 of concerted evolution occurred before the split between *S. cerevisiae* and *S. mikatae* (fig. 3B).
220 24 RP families have ended their concerted evolution before the divergence of the
221 *Saccharomyces sensu stricto* group, which including *S. cerevisiae*, *S. mikatae*, *S. eubayanus*
222 (fig. 3C). Only seven RP gene pairs do not show strong evidence of gene conversion (fig. 3 D).
223 A summary of the termination time of concerted evolution in 59 RP pairs in *S. cerevisiae* was
224 provided in fig. 3E.

225 **Duplication and concerted evolution of RP genes in the fission yeasts**

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

226 We identified all RP genes for five species in the subphylum of Taphrinomycotina, including
227 the four fission yeasts and *Pneumocystis murina*. *P. murina* belongs to the class of
228 Pneumocystidomycetes, which is probably the most closely related lineage to the fission yeasts,
229 and it was used as an outgroup to infer the evolutionary history of RP genes. 142 to 145 RPs
230 are present in the four fission yeast species. The number of RP families with duplicate copies
231 range from 58 to 59 in the four species (fig. 1 and Supplementary table 2). Most of them have
232 two gene copies, but three copies of RP genes are present in 6 RP families (fig. 1). Only 78 RP
233 genes were identified in *P. murina*. Thus, it is reasonable to assume that the massive expansion
234 of RPs genes in the fission yeasts occurred after their divergence with *P. murina*.

235 Similar to the budding yeasts, most paralogous RP genes in each fission yeast are more
236 similar to each other than to their orthologous genes (fig. 2 and Supplementary File 3). The tree
237 topologies indicate that these RP genes were duplicated independently in each fission yeast
238 after their divergence. However, we should consider the possibility of gene duplication which
239 resulted in underestimation the ages of duplicate genes. To infer when gene duplication
240 occurred, we conducted gene collinearity (microsynteny) analysis for all duplicate RP genes in
241 the four fission yeasts (see Methods and Materials). If an RP gene was duplicated
242 independently in each species after their divergence, the daughter genes are expected to be
243 found in different genomic regions among these species. Under this scenario, only the parental
244 copy of RP genes share microsynteny by them, and it is extremely unlikely that they also share
245 conserved regions of microsynteny around the daughter genes. On the other hand, if these
246 fission yeasts share microsynteny in both copies of RP genes, the two copies should be created
247 by a single gene duplication event in their common ancestor.

248 We first identified all orthologous groups in the four fission yeasts (Supplementary table 4).
249 Based on gene order and ortholog group information, we analyzed microsynteny for each pair of
250 RP genes. Herein, we defined a conserved region of microsynteny as a block containing three
251 or more conserved homologs within five genes downstream and upstream of an RP gene (fig.
252 4A). In *Sch. pombe,* 58 RP families have at least two gene copies. In each of the RP families,
253 the duplicate genes share microsynteny in *Sch. pombe*, *Sch. cryophilus*, and *Sch. octosporus*
254 (Supplementary table 5). This result suggests that duplicate pairs in the 58 RP families were
255 generated at least before the divergence of the three fission yeasts, which have occurred ~119
256 million years ago (mya) (Rhind, et al. 2011). We then inferred how many of gene duplication
257 events occurred even before the split of *Sch. japonicas*, approximately 220 mya (Rhind, et al.
258 2011). Thanks to the highly conserved gene order in fission yeasts (Rajeh, et al. 2018), we were

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

259 able to detect the presence of microsynteny in both RP duplicates in 49 families in *Sch.*
260 *japonicus*, suggesting that gene duplication of these RP families have occurred before the
261 divergence of the four fission yeasts (Supplementary table 5). For example, two copies of
262 RPL11 genes are found in each *Schizosaccharomyces* species. Highly conserved regions of
263 microsynteny surrounding RPL11A genes were found in all four fission yeasts, and so were the
264 RPL11B genes (fig. 4A), supporting that RPL11 was duplicated in their common ancestor and
265 both copies have been maintained in each fission yeast after their divergence.

266 For the rest 9 RP families, we did not obtain conclusive evidence to determine whether they
267 were duplicated before the split of *Sch. japonicus*. Four of them (RPL10, RPL30, RPS12, and
268 RPS25) has only a single RP copy present in *Sch. japonicus*. These genes could be duplicated
269 in the common ancestor of fission yeasts, and a duplicate copy has subsequently lost in *Sch.*
270 *japonicus*. Alternatively, the duplication events have occurred after the split of *Sch. japonicus*
271 from the other species. In the other five RP families (RPL3, RPL17, RPL18, RPL21, and
272 RPS19), only one RP copy in *Sch. japonicus* share microsynteny with the other three species.
273 Similarly, the duplication events of these RP families could be predated to the divergence of
274 fission yeasts, following by genome rearrangements in *Sch. japonicus* that resulted in the loss of
275 its gene collinearity. However, we cannot exclude the possibility that they were generated by
276 independent duplication events in *Sch. japonicus*.

277 To determine the number of RP families has an incompatibility between gene phylogenetic
278 tree and duplication history, we constructed a phylogenetic tree for each RP family with
279 duplicates in the fission yeasts. In the case of RPL11, contradict to the gene true duplication
280 history as inferred by microsynteny analysis (fig. 4A), the phylogenetic tree shows that RPL11
281 paralogs from species-specific clades in each fission yeast (Fig. 4B). Such incompatibility
282 suggests that gene conversion has occurred between RPL11 paralogous genes in each fission
283 yeast after their divergence. A total number of 45 RP families (77.6%) in fission yeasts have a
284 similar tree topology to RPL11 (Supplementary File 3). In other families, such as RPS17, the
285 two copies of RP genes from *Sch. pombe*, *Sch. octosporus* and *Sch. cryophilus* form two clades
286 and each clade have one gene copy from the three species. We observed nine RP families
287 similar to RPS17, suggesting that concerted evolution of these RP genes might have been
288 terminated before their divergence (fig. 4C). However, we did not find evidence of gene
289 conversion in four RP families, including RPL30, RPS5, RPS12 and RPS28 (fig. 4D and E).

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

290 **Retroposition as a major mechanism for massive duplication of RP genes in the**
291 **ancestral fission yeast**

292 Because no WGD was detected during the evolution of *Sch. pombe* (Rhind, et al. 2011), we
293 then inferred other mechanisms that resulted in massive duplication of RP genes in the fission
294 yeasts, such as unequal crossing-over and retroposition. Unequal crossing-over typically
295 generates segmental or tandem gene duplicates. If a pair of genes was generated by segmental
296 duplication, we expect to observe microsynteny between regions of paralogous RP genes within
297 a species. However, we did not find any case in these RP families (Supplementary table 5).
298 Furthermore, we did not detect tandemly arranged RP paralogous genes, suggesting that
299 unequal crossing-over is not a main contributor for RP duplications in the fission yeasts either.

300 Retroposition generates retroduplicates through random insertion of a retrotranscribed
301 cDNA from parental source genes, resulting in intron-less retroduplicate genes (Kaessmann, et
302 al. 2009). We examined the exon-intron structure for all RP paralogous genes in *Sch. pombe*.
303 Among the 21 singleton RP families in *Sch. pombe*, only 7 of them (33.3%) are intron-less
304 (Supplementary table 6). In contrast, 33 of 58 duplicate RP families (56.9%) have at least one
305 copy of intron-less gene, which is significantly higher than the group of singleton RPs (p =
306 0.006, Fisher exact test). This ratio is also significantly higher than 27.3% of RP duplicates
307 generated by WGD in *S. cerevisiae*. Thus, the enrichment of intron-less RP genes in the
308 duplicate RP families in fission yeast suggests that they were likely generated by retroposition.
309 For those RP duplicates with intron in both copies, the possibility that they may be created by
310 retroposition following by insertion of intron cannot be excluded, because the locations and
311 phases of introns between these paralogous RP genes in *Sch. pombe* are usually different.

312 **Duplication and concerted evolution of RP genes in the Mucoromycota species**

313 Four Mucoromycota species examined demonstrate massive duplication of RP genes.
314 Three of them belong to the order of Mucorales (pin molds) in subphylum of Mucoromycotina
315 (311 RP genes in *R. delemar*, 182 in *R. microspores*, and 217 in *P. blakesleeanus*). In their
316 distantly related species in the same subphylum, *Bifiguratus adelaidae*, only 89 RP genes were
317 found. Massive duplication of RP genes (137 RP genes) was also observed in *Lobosporangium*
318 *transversale*, which is a distantly related species belonging to another subphylum
319 Mortierellomycotina. The earliest diverging species among all Mucoromycota species examined
320 is *Rhizophagus irregularis*, which has only 78 RPs genes (fig. 1).

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

321 Based on RP gene copy numbers and the evolutionary relationships of these Mucoromycota
322 species, it is most parsimonious to conclude that massive expansion of RP genes in the three
323 pin mold species and *L. transversale* occurred independently*. L. transversal* is a rare species
324 that having only been reported by a few isolations in North American (Benny and Blackwell
325 2004). The genomic studies and physiological characterizations *L. transversale* are scarce. Due
326 to lack of genomic data from closely related species, we cannot provide a systematic inference
327 of the timing and nature of massive RP duplication in *L. transversale*. Thus, our subsequent
328 analysis only focused on the origin and evolution of RP duplicate genes in pin molds.

