

1 Analysis of fungal genomes reveals commonalities of intron loss/gain and functions in 2 intron-poor species

3

4 Chun Shen Lim¹, Brooke N. Weinstein^{2,3}, Scott W. Roy^{2,3,*} and Chris M. Brown^{1,*}

5

6 ¹ Department of Biochemistry, School of Biomedical Sciences, University of Otago, Dunedin, New
7 Zealand

8 ² Quantitative & Systems Biology, School of Natural Sciences, University of California-Merced,
9 Merced, CA, USA

10 ³ Department of Biology, San Francisco State University, San Francisco, CA, USA

11 * Corresponding authors. Emails: chris.brown@otago.ac.nz, scottwroy@gmail.com

12

13 **ABSTRACT** (249 words)

14 Current evolutionary reconstructions predict that early eukaryotic ancestors including both the last
15 common ancestor of eukaryotes and of all fungi had intron-rich genomes. However, some extant
16 eukaryotes have few introns, raising the question as to why these few introns are retained. Here we
17 have used recently available fungal genomes to address this question. Evolutionary reconstruction of
18 intron presence and absence using 263 diverse fungal species support the idea that massive intron loss
19 has occurred in multiple clades. The intron densities estimated in the fungal ancestral states differ
20 from zero to 8.28 introns per one kbp of protein-coding gene. Massive intron loss has occurred not
21 only in microsporidian parasites and saccharomycetous yeasts (0.01 and 0.05 introns/kbp on average,
22 respectively), but also in diverse smuts and allies (e.g. *Ustilago maydis*, *Meira miltonrushii* and
23 *Malassezia globosa* have 0.06, 0.10 and 0.20 introns/kbp, respectively). To investigate the roles of
24 introns, we searched for their special characteristics using 1302 orthologous genes from eight
25 intron-poor fungi. Notably, most of these introns are found close to the translation initiation codons.
26 Our transcriptome and translatoome data analyses showed that these introns are from genes with both
27 higher mRNA expression and translation efficiency. Furthermore, these introns are common in
28 specific classes of genes (e.g. genes involved in translation and Golgi vesicle transport), and rare in
29 others (e.g. base-excision repair genes). Our study shows that fungal introns have a complex
30 evolutionary history and underappreciated roles in gene expression.

31

32 **INTRODUCTION**

33 Spliceosomal introns are ubiquitous in eukaryotes. They are removed from all regions of the
34 transcripts including the untranslated regions (UTRs) as well as coding sequences (CDS) (De Conti et
35 al. 2013; Shi 2017; Lim et al. 2018). Early studies proposed that introns may be involved in

generating multi-domain genes by exon shuffling (Logsdon et al. 1995; Patthy 2003; Stoltzfus 2004; Sverdlov et al. 2005), and promoting intragenic recombination for higher fitness (Gilbert 1978; Tonegawa et al. 1978; Comeron and Kreitman 2000; Duret 2001). Notable experimentally supported roles of introns in eukaryotes include: (i) generating protein diversity by alternative splicing (Kempken 2013; Irimia and Roy 2014), (ii) harboring noncoding RNA (ncRNA) genes, such as snoRNAs and microRNAs (Chorev and Carmel 2012; Jo and Choi 2015), (iii) maintaining genome stability by decreasing the formation of DNA-RNA hybrids called R-loops (Niu 2007; Bonnet et al. 2017), (iv) intron-mediated enhancement of gene expression (Niu and Yang 2011; Gallegos and Rose 2015; Laxa 2016; Shaul 2017), (v) harboring binding sites for transcriptional or posttranscriptional regulators of gene expression (Rose 2018), (vi) allowing for an additional level of post-transcriptional regulation through regulation of RNA splicing (Witten and Ule 2011), and (vii) triggering nonsense-mediated decay in unspliced or partially spliced mRNAs through exon junction complexes (EJCs) (Mekouar et al. 2010; Grützmann et al. 2014; Zhang and Sachs 2015; Hellens et al. 2016). Recently, we have uncovered an unexpected relationship between introns and translation, suggesting a role of 5'UTR introns in promoting translation of upstream open reading frames (Lim et al. 2018).