329 A WGD event has been proposed in ancestral *R. delemar* (Ma, et al. 2009). Another WGD
330 was speculated to have occurred in *P. blakesleeanus* prior to its divergence from *R.*
331 *microspores* and *R. delemar* (Corrochano, et al. 2016). Therefore, *R. delemar* might have
332 experienced two rounds of WGDs, which correlates with the largest RP repertoire (311)
333 identified in *R. delemar*, while *P. blakesleeanus* and *R. microspores* have 217 and 182 RP
334 genes respectively. Based on RP gene numbers, it is reasonable to conclude that the second
335 WGD occurred after the divergence of *R. delemar* from *R. microspores*.

336 We conducted microsynteny analysis to infer which RP gene pairs were generated by the
337 WGDs in the pin molds. The estimated divergence time between *Phycomyces* and *Rhizopus* is
338 over 750 mya (Mendoza, et al. 2014). Most, if not all, microsynteny blocks generated by the first
339 WGD might have lost during the evolution of these pin molds. Even though we have used a less
340 strict definition of microsynteny (a minimum of 3 shared homologs in a block of ± 10 neighboring
341 genes surrounding RP), we only identified 3 and 10 pairs of microsynteny blocks between
342 paralogous RP genes in *R. microspores* and *P. blakesleeanus* respectively*.* In contrast, in *R.*
343 *delemar*, which has experienced a second round of WGD after its divergence from *R.*
344 *microspores*, we detected microsynteny for 63 pairs of RP paralogous genes (Supplementary
345 table 7), supporting the recent WGD as a major contributor to the expansion of RP genes in *R.*
346 *delemar*.

347 We attempted to identify microsynteny for orthologous RP genes to infer the evolutionary
348 history of each RP family in pin molds (Supplementary table 8). Due to the large divergence
349 times between these species, most RP orthologous genes lack well-supported microsynteny
350 (Supplementary table 7). The most well-supported example is probably the RPL3 family (Fig.
351 5A). Based on the shared gene orders between paralogous and orthologous RPL3 genes, it is
352 reasonable to infer that RPL3 was duplicated before the divergence of the three pin mold

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

353 species, probably due to the first WGD. In *R. delemar*, the two RPL3 genes have been further
354 duplicated by the recent WGD, generating four copies. However, their gene tree (fig. 5B)
355 demonstrates that the RPL3 paralogous genes in each species form a species-specific clade,
356 suggesting the occurrence of gene conversion between paralogous RPL3 genes in each
357 species. A total number of 57 RP families have a similar tree topology (Supplementary File 4).
358 Although there is no conclusive microsynteny evidence to support that these RP families have
359 the same evolutionary history as RPL3, we believed that it would be the most likely scenario. In
360 some RP families, such as RPL38 (fig. 5C), the genes from *R. delemar* and *R. microspores*
361 form two clades, and each clade has members from both species. There are 17 RP families
362 have a similar tree topology to RPL38. If these RP genes were the product of the ancient WGD,
363 their concerted evolution had terminated prior to the divergence of the two *Rhizopus* species.
364 The last type of tree topology, such as RPS20 (fig. 5D), whose members form two clades, and
365 each clade include genes from all the three pin mold species. Such tree topology does not
366 support the occurrence of gene conversion. We observed two RP families belonging to this
367 type. In summary, our results implied that most RP paralogous genes in pin molds might have
368 also experienced gene conversion, similar to what happened in budding yeasts and fission
369 yeasts.

370 **cDNA as the probable donor for gene conversion between RP paralogous genes**

371 During gene conversion, the genomic sequence of the ‘acceptor’ locus is replaced by a
372 'donor' sequence through recombination (Chen, et al. 2007). The donor can be genomic DNA or
373 cDNA derived from an mRNA intermediate (Derr and Strathern 1993; Storici, et al. 2007). If
374 genomic DNA is the donor, the sequences of both intron and exon can be homogenized. In
375 contrast, if cDNA is the donor, only the exon sequences of the acceptor are replaced.
376 Considering that synonymous mutations are largely free from natural selection, it is possible to
377 determine the donor of gene conversion by comparing the substitution rates between intron and
378 synonymous sites of exons. If the synonymous substitution rates (d_S) are significantly lower than
379 intron mutation rates (μ_{intron}), supporting cDNA as a donor. We calculated d_S and μ_{intron} for all RP
380 duplicate genes for presentative species from each fungal lineage: *S. cerevisiae*, *Sch. pombe*
381 and *R. microspores* (Supplemental table 9). Overall, the d_S values of all paralogous RP genes
382 are significantly lower than μ_{intron} in each species examined (fig. 6A-C, Student’s t-test, $p < 0.01$).
383 Considering that different genomic regions might have different mutation rates, we then
384 compared the d_S and μ_{intron} between each pair of RP duplicate genes (fig. 6D-E). Consistently,
385 most of RP duplicate gene pairs have lower d_S values than μ_{intron}. In a small number of cases,

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

386 high d_S are observed, probably because the concerted evolution between a pair of orthologous
387 genes have terminated long time ago, resulting accumulation of many synonymous mutations.
388 These results suggest that, in most cases, only the coding sequences have been homogenized
389 by gene conversion, supporting cDNA as the probable gene conversion donor.

390 **The retention of RP gene duplicates was associated with the evolution to fermentative**
391 **ability in fungi**

392 Most eukaryotic species fully oxidize glucose, their primary carbon and energy source,
393 through mitochondrial oxidative phosphorylation in the presence of oxygen for maximum energy
394 production. In contrast, post-WGD budding yeasts and fission yeasts predominantly ferment
395 sugar to ethanol in the presence of excess sugars, even under aerobic conditions, which was
396 called aerobic fermentation (Alexander and Jeffries 1990; Lin and Li 2011a). Aerobic
397 fermentation has independently evolved in the budding yeasts and fission yeasts (de Jong-
398 Gubbels, et al. 1996). In addition, the domesticated form of *R. microspores* has been a widely
399 used starter culture for the production of tempeh from fermented soybean (Hachmeister and
400 Fung 1993). Its close relative, *R. delemar*, was also well known as efficient ethanol and fumaric
401 acid producer by fermentation (Kito, et al. 2009; Straathof and van Gulik 2012). *P.*
402 *blakesleeanus*h was known for capable of fermenting sugar into β-carotene at an industrial
403 scale, which is derived from the end product of glycolysis (Kaessmann, et al. 2009).

404 We speculated that the massive duplication and retention of RP genes have contributed to
405 the evolution of fermentative ability in these species. Increased gene dosage could lead to a
406 quantitative increase in gene expression and production of protein. To determine the impact of
407 gene duplication on the production of RP transcripts, we calculated the total transcription
408 abundance of all RP genes using our transcriptomic data generated by Cap Analysis of Gene
409 Expression (CAGE) (McMillan, et al. 2019). The CAGE technique captures and sequences the
410 first 75 bp of transcripts, which quantifies the transcription abundance based on numbers of
411 mapped reads (Murata, et al. 2014). Among the 34 species, 11 of them have available CAGE
412 data, including nine budding yeasts and two fission yeasts (Supplementary table 10). As shown
413 in fig. 7A, the RP copy numbers are positively corrected with total transcription abundance value
414 of all RP genes (Supplemental table 10, Pearson correlation $r = 0.72$), supporting that the
415 increased RP gene dosage might have increased ribosome biogenesis by generating more RP
416 transcripts.

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

417 We then infer whether increased RP gene dosage is associated with better fermentative
418 ability. A previous study has measured various physiological characteristics for over 40 yeast
419 species (Hagman, et al. 2013), including 19 species examined in this study. We observed a
420 positive correlation between RP gene number and ethanol production efficiency ($r = 0.80$), and
421 glucose consumption rate ($r = 0.76$) (fig. 7B and C). We also observed a significant positive
422 correlation between total RP expression and both ethanol production efficiency ($r = 0.87$) and
423 glucose consumption rate ($r = 0.88$) (Supplementary fig. 2). These results suggest that the
424 increased RP expression by gene duplication might have enhanced these organisms' ability to
425 rapidly consuming glucose through the fermentation pathway.

426 ## Discussion

427 ### The preferential retention of RP duplicate genes was selection-driven

428 Our survey of 295 fungal genomes revealed that massive duplications of RP genes are not
429 prevalent. However, a significant increase in RP gene copy numbers had independently
430 occurred a small number of species in three distantly related lineages in fungi. WGD events
431 have played an important role in the expansion RP repertoire in the budding yeasts and pin
432 molds. In the budding yeasts, only ~10% of WGD ohnologs have survived, while 70.5% of RP
433 duplicates generated by WGD have been maintained in *S. cerevisiae*. As indicated by previous
434 studies, the survival rate of RP ohnologs is significantly higher than the other WGD ohnologs
435 (Papp, et al. 2003).

436 Our results suggested that RP genes in the fission yeasts were likely individually duplicated
437 by small-scale duplication events (SSD), such as retroposition. In general, the gene retention
438 rate of SSDs is much lower than ohnologs (half-life of 4 million years vs 33 million years)
439 (Hakes, et al. 2007), it is even lower in genes encoding macromolecular complexes due tp
440 evolutionary constraints imposed by gene dosage balance (Li, et al. 1996; Conant and Wolfe
441 2008). Similar to the fission yeasts, 99.8% of RP duplicates in mammals were found to be
442 generated by retroposition (Dharia, et al. 2014). However, almost all RP retroduplicates in
443 mammals become pseudogenes (Dharia, et al. 2014). Therefore, the high retention rates of
444 functional RP duplicates generated by SSDs in each fission yeasts are indeed unexpected. A
445 reasonable explanation is that the increased RP gene dosage have provided selective
446 advantages to these species, natural selection favored the retention of RP duplicate genes.

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

447 There is another line of evidence supporting that the retention of RP duplicate genes in
448 fission yeasts was selection-driven. The fission yeasts have been known to maintain a single
449 copy of genes in most gene families (Rhind, et al. 2011; Rajeh, et al. 2018). Based on our
450 orthologous group data (Supplementary table 4), in 86% (4069/4734) of fission yeast ortholog
451 groups, only a single copy gene is present in each of the four species (or 1:1:1:1 ortholog). Of
452 the gene families with gene duplication or loss in at least one fission yeast species, ribosomal
453 proteins account for 9.3% (62/665) of them, which is significantly overrepresented in this group
454 ($p < 10^{-5}$, Fisher exact test).

455 **How duplication and retention of RP genes contributed to the evolution of fermentative**
456 **ability in fungi**

457 Our data suggested that the retention of RP duplication genes might have been driven by
458 their contributions to the evolution of strong fermentative ability in these organisms. The
459 fermentative yeasts were believed to have gained a growth advantage through rapid glucose
460 fermentation in the presence of excess sugars (Piskur, et al. 2006). It was found that *S.*
461 *cerevisiae* often outgrew its non-fermentative competitors in co-culture experiments (Pérez-
462 Nevado, et al. 2006; Williams, et al. 2015). Fermentation is a much less efficient way to
463 generate energy. Through fermentation, each glucose molecule only yields 2 ATP from
464 glycolysis, compared to 32 ATP through mitochondrial oxidative phosphorylation pathway. In
465 sugar rich environments, fermentative organisms are able to produce more ATP per unit time by
466 rapidly consuming sugars through fermentation, providing selective advantages (Pfeiffer, et al.
467 2001; Pfeiffer and Morley 2014). The rapid glucose consumption was facilitated by the
468 increased glycolysis flux and more efficiently transporting glucose across cellular membranes.
469 The enhanced glycolytic activity is also present in many tumor cells, known as the "Warburg
470 effect" (Vander Heiden, et al. 2009; Diaz-Ruiz, et al. 2011). It has been shown that increased
471 copy number of genes related to glycolysis (Conant and Wolfe 2007) and glucose transporters
472 (Lin and Li 2011b) have played an important role in the switch of glucose metabolism. Similarly,
473 it is reasonable to propose that the increased RP gene dosage increased their transcript
474 abundance and ribosome biogenesis, resulting in increased biogenesis of glycolysis enzymes,
475 glucose transporters, and other building blocks for cell growth and proliferation.

476 Ribosome biosynthesis, however, comes with the opportunity cost of higher expression of
477 other cellular processes needed for cell viability and function. Maintaining one RP gene per
478 family may be advantageous for most other species to allow for greater Pol II transcription

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

479 potential for other genes. However, massive duplication of RP genes made it possible to rapidly
480 consume glucose through the low-efficient fermentative pathway, and at the same time maintain
481 a high growth rate. Such physiological chrematistic provided selective advantages to the
482 organisms in sugar rich environments, which well explained the high retention rate of RP
483 duplicates in fungal species.

484 **Gene duplication allows further increase of transcription abundance from highly**
485 **expressed RP genes**

486 One may argue that increased of ribosome biosynthesis can also be achieved by elevated
487 transcription activities of RP genes. It is probably true because we also observed elevated
488 expression level of RP genes in the two non-WGD yeasts with an intermediate level of ethanol
489 fermentation ability: *Lachancea thermotolerans* and *Lachancea waltii* (Hagman, et al. 2013).
490 Across all eukaryotic species, RP genes are the group of most abundantly transcribed genes in
491 eukaryotic cells, accounting for 50% of RNA polymerase II (Pol II) transcription (Warner 1999).
492 RP genes have the highest density of bound Pol II. 100 RP genes have on average >60% of the
493 maximum Pol II occupancy, while a majority of the genome only has <5% of the maximum Pol II
494 density (Venters and Pugh 2009). Therefore, there is limited room for further increase the
495 transcription abundance of RP genes. However, duplication of RP genes provides an additional
496 substrate on which Pol II can transcribe into RP mRNAs, reducing transcription as a rate-limiting
497 step in the process of ribosome biogenesis.

498 **Gene conversion facilitated retentions of RP duplicate genes**

499 The sequences of all RP families are highly conserved during the evolution of eukaryotes
500 due to their vital roles in many cellular functions (Korobeinikova, et al. 2012). Because
501 misfolding and misinteractions of highly abundant proteins can be more costly, proteins likely
502 RPs should have been under more functional constraints and evolved even slower than other
503 important proteins (Zhang and Yang 2015). It has shown that there is a strong evolutionary
504 constraint posed on the duplicability of genes encoding core components of protein complexes
505 (Li, et al. 2006). Accumulation of new mutations in duplicate genes could impact the stability of
506 protein complexes, posing selective disadvantages. The other structure component of
507 ribosomes, rRNA, is one of the best-known examples of concerted evolution (Liao 1999; Nei
508 and Rooney 2005).

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

509 Our study confirms that gene conversion following the duplication of RP genes appears to
510 be a universal path in fungal species. Through gene conversion, the sequences of paralogous
511 genes are homogenized, easing the new mutations accumulated in one of the paralogous
512 genes. It has been shown that highly-expressed genes are more likely to experience mRNA-
513 mediated gene conversion (Weng, et al. 2000; Schildkraut, et al. 2006). Thus, as a group of
514 most actively transcribed genes (Warner 1999), the repeated occurrence of gene conversion on
515 RP genes might be due to the highly expressed nature of RP genes. It was also found that the
516 promoter or flanking genomic sequences are much more divergence than coding sequences
517 between paralogous RP genes (Evangelisti and Conant 2010), further supporting gene
518 conversion in RPs was mediated by cDNA. Gene conversion could provide an additional layer
519 of protection on top of purifying selection to remove newly accumulated mutations (Evangelisti
520 and Conant 2010; Scienski, et al. 2015). Increased RP genes copies could be advantageous to
521 fungal species in sugar-rich environments through fermentative growth. Therefore, the repeated
522 occurrence of gene conversions between RP paralogous genes have contributed to the
523 maintenance of functional RP duplicates in these organisms by removing newly accumulated
524 mutations.

525 ## Materials and Methods

526 ### Data sources, identification and manual curation of RP repertoire

527 We obtained a list of RP genes from *S. cerevisiae* and *Sch. pombe* from the Ribosomal
528 Protein Gene Database (RPG)(Nakao, et al. 2004). We downloaded RefSeq protein sequence
529 data of 285 fungal genomes from NCBI (Table S1). In the first round of homologous sequences,
530 we used RP sequences from *S. cerevisiae* and *Sch. pombe* as queries to run BLASTP search
531 against the 285 proteomic data (Camacho, et al. 2009). For BLASTP search, we used e-value
532 cutoff of 1e-10, and only hits with a minimum alignment length of 50% of query sequences were
533 considered as homologous RP genes.

534 In the second round of homologous searches, we obtained the protein and genome
535 sequences of 34 fungal species from NCBI WGS, JGI and Yeast Gene Order Browser (YGOB)
536 (Byrne and Wolfe 2005; Maguire, et al. 2013) (Supplementary table 2). We first used 79 RP
537 protein sequences from *S. cerevisiae* and *Sch. pombe* as queries to search for homologous
538 sequences from the ten newly added fungal species using BLASTP. To identify RP sequences
539 not predicted by existing genome annotations, we conducted TBLASTN searches against all 34

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

540 genomic sequences. Manual inspections were performed to compare the BLASTP and
541 TBLASTN results to identify discrepant hits. For hits obtained by TBLASTN by not BLASTP, we
542 predicted the coding sequences (CDS) based on six frame translations of genomic sequences.
543 The exon-intron boundaries were determined based on the TBLASTN alignments and the
544 presence of GT/AG splice sites in flanking intron sequences. We also revised the predicted
545 protein sequences if there is a discrepancy in aligned regions between BLASTP and TBLASTN
546 results. The same gene prediction method was used to revise misannotated ORF.

547 **Construction of phylogenetic tree for RP genes**

548 We inferred the phylogeny for each of the 79 RP families using the RP protein sequences
549 collected from the 34 fungal species. Sequences were aligned through MUSCLE (Edgar 2004).
550 The molecular phylogenetic tree was inferred by the Maximum likelihood method using RAxML
551 with 100 bootstrap pseudo-replicates (Stamatakis 2006). The best-fit substitution model was
552 inferred by using ProtTest (Abascal, et al. 2005). As the best substitution models are the LG
553 model were identified for the majority of RP families, it was used in our phylogenetic
554 reconstruction. A discrete Gamma distribution [+G] and invariable sites [+I] was used to model
555 evolutionary rate differences among sites. We also constructed lineage-specific phylogenetic
556 trees for duplicate RP families using representative species from each of the three fungal
557 lineages using Neighbor-Joining method using MEGA 7 (Kumar, et al. 2016). For those RP
558 families with almost identical amino acid sequences, we used their CDS for construction of gene
559 trees to obtain better resolved phylogenetic trees (Supplementary Files 2-4).

560 **Homology microsynteny analysis**

561 We conducted microsynteny analysis for each RP gene family in the four
562 Schizosaccharomyces species and four Mucoromycotina species, including *R. delemar*, *R.*
563 *microspores*, *P. blakesleeanus* and *B. adelaidae*. Gene order information was retrieved from
564 genome annotations of each species obtained from NCBI. The orthologous gene groups in the
565 fours fission yeast species and four Mucoromycotina species were respectively identified using
566 the OrthoDB (Kriventseva, et al. 2014). For fission yeasts, we obtained a list of ten genes
567 surrounding each RP gene (five upstream and five downstream of RP gene). For the Mucorales
568 species, we extended our microsynteny analysis to a block of ten genes upstream and ten
569 downstream of RP gene due to their divergent genome structures.