The most well-studied introns are those that interrupt the protein-coding regions of genes. Extensive computational studies suggest that the last eukaryotic common ancestor (LECA) had a density of introns of about 4 introns/kbp (the number of introns per one kbp of CDS on average) (Stajich et al. 2007; Csuros et al. 2011; Koonin et al. 2013; Irimia and Roy 2014). Notably, a study of 99 eukaryotic genomes has revealed a surprising variability of intron densities, ranging from 0.1 introns/kbp in the baker's yeast *Saccharomyces cerevisiae* to 7.8 introns/kbp in *Trichoplax adhaerens* (Csuros et al. 2011). Counterintuitively, *T. adhaerens* is one of the simplest free-living multicellular animals (Srivastava et al. 2008). The large variability of intron densities owes to remarkable differences in rates of intron loss through eukaryotic evolution (Roy and Gilbert 2005; Csuros et al. 2011) and may, in part, be due to the transposable properties of some spliceosomal introns (Roy 2004; Worden et al. 2009; Huff et al. 2016; Wu et al. 2017). Several models have also been proposed for intron loss, in particular, through genomic deletion (Loh et al. 2008; Yenerall et al. 2011; Zhu and Niu 2013a) and recombination of cDNA with genomic DNA (Fink 1987; Roy and Gilbert 2005; Zhang et al. 2010; Zhu and Niu 2013b).

As of April 2020, a total of 6,337 fungal genome assemblies were available in NCBI Genome. Fungi and their genomes are of interests for many reasons, notably as food, and plant/animal pathogens/symbionts, and for biotechnology applications (Sapountzis et al. 2015; Wheeler et al. 2017; Chan et al. 2018; Kijpornyongpan et al. 2018; Uhse et al. 2018). As fungi belong to a diverse group of organisms evolving over the past 900 million years (Dornburg et al. 2017; Kumar et al. 2017), some fungal clades have undergone massive loss of introns, in particular, the intracellular parasites

71 microsporidia as well as saccharomycetous yeasts (Byrne and Wolfe 2005; Neuvéglise et al. 2011;
72 Hooks et al. 2014; Corradi 2015; Han and Weiss 2017; Whelan et al. 2019; Priest et al. 2020; Wang et
73 al. 2020). For instance, only 4% of *S. cerevisiae* genes have introns. In contrast, some other fungi, for
74 example, the facultative pathogen *Cryptococcus neoformans*, have a relatively high intron density of 4
75 introns/kbp (Stajich et al. 2007; Csuros et al. 2011).

76 Previous results have suggested that frequent intron loss events, relatively few instances of
77 intron gain, and the retention of ancestral introns characterize the evolution of introns throughout most
78 fungal lineages (Csurös et al. 2007; Stajich et al. 2007; Csuros et al. 2011). With thousands of fungal
79 genomes available to date (Priest et al. 2020), it is timely to revisit the ancestral states and scale of
80 intron gain or loss in the fungal kingdom. Our analysis includes representatives from nearly all
81 phylum-level clades, including the early-diverging Blastocladiomycota, Chytridiomycota,
82 Mucoromycota, Zoopagomycota, Cryptomycota, and Microsporidia phyla. The diversity of
83 intron-exon structures and the wealth of kingdom-wide genomic resources of fungi make them
84 excellent models for studying the intron gain and loss dynamics and the functional roles of introns
85 (Priest et al. 2020). Here we analyzed the introns of 644 fungal genomes, inferring ancestral states,
86 conservation, and functional processes.

87

88 RESULTS

89 Evolutionary reconstruction reveals high ancestral intron densities and a general bias towards 90 intron loss over intron gain

91 We aligned protein sequences and mapped corresponding intron positions for 1445 sets of orthologous
92 genes from 263 fungal species. We reconstructed the evolutionary history of intron gain or loss
93 among these species ([Figure 1](#); see Materials and Methods). These 263 species represented a wide
94 variety of intron densities, from various intronless Microsporidia to 4.8 introns/kbp in the chytrid
95 *Gonapodya prolifera*. This reconstruction revealed a remarkably dynamic and diverse history of
96 intron loss and, with many episodes of massive intron loss and/or gain coupled to general stasis within
97 large clades of organisms (e.g., very low intron densities within all Microsporidia and similar intron
98 densities among nearly all Pezizomycotina). Most strikingly, we reconstructed very high intron
99 densities (Figure 2; Table 1), with some 8.3 introns/kbp reconstructed in the fungal ancestor. While it
100 may be counterintuitive that the ancestral fungus harbored nearly twice as many introns as any
101 modern fungus in this dataset, this finding is in keeping with previous results showing a general bias
102 towards intron loss over intron gain in many lineages, and echoes the finding of considerably higher
103 intron densities in alveolate ancestors than in modern alveolates (Csuros et al. 2011). While these
104 results are in general agreement with previous studies that inferred intron-rich ancestral fungi (Stajich
105 et al. 2007; Csuros et al. 2011; Grau-Bové et al. 2017), our inferred densities are considerably higher,