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

570 **Estimation of substitution rates in intron and synonymous sites**

571 We calculated the substitution rates for every pair of duplicate RP genes for three species
572 representing the three fungal lineages with massive RP duplications, including *S. cerevisiae*,
573 *Sch. pombe*, *R. microspores*. The CDS and intron sequences were retrieved from NCBI, and
574 were aligned using MUSCLE. Synonymous substitution rate was calculated using Li-Wu-Luo
575 method with Kimura 2-parameter model (Li, et al. 1985) in MEGA 7 (Kumar, et al. 2016).
576 Nucleotide substitution rates in intron sequences were calculated using the Kimura 2-parameter
577 model in MEGA 7.

578 **Analysis of RP gene transcriptomic and physiological data**

579 The transcriptomic data of nine budding yeasts and two fission yeasts examined were
580 obtained from (McMillan, et al. 2019) based on CAGE. The expression abundance of an RP
581 gene was defined as the sum of transcripts initiated from all core promoters within 500 base
582 pairs upstream of its annotated start codon, which were normalized as TPM (tags per million
583 mapped reads, and each tag represent one sequenced transcript). The total transcription
584 abundance of RP genes in a species was calculated as the sum of TPM of all RP genes
585 identified in this species. For newly predicted RP genes that were not annotated in CAGE
586 datasets, we performed TBLASTN searches to determine their genomic locations of CDS and
587 obtain the expression abundance data using the same criteria. The ethanol production efficiency
588 and glucose consumption rate of 19 budding and fission yeast species were obtained from
589 (Hagman, et al. 2013). The ethanol production efficiency was measured as grams of ethanol
590 produced per gram of biomass per gram of glucose consumed. The glucose consumption rate
591 was measured as grams of glucose consumed per gram of biomass per hour. If multiple
592 biological replicates were measured for a single species, their average values were used for our
593 analysis.

594 **Acknowledgements**

595 This study was supported by the start-up fund and President's Research Fund from Saint
596 Louis University to ZL. We would like to thank Dr. Zhenglong Gu and Dr. Dapeng Zhang for
597 valuable discussions and suggestions to this work.

598

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

599 **References**

600

601 Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein
602 evolution. Bioinformatics 21:2104-2105.

603 Alexander MA, Jeffries TW. 1990. Respiratory efficiency and metabolite partitioning as
604 regulatory phenomena in yeasts. Enzyme Microb Technol 12:2-19.

605 Arnheim N, Krystal M, Schmickel R, Wilson G, Ryder O, Zimmer E. 1980. Molecular evidence
606 for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and
607 apes. Proc Natl Acad Sci U S A 77:7323-7327.

608 Barakat A, Szick-Miranda K, Chang F, Guyot R, Blanc G, Cooke R, Delseny M, Bailey-Serres J.
609 2001. The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome.
610 Plant Physiol 127:398-415.

611 Benny GL, Blackwell M. 2004. Lobosporangium, a new name for Echinosporangium Malloch,
612 and Gamsiella, a new genus for Mortierella multidivaricata. Mycologia 96:143-149.

613 Birchler JA, Veitia RA. 2012. Gene balance hypothesis: Connecting issues of dosage sensitivity
614 across biological disciplines. Proceedings of the National Academy of Sciences 109:14746-
615 14753.

616 Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner
617 JD, Rode CK, Mayhew GF, et al. 1997. The complete genome sequence of Escherichia coli K-
618 12. Science 277:1453-1462.

619 Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC, Gresham D, Boer VM,
620 Troyanskaya OG, Botstein D. 2008. Coordination of growth rate, cell cycle, stress response,
621 and metabolic activity in yeast. Mol Biol Cell 19:352-367.

622 Brown DD, Wensink PC, Jordan E. 1972. A comparison of the ribosomal DNA's of Xenopus
623 laevis and Xenopus mulleri: the evolution of tandem genes. J Mol Biol 63:57-73.

624 Byrne KP, Wolfe KH. 2005. The Yeast Gene Order Browser: combining curated homology and
625 syntenic context reveals gene fate in polyploid species. Genome research 15:1456-1461.

626 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
627 BLAST+: architecture and applications. BMC Bioinformatics 10:421.

628 Casola C, Conant GC, Hahn MW. 2012. Very low rate of gene conversion in the yeast genome.
629 Mol Biol Evol 29:3817-3826.

630 Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. 2007. Gene conversion:
631 mechanisms, evolution and human disease. Nat Rev Genet 8:762-775.

632 Conant GC, Wolfe KH. 2006. Functional partitioning of yeast co-expression networks after
633 genome duplication. PLoS Biol 4:e109.

634 Conant GC, Wolfe KH. 2007. Increased glycolytic flux as an outcome of whole-genome
635 duplication in yeast. Mol Syst Biol 3:129.

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

636 Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new
637 functions. Nat Rev Genet 9:938-950.

638 Corrochano LM, Kuo A, Marcet-Houben M, Polaino S, Salamov A, Villalobos-Escobedo JM,
639 Grimwood J, Alvarez MI, Avalos J, Bauer D, et al. 2016. Expansion of Signal Transduction
640 Pathways in Fungi by Extensive Genome Duplication. Curr Biol 26:1577-1584.

641 de Jong-Gubbels P, van Dijken JP, Pronk JT. 1996. Metabolic fluxes in chernostat cultures of
642 Schizosaccharomyces pombe grown on mixtures of glucose and ethanol. Microbiology
643 142:1399-1407.

644 Derr LK, Strathern JN. 1993. A role for reverse transcripts in gene conversion. Nature 361:170.

645 Dharia AP, Obla A, Gajdosik MD, Simon A, Nelson CE. 2014. Tempo and mode of gene
646 duplication in mammalian ribosomal protein evolution. PLoS ONE 9:e111721.

647 Diaz-Ruiz R, Rigoulet M, Devin A. 2011. The Warburg and Crabtree effects: On the origin of
648 cancer cell energy metabolism and of yeast glucose repression. Biochim Biophys Acta
649 1807:568-576.

650 Dudov KP, Perry RP. 1984. The gene family encoding the mouse ribosomal protein L32
651 contains a uniquely expressed intron-containing gene and an unmutated processed gene. Cell
652 37:457-468.

653 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
654 throughput. Nucleic Acids Res 32:1792-1797.

655 Evangelisti AM, Conant GC. 2010. Nonrandom survival of gene conversions among yeast
656 ribosomal proteins duplicated through genome doubling. Genome Biol Evol 2:826-834.

657 Hachmeister KA, Fung DY. 1993. Tempeh: a mold-modified indigenous fermented food made
658 from soybeans and/or cereal grains. Crit Rev Microbiol 19:137-188.

659 Hagman A, Sall T, Compagno C, Piskur J. 2013. Yeast "make-accumulate-consume" life
660 strategy evolved as a multi-step process that predates the whole genome duplication. PLoS
661 ONE 8:e68734.

662 Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. 2005. Estimating the tempo and mode
663 of gene family evolution from comparative genomic data. Genome Res 15:1153-1160.

664 Hakes L, Pinney JW, Lovell SC, Oliver SG, Robertson DL. 2007. All duplicates are not equal:
665 the difference between small-scale and genome duplication. Genome Biol 8:R209.

666 Jorgensen P, Nishikawa JL, Breitkreutz B-J, Tyers M. 2002. Systematic identification of
667 pathways that couple cell growth and division in yeast. Science 297:395-400.

668 Kaessmann H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and
669 evolutionary insights. Nat Rev Genet 10:19-31.

670 Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome
671 duplication in the yeast Saccharomyces cerevisiae. Nature 428:617-624.

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

672 Kenmochi N, Kawaguchi T, Rozen S, Davis E, Goodman N, Hudson TJ, Tanaka T, Page DC.
673 1998. A map of 75 human ribosomal protein genes. Genome Res 8:509-523.

674 Kito H, Abe A, Sujaya IN, Oda Y, Asano K, Sone T. 2009. Molecular characterization of the
675 relationships among Amylomyces rouxii, Rhizopus oryzae, and Rhizopus delemar. Biosci
676 Biotechnol Biochem 73:861-864.

677 Kondrashov FA, Kondrashov AS. 2006. Role of selection in fixation of gene duplications. J
678 Theor Biol 239:141-151.

679 Korobeinikova AV, Garber MB, Gongadze GM. 2012. Ribosomal proteins: structure, function,
680 and evolution. Biochemistry (Mosc) 77:562-574.

681 Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, Pozdnyakov IA, Ioannidis
682 P, Zdobnov EM. 2014. OrthoDB v8: update of the hierarchical catalog of orthologs and the
683 underlying free software. Nucleic Acids Res 43:D250-D256.

684 Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis
685 Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870-1874.

686 Kuzumaki T, Tanaka T, Ishikawa K, Ogata K. 1987. Rat ribosomal protein L35a multigene
687 family: molecular structure and characterization of three L35a-related pseudogenes. Biochimica
688 et Biophysica Acta (BBA)-Gene Structure and Expression 909:99-106.

689 Li B, Vilardell J, Warner JR. 1996. An RNA structure involved in feedback regulation of splicing
690 and of translation is critical for biological fitness. Proceedings of the National Academy of
691 Sciences 93:1596-1600.

692 Li L, Huang Y, Xia X, Sun Z. 2006. Preferential duplication in the sparse part of yeast protein
693 interaction network. Mol Biol Evol 23:2467-2473.

694 Li WH, Wu CI, Luo CC. 1985. A new method for estimating synonymous and nonsynonymous
695 rates of nucleotide substitution considering the relative likelihood of nucleotide and codon
696 changes. Mol Biol Evol 2:150-174.

697 Liao D. 1999. Concerted evolution: molecular mechanism and biological implications. Am J Hum
698 Genet 64:24-30.

699 Lin Z, Kong H, Nei M, Ma H. 2006. Origins and evolution of the recA/RAD51 gene family:
700 evidence for ancient gene duplication and endosymbiotic gene transfer. Proc Natl Acad Sci U S
701 A 103:10328-10333.

702 Lin Z, Li WH. 2011a. The evolution of aerobic fermentation in Schizosaccharomyces pombe
703 was associated with regulatory reprogramming but not nucleosome reorganization. Mol Biol
704 Evol 28:1407-1413.

705 Lin Z, Li WH. 2011b. Expansion of hexose transporter genes was associated with the evolution
706 of aerobic fermentation in yeasts. Mol Biol Evol 28:131-142.

707 Lin Z, Nei M, Ma H. 2007. The origins and early evolution of DNA mismatch repair genes -
708 Multiple horizontal gene transfers and co-evolution. Nucleic Acids Res 35:7591-7603.

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

709 Ma L-J, Ibrahim AS, Skory C, Grabherr MG, Burger G, Butler M, Elias M, Idnurm A, Lang BF,
710 Sone T. 2009. Genomic analysis of the basal lineage fungus Rhizopus oryzae reveals a whole-
711 genome duplication. PLoS Genet 5:e1000549.

712 Maguire SL, ÓhÉigeartaigh SS, Byrne KP, Schröder MS, O'Gaora P, Wolfe KH, Butler G. 2013.
713 Comparative genome analysis and gene finding in Candida species using CGOB. Mol Biol Evol
714 30:1281-1291.

715 McMillan J, Lu Z, Rodriguez JS, Ahn TH, Lin Z. 2019. YeasTSS: an integrative web database of
716 yeast transcription start sites. Database (Oxford) 2019.

717 Mendoza L, Vilela R, Voelz K, Ibrahim AS, Voigt K, Lee SC. 2014. Human Fungal Pathogens of
718 Mucorales and Entomophthorales. Cold Spring Harb Perspect Med 5.

719 Montagne J, Stewart MJ, Stocker H, Hafen E, Kozma SC, Thomas G. 1999. Drosophila S6
720 kinase: a regulator of cell size. Science 285:2126-2129.

721 Murata M, Nishiyori-Sueki H, Kojima-Ishiyama M, Carninci P, Hayashizaki Y, Itoh M. 2014.
722 Detecting expressed genes using CAGE. Methods Mol Biol 1164:67-85.

723 Nakao A, Yoshihama M, Kenmochi N. 2004. RPG: the ribosomal protein gene database.
724 Nucleic acids research 32:D168-D170.

725 Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. Annu
726 Rev Genet 39:121-152.

727 Ohno S. 1970. Evolution by gene duplication Berlin, New York: Springer-Verlag.

728 Panchy N, Lehti-Shiu M, Shiu SH. 2016. Evolution of Gene Duplication in Plants. Plant Physiol
729 171:2294-2316.

730 Papp B, Pal C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast.
731 Nature 424:194-197.

732 Pérez-Nevado F, Albergaria H, Hogg T, Girio F. 2006. Cellular death of two non-
733 Saccharomyces wine-related yeasts during mixed fermentations with Saccharomyces
734 cerevisiae. Int J Food Microbiol 108:336-345.

735 Pfeiffer T, Morley A. 2014. An evolutionary perspective on the Crabtree effect. Frontiers in
736 molecular biosciences 1:17.

737 Pfeiffer T, Schuster S, Bonhoeffer S. 2001. Cooperation and competition in the evolution of
738 ATP-producing pathways. Science 292:504-507.

739 Piskur J, Rozpedowska E, Polakova S, Merico A, Compagno C. 2006. How did Saccharomyces
740 evolve to become a good brewer? Trends Genet 22:183-186.

741 Rajeh A, Lv J, Lin Z. 2018. Heterogeneous rates of genome rearrangement contributed to the
742 disparity of species richness in Ascomycota. BMC Genomics 19:282.

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

743 Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ, Habib N, Wapinski I, Roy S, Lin MF,
744 Heiman DI, et al. 2011. Comparative functional genomics of the fission yeasts. Science
745 332:930-936.

746 Sanchez-Gracia A, Vieira FG, Rozas J. 2009. Molecular evolution of the major chemosensory
747 gene families in insects. Heredity (Edinb) 103:208-216.

748 Schildkraut E, Miller CA, Nickoloff JA. 2006. Transcription of a donor enhances its use during
749 double-strand break-induced gene conversion in human cells. Mol Cell Biol 26:3098-3105.

750 Schlotterer C, Tautz D. 1994. Chromosomal homogeneity of Drosophila ribosomal DNA arrays
751 suggests intrachromosomal exchanges drive concerted evolution. Curr Biol 4:777-783.

752 Scienski K, Fay JC, Conant GC. 2015. Patterns of Gene Conversion in Duplicated Yeast
753 Histones Suggest Strong Selection on a Coadapted Macromolecular Complex. Genome Biol
754 Evol 7:3249-3258.

755 Sidow A. 1996. Gen (om) e duplications in the evolution of early vertebrates. Curr Opin Genet
756 Dev 6:715-722.

757 Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
758 thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

759 Storici F, Bebenek K, Kunkel TA, Gordenin DA, Resnick MA. 2007. RNA-templated DNA repair.
760 Nature 447:338.

761 Straathof AJ, van Gulik WM. 2012. Production of fumaric acid by fermentation. Subcell Biochem
762 64:225-240.

763 Vander Heiden MG, Cantley LC, Thompson CB. 2009. Understanding the Warburg effect: the
764 metabolic requirements of cell proliferation. Science 324:1029-1033.

765 Venters BJ, Pugh BF. 2009. A canonical promoter organization of the transcription machinery
766 and its regulators in the Saccharomyces genome. Genome Res 19:360-371.

767 Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in Arabidopsis.
768 Science 290:2114-2117.

769 Warner JR. 1999. The economics of ribosome biosynthesis in yeast. Trends Biochem Sci
770 24:437-440.

771 Weng Y-s, Xing D, Clikeman JA, Nickoloff JA. 2000. Transcriptional effects on double-strand
772 break-induced gene conversion tracts. Mutation Research/DNA Repair 461:119-132.

773 Williams KM, Liu P, Fay JC. 2015. Evolution of ecological dominance of yeast species in high-
774 sugar environments. Evolution 69:2079-2093.

775 Wimberly BT, Brodersen DE, Clemons WM, Morgan-Warren RJ, Carter AP, Vonrhein C,
776 Hartsch T, Ramakrishnan V. 2000. Structure of the 30S ribosomal subunit. Nature 407:327-339.

777 Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast
778 genome. Nature 387:708-712.

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

779 Wool IG. 1979. The structure and function of eukaryotic ribosomes. Annu Rev Biochem 48:719-
780 754.

781 Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. 2014. Resolution of deep angiosperm
782 phylogeny using conserved nuclear genes and estimates of early divergence times. Nat
783 Commun 5:4956.

784 Zhang J. 2003. Evolution by gene duplication: an update. Trends in Ecology & Evolution 18:292-
785 298.

786 Zhang J. 2013. Gene duplication. In: Losos J, editor. The Princeton Guide to Evolution.
787 Princeton, New Jersey 08540 USA: Princeton University Press. p. 397-405.

788 Zhang J, Yang JR. 2015. Determinants of the rate of protein sequence evolution. Nat Rev
789 Genet 16:409-420.
790

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

791 **Figure Legends:**

792 **Figure 1: Schematic illustration of gene duplication patterns of 79 RP families in 34**
793 **fungal species.** Each row represents a fungal species, and each column represents an RP
794 gene family. The colors of a cell represent the numbers of gene copies identified in an RP family
795 of a species. The evolutionary relationship for 34 species, inferred based on amino acid
796 sequences of RNA polymerase II were shown to the left side of the matrix. The species names
797 were provided to the right of the matrix.

798

799 **Figure 2: Phylogenetic trees of representative RP gene families.** Phylogenetic trees of the
800 RPS6 gene family (A) and the RPS19 (B) gene family in 34 fungal species. The phylogenetic
801 trees were inferred by maximum likelihood (ML) method with 100 bootstrap tests. Only
802 bootstrap values above 50 are shown next to each node. The branches of budding yeast, fission
803 yeast, and Mucoromycota species are colored in red, blue and green respectively. The full
804 species names in taxa were provided in Supplementary table 2.

805

806 **Figure 3: Major types of tree topologies observed from 55 RP families with duplicates in**
807 **the budding yeasts. (A)** Phylogenetic relationships of RPL12 genes from five representative
808 budding yeast species. In this case, the two copies of RPL12 genes in *S. cerevisiae* for a
809 species-specific clade. 21 RP gene families demonstrate similar tree topology, as indicated by
810 number “21” in a yellow dot; (B) Phylogenetic relationships of RPL11 genes. The two copies of
811 RPL11 genes in *S. cerevisiae* and *S. mikatae* form a well-supported clade, and the orthologous
812 genes between *S. cerevisiae* and *S. mikatae* are more closely related to each other. Seven RP
813 gene families have a similar tree topology; (C) Phylogenetic relationships of RPL16 genes. Each
814 of RPL16 duplicate genes in *S. cerevisiae* is more closely related to their orthologous genes in
815 *S. mikatae* and *S. eubayanus*. 24 RP gene families share a similar tree topology; (D)
816 Phylogenetic relationships of RPP1 genes. Each of RPP1 duplicate genes in *S. cerevisiae* is
817 more closely related to their orthologous genes in the five WGD species. Seven RP gene
818 families demonstrate a similar tree topology. (E) The distribution of RP families with different
819 termination points of concerted evolution based on evolutionary relationships of RP duplicate
820 genes in *S. cerevisiae*. The numbers on each tree branch represent the numbers of RP families
821 that have terminated concerted evolution at the indicated evolutionary stages.