likely due to improved model specification made possible by greater species density. Interestingly, our reconstructed value is close to the inferred intron content of the animal ancestor (8.8 introns/kbp) in a study using the same reconstruction method on a smaller, eukaryote-wide dataset (Csuros et al. 2011). In contrast to intron-rich ancestral states, almost three-quarters of fungi have maintained less than 10% of the intron density of the last fungal ancestor (191 of 263 species; Figure 1; see also Supplementary Table S1 and S2).

These results also illuminate the history of massive intron loss in these same two lineages. Many studies have found that the obligate intracellular microsporidian parasites have zero or few introns (Keeling et al. 2010; Cuomo et al. 2012; Peyretailade et al. 2012; Corradi 2015; Desjardins et al. 2015; Han and Weiss 2017; Mikhailov et al. 2017; Ndikumana et al. 2017) and that saccharomycetous yeasts have lost most of their introns (Stajich et al. 2007; Csuros et al. 2011; Hooks et al. 2014). For both remarkable groups, our analysis includes newly available genomes including relatively intron-rich sister species (*Rozella allomyces* (2.7 introns/kbp) for Microsporidia and *Lipomyces starkeyi* (1.2 introns/kbp) for Saccharomycotina), allowing for improved resolution of the history of these organisms. In both lineages our reconstructions reveal a massive intron loss event leading to the ancestor of a large clade of intron-poor organisms. However, whereas in Microsporidia this loss event occurred in the ancestor of the group after divergence from Cryptomycota, for saccharomycetous yeast this massive loss event occurs within the group, after divergence of *L. starkeyi*.

A general bias towards intron loss punctuated by several independent episodes of intron gain

A bias towards intron loss over intron gain is seen across the fungal tree. This is evident not only in Microsporidia and Saccharomycotina but also in groups with more moderate intron densities, including the filamentous ascomycetes Pezizomycotina (120 of 122 species) and smuts/allies Ustilaginomycotina (15 of 20 species). Indeed, we find a striking bias towards intron loss over gain. Among branches estimated to have undergone at least 5% change in intron density, ten times as many have more loss than gain. Remarkably, a bias is seen even for lineages with very little change, in which intron loss outweighs gain three-fold (Supplementary Figure S1).

While ongoing intron loss is characteristic of most lineages, our results indicate several substantial episodes of intron gain. Within Basidiomycotina, we estimated a 26% increase in intron density leading to the ancestor of Ustilaginomycotina and an 18% increase in the ancestor of Pucciniomycotina. The most substantial intron gains occurred, unexpectedly, within the famously intron-poor lineages Microsporidia and Saccharomycotina. We inferred substantial, secondary independent intron gain in two extant microsporidians, (*Nosema bombycis* and *Nosema apis*) and four saccharomycetous yeasts (*Scheffersomyces stipitis*, *Candida maltosa*, *Pichia kudriavzevii*, and

141 *Spathaspora passalidarum*). While preliminary analysis suggests the reality of some of these gains, it
142 is worthy of note that, given the small absolute number of gains involved (leading to <1 intron/kbp),
143 further detailed analysis will be necessary to confirm these episodes.

144

145 **No relationship of genome size to intron number**

146 Given that genome size has been argued to relate to intron number, organismal complexity, population
147 size and generation time (Lynch and Conery 2003; Koonin et al. 2013), we examined the relationship
148 between genome size and intron density using phylogenetic independent contrasts. Remarkably, we
149 found no evidence for a positive relationship between genome size and intron number — indeed, the
150 correlation is slightly and non-significantly negative (Figure 3; Spearman's $\rho = -0.070$, p -value =
151 0.41).

152

153 **Functional biases of intron-containing genes in intron-poor species**

154 To better understand the evolutionary forces responsible for maintenance of introns through evolution,
155 we chose eight intron-poor species (with intron densities <10% of the fungal ancestor), identified
156 orthologous genes, and analyzed the selective pressures on intron-containing and intronless genes.
157 These species comprised of *S. cerevisiae* and *Candida dubliniensis* in Saccharomycotina,
158 *Cyphellophora europaea* and *Sporothrix schenckii* in Pezizomycotina, and *Ustilago maydis*,
159 *Pseudozyma hubeiensis*, *Meira miltonrushii* and *Malassezia sympodialis* in Ustilaginomycotina
160 (Figure 1, green filled circles), representing six separate massive reductions in intron number.