822 **Figure 4: The origin and evolution of RP genes in the fission yeasts.** (A) A schematic
823 illustration of microsynteny structures of RPL11 genes in four budding yeasts. The microsynteny
824 blacks of RPL11A are shared by all fission yeasts, so are the RPL11 genes, supporting that the
825 duplication of RPL11 occurred prior to the divergence of the fission yeasts. The number in each
826 box represents its orthologous group ID. (B) Phylogenetic relationships of RPL11 genes in four
827 fission yeast species. The RPL11 duplicate genes in *Sch. pombe* are more closely related to
828 each other than to any orthologous genes. 45 RP gene families demonstrate similar tree
829 topology; (C) Phylogenetic relationships of RPS17 genes. Each of RPL17 duplicate genes in
830 *Sch. pombe* are more closely related to their orthologous genes in *Sch. octosporus* and *Sch.*
831 *cryophilus*. Nine RP gene families demonstrate similar tree topology; (D) Phylogenetic
832 relationships of RPS5 genes. Each copy of RPS5 duplicates in *Sch. pombe* are more closely
833 related to their orthologous genes. Four RP gene families demonstrate similar tree topology. (E)
834 The distribution of RP families with different termination points of concerted evolution in *Sch.*
835 *pombe*. The numbers on each tree branch indicate the numbers of RP families that have
836 terminated concerted evolution at the indicated evolutionary stages.

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

837 **Figure 5: The origin and evolution of RP genes in pin molds.** (A) A schematic illustration of
838 microsynteny structures of RPL3 genes in three pin molds. The shared microsynteny structure
839 suggested that the first duplication event of RPL3 genes have occurred prior to the divergence
840 of the pin molds species. The two RPL3 genes have experienced a second round of duplication
841 by WGD in *R. delemar.* (B) Phylogenetic relationships of RPL3 genes in three pin molds. The
842 RPL3 paralogous genes in *R. microsporus* are more closely related to each other than to any
843 orthologous genes. 57 RP gene families demonstrate similar tree topology; (C) Phylogenetic
844 relationships of RPL38 genes in Mucorales. Each of RPL38 duplicate genes in *R. microsporus*
845 is more closely related to their orthologous genes in *R. delemar*. 17 RP gene families
846 demonstrate similar tree topology; (D) Phylogenetic relationships of RPS20 genes in three pin
847 molds. Each RPS20 duplicate gene in *R. microsporus* is more closely related to their
848 orthologous genes. Two RP gene families demonstrate similar tree topology. (E) The
849 distribution of RP families with different termination points of concerted evolution during the
850 evolution of *R. microsporus*. The numbers on each tree branch represent the estimated
851 numbers of RP families that have terminated concerted evolution at the indicated evolutionary
852 stage.

853

854 **Figure 6: Distinct mutations rates in substitution sites and introns between RP**
855 **paralogous genes.** The distributions of mutation rates in intron and synonymous sites between
856 RP paralogous genes in *S. cerevisiae* (A), *Sch. pombe* (B) and *R. microspores* (C). Scatter plots
857 of mutation rates in intron against synonymous sites between RP paralogous genes in in *S.*
858 *cerevisiae* (D), *Sch. pombe* (E) and *R. microspores* (F).

859

860 **Figure 7: RP gene copy numbers are positively corrected with RP transcription**
861 **abundance, ethanol production ability, and glucose consumption rates.** (A) A scatter plot
862 between the RP copy number and total RP transcription abundance in 11 yeast species. (B) A
863 scatter plot between the RP copy number and ethanol production efficiency in 19 yeast species.
864 (C) A scatter plot between the RP copy number and glucose consumption rate in 19 yeast
865 species.

866

bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

867 **Supplementary Figures:**

868 **Supplementary fig. 1: The phylogenetic relationships of 34 representative fungal species.**
869 The phylogenetic tree was inferred based on amino acid sequences of RPA polymerase II using
870 ML method by RAxML with 100 bootstrap replicates.

871 **Supplementary fig. 2: The total transcription abundance of RP genes is positively**
872 **correlated with ethanol production efficiency and glucose consumption rates.** (A) A
873 scatter plot between total RP transcription abundance and ethanol production efficiency in 9
874 yeast species. (B) A scatter plot between the total RP transcription abundance and glucose
875 consumption rate in 9 yeast species.

876 **Supplementary Tables:**

877 Table S1: List of numbers of RP genes identified in 285 fungal species

878 Table S2: List of numbers of RP gene identified in 34 representative fungal species based on
879 manual curation.

880 Table S3: List of RP genes identified in 34 representative fungal species

881 Table S4: List of orthologous groups identified in four fission yeasts

882 Table S5: Microsynteny data of the 79 RP families in fission yeasts

883 Table S6: Statistics of intron number in RP genes in *S. cerevisiae* and *Sch. pombe*.

884 Table S7. Number of share homologous genes between microsynteny blocks near the RP
885 paralogous genes in three Mucorales species

886 Table S8. Number of share homologous genes between microsynteny blocks near the RP
887 orthologous genes in three Mucorales species

888 Table S9: Mutation rates of synonymous sites and intron in *S. cerevisiae*, *Sch. pombe* and *R.*
889 *microspores*

890 Table S10: Transcription abundance of RP genes in 11 species based on CAGE data

891
892 **Supplementary Files:**
893
894 Supplementary File 1: Phylogenetic trees of RP families from the 34 representatives fungal
895 species.

896 Supplementary File 2: Phylogenetic trees of RP families in the five WGD budding yeast species.

897 Supplementary File 3: Phylogenetic trees of RP families in the four fission yeast species.

898 Supplementary File 4: Phylogenetic trees of RP families in the four Mucoromycotina species.

899

900

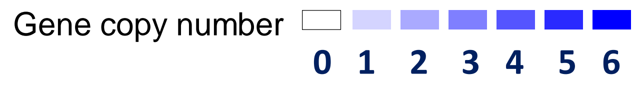

Saccharomyces cerevisiae
Saccharomyces mikatae
Saccharomyces eubayanus
Saccharomyces uvarum
Naumovozyma castellii
Naumovozyma dairenensis
Candida glabrata
Nakaseomyces bacillisporus
Tetrapisispora phaffii
Vanderwaltozyma polyspora
Zygosaccharomyces rouxii
Lachancea waltii
Lachancea thermotolerans
Lachancea kluyverii NRRL
Kluyveromyces dobzhanskii
Kluyveromyces lactis
Kluyveromyces marxianus
Eremothecium sinecaudum
Komagataella pastoris
Dekkera bruxellensis
Candida albicans
Yamadazyma tenuis
Yarrowia lipolytica
Schizosaccharomyces octosporus
Schizosaccharomyces cryophilus
Schizosaccharomyces pombe
Schizosaccharomyces japonicus
Pneumocystis murina
Rhizopus delemar
Rhizopus microsporus
Phycomyces blakesleeanus
Bifiguratus adelaidae
Lobosporangium transversale
Rhizophagus irregularis
RPP0 RPP1 RPP2 RPL1 RPL2 RPL3 RPL4 RPL5 RPL6 RPL7 RPL8 RPL9 RPL10 RPL11 RPL12 RPL13 RPL14 RPL15 RPL16 RPL17 RPL18 RPL19 RPL20 RPL21 RPL22 RPL23 RPL24 RPL26 RPL25 RPL27 RPL28 RPL28e RPL29 RPL30 RPL31 RPL32 RPL33 RPL34 RPL35 RPL36 RPL37 RPL38 RPL39 RPL40 RPL41 RPL42 RPL43 RPS0 RPS1 RPS2 RPS3 RPS4 RPS5 RPS6 RPS7 RPS8 RPS9 RPS10 RPS11 RPS12 RPS13 RPS14 RPS15 RPS16 RPS17 RPS18 RPS19 RPS20 RPS21 RPS22 RPS23 RPS24 RPS25 RPS26 RPS27 RPS28 RPS29 RPS30 RPS31
Gene copy number
0 1 2 3 4 5 6
Whole genome duplication
Fermentative lifestyle

Fig. 2

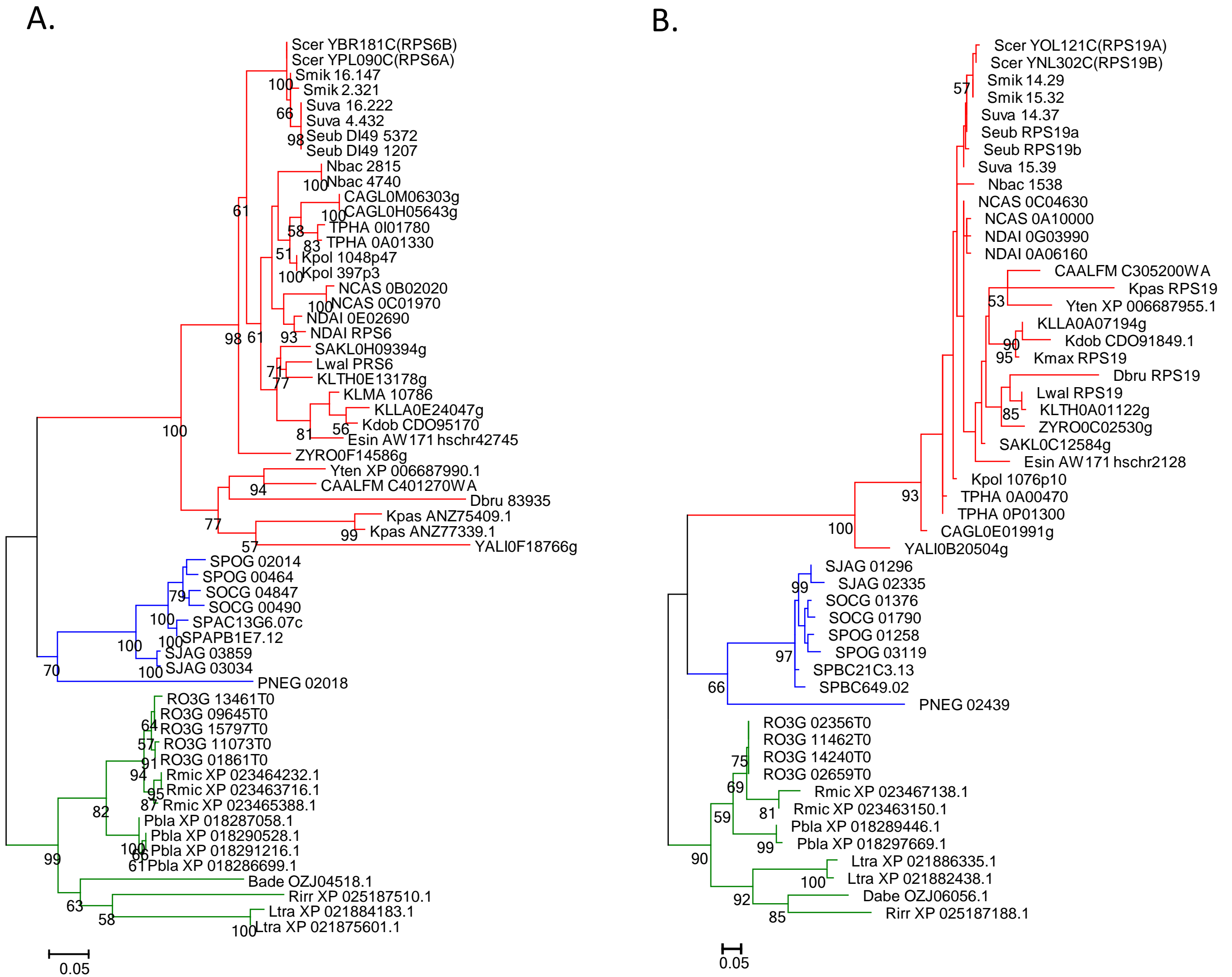


Fig. 3

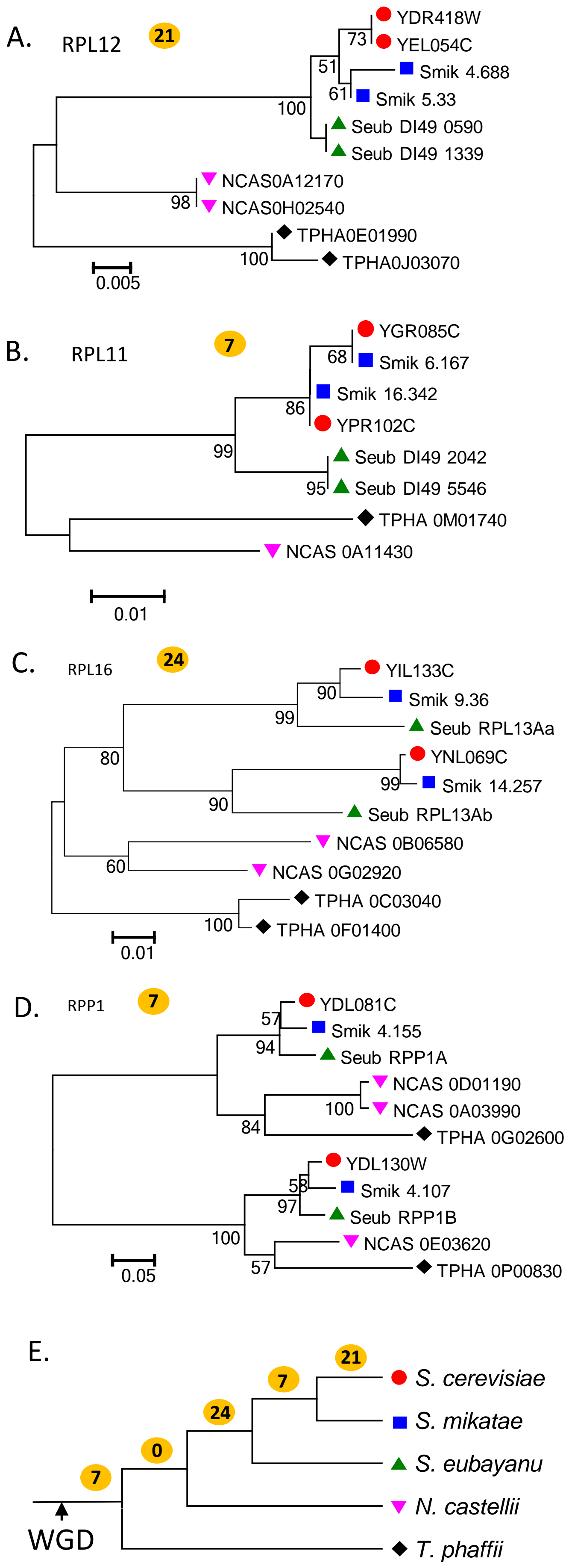
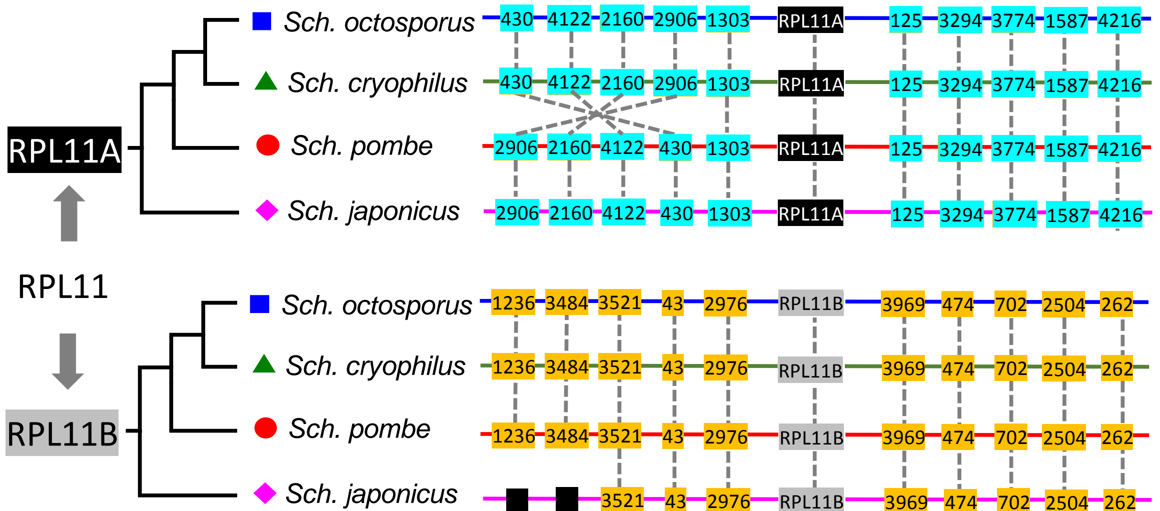
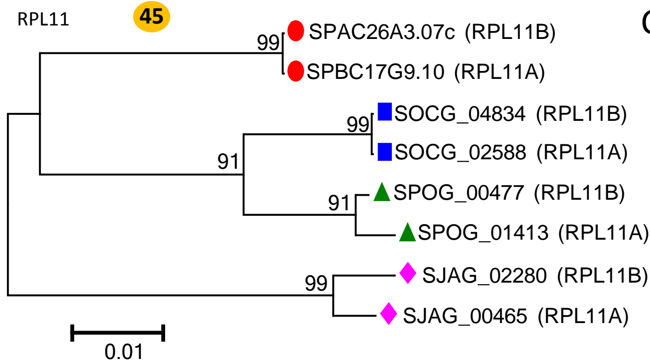


Fig. 4

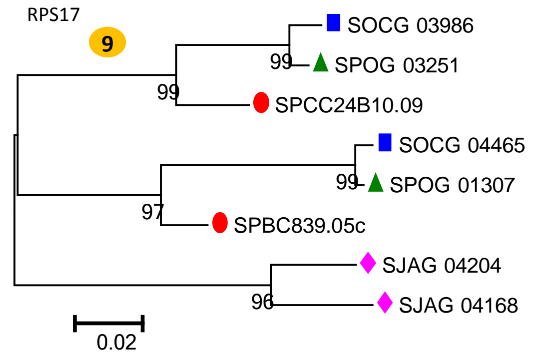
A.



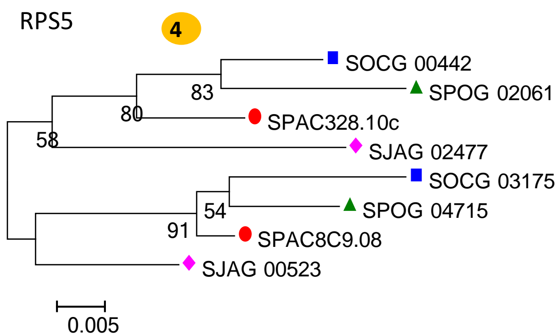
B.



C.



D.



E.

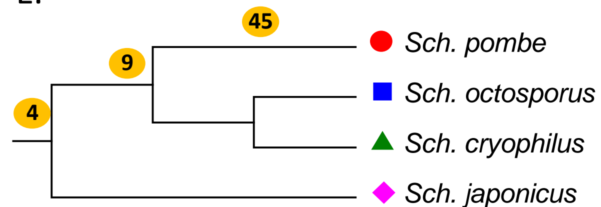
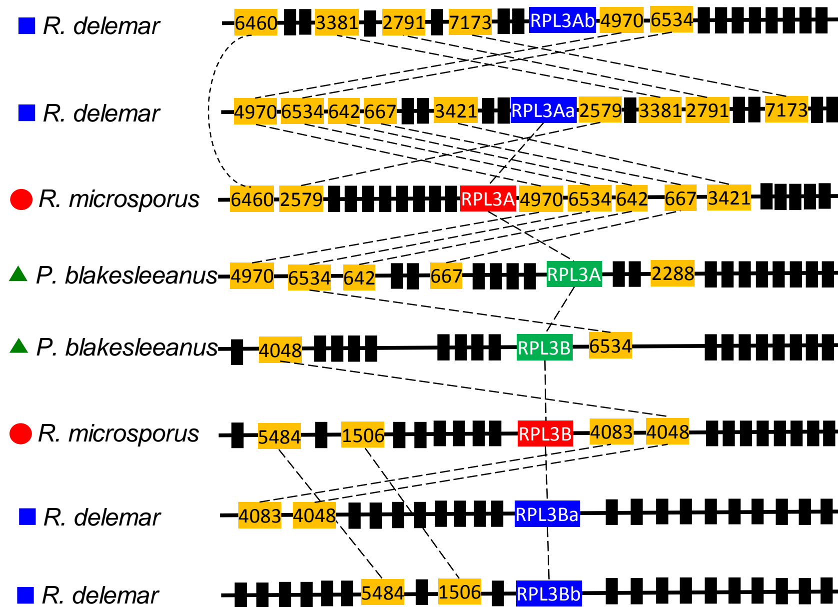
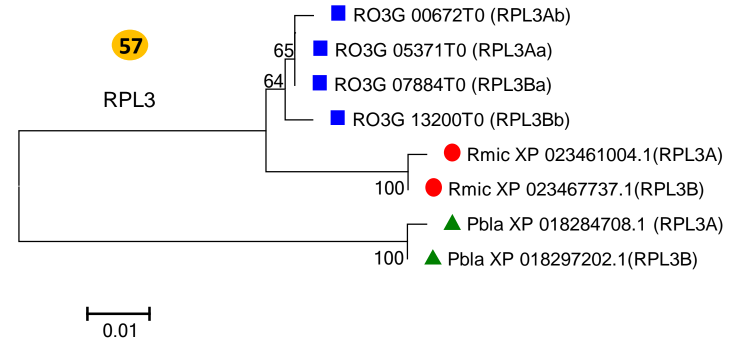
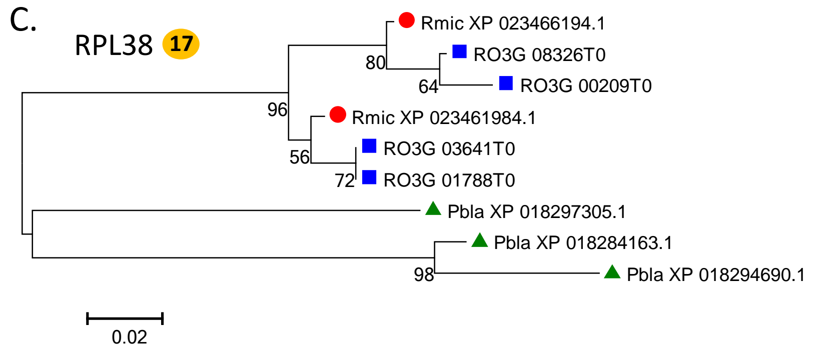
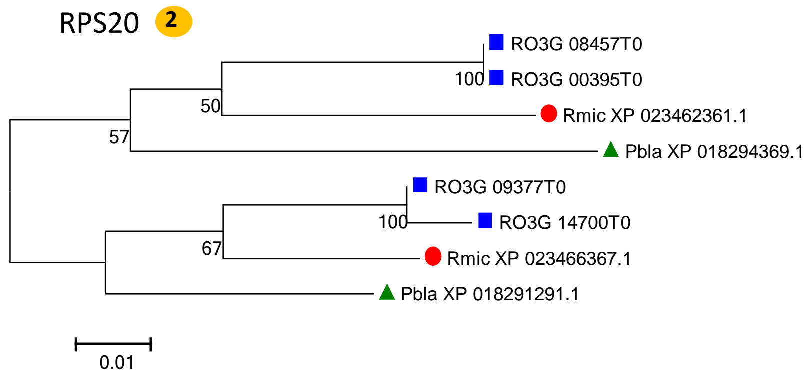


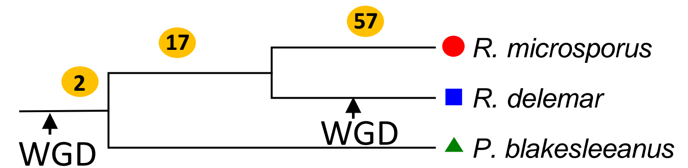
Fig. 5



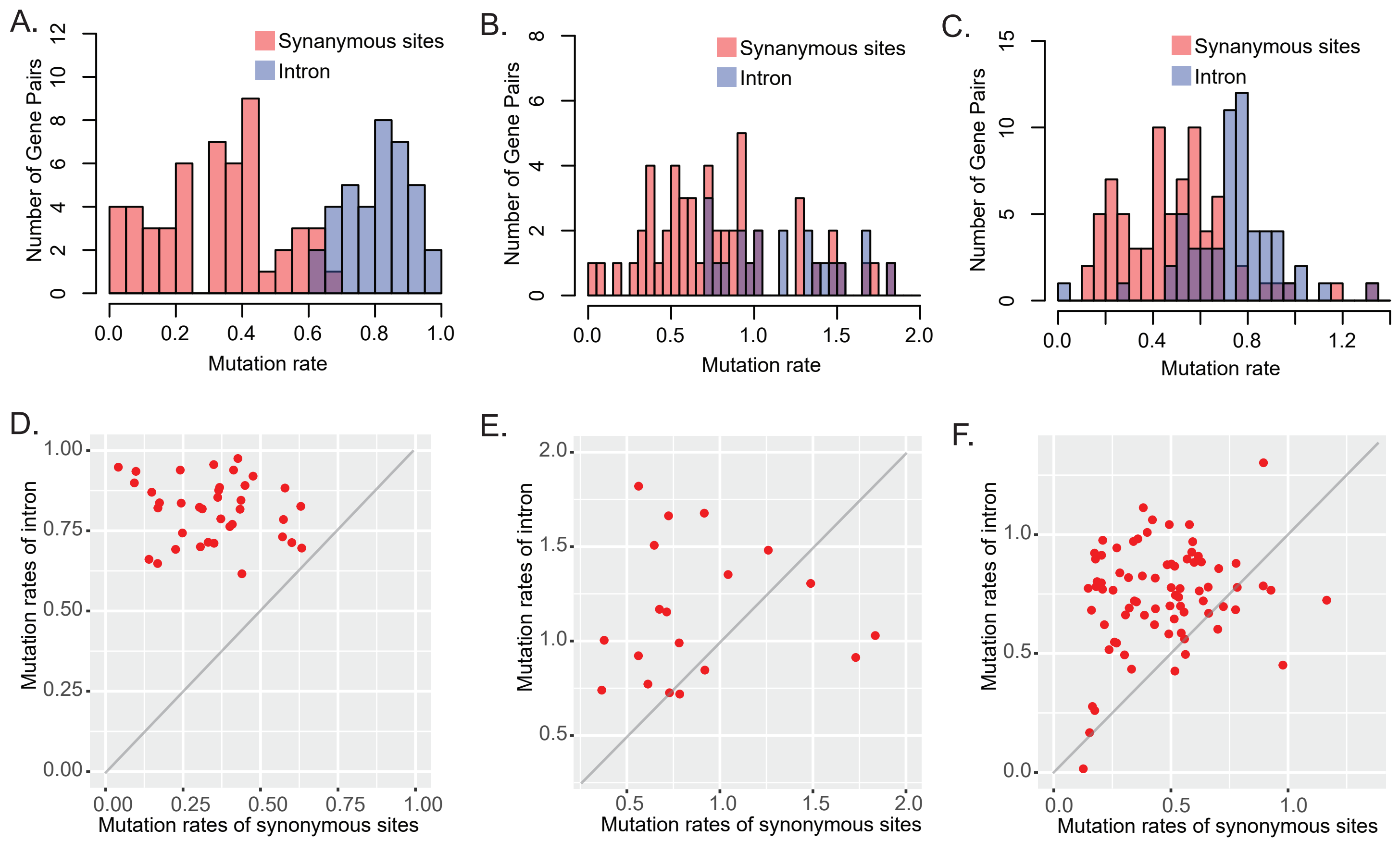








bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/751792; this version posted August 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.