161 Comparison of intron-containing genes with intronless genes revealed a number of
162 differences. We found that intron-containing genes are less likely to have undergone recent positive
163 selection on their protein-coding meaning (Figure 4A, one-sided Fisher's exact tests, p -value < 0.05).

164 We also found an association with gene duplication. Significantly higher proportions of the
165 intron-containing genes in *S. cerevisiae* and *U. maydis* are duplicated (Figure 4B, two-sided Fisher's
166 exact tests, p -value < 0.05). While this finding in *S. cerevisiae* could largely be explained by the
167 previously-noted concentration of introns in ribosomal protein genes, in which introns facilitate
168 cross-regulation among paralogous genes (Pleiss et al. 2007; Parenteau et al. 2011; Petibon et al.
169 2016; Parenteau and Abou Elela 2019), it cannot explain the bias in *U. maydis*.

170

171 **Concordance of the presence of introns in orthologs in species with independent massive intron** 172 **loss**

173 If introns carry useful functions, we hypothesize that they should be maintained in the orthologs of the
174 intron-poor species. We determined ratios of the intron-containing orthologs among the intron-poor
175 species and compared these ratios with null expectations (i.e. the proportions of orthologs with introns

for the intron-poor species; Figure 5A). Indeed, these orthologs tend to harbor introns concordantly (Figure 5B). Strikingly, two orthologous genes, *RPL7B* and *NOG2* have conserved intron positions in all eight studied intron-poor species (Figure 6). *RPL7B* encodes a ribosomal 60S protein whereas *NOG2* encodes a putative GTPase-associated pre-60S ribosomal subunit, in which their introns harbor a box C/D and a box H/ACA snoRNAs, respectively [snR59 and snR191 in *S. cerevisiae*, respectively; *Saccharomyces* Genome Database (Cherry et al. 2012)]. This suggests that introns with conserved positions have functions (e.g. as snoRNAs). Interestingly, this conservation may not be trackable to LECA as snoRNA genes are mobile (Weber 2006; Luo and Li 2007; Schmitz et al. 2008; Hoepfner and Poole 2012), however this suggests an ancient association within fungi.

We next asked whether the distances of introns with respect to translation initiation codons are conserved. We compared the distribution of the first introns in the coding genes with two null distributions (Figure 7). Both null distributions were generated using *S. cerevisiae* genes. The first null distribution was generated based on our observation that *S. cerevisiae* has only 1 or 2 introns in its coding genes (GCA_000146045, total 282 spliceosomal introns, 273 of 6619 coding genes have introns). Therefore, the expected distances of the first introns from the start codons are either half or one-third of the CDS lengths (Figure 7, dotted lines in light gray). The second null distribution was generated in a similar way but including UTR introns and centering at the transcription start or termination sites (Figure 8, dashed lines in dark gray). We then divided the genes into two groups, genes related to translation and other classes of genes using gene ontology (GO) terms (Cherry et al. 2012).

We observed that introns are closer to the initiation codons than these null distributions (Figure 7), which is in agreement with previous studies (Bon et al. 2003; Mourier and Jeffares 2003; Russell et al. 2005; Franzén et al. 2013). This observation is consistent irrespective of the roles of intron-containing genes in translation, supporting the idea that introns may have regulatory roles in both transcription and translation (Lim et al. 2018).

Roles of introns in gene expression

Previous studies have shown that introns are common in the ribosomal protein genes (e.g. *RPL7B*) of intron-poor protozoa and saccharomycetous yeasts (Bon et al. 2003; Russell et al. 2005; Franzén et al. 2013). However, the abundance of introns in other classes of genes is less well-known. We examined the GO terms of the orthologs of the intron-poor species. We found that introns are highly abundant not only in genes involved in cytoplasmic translation (e.g. ribosomal proteins) but also in genes involved in Golgi vesicle transport (Figure 7). In contrast, introns are depleted in genes involved in base-excision repair and peptidyl-proline modification. The reasons for these biases are still unclear.

These findings prompted us to compare the transcription level and translation efficiency between intron-containing and intronless genes. We analyzed the matched RNA-seq and ribosome profiling datasets for the fungal species that are publicly available — *S. cerevisiae* (Heyer and Moore 2016), *Candida albicans* (Muzzey et al. 2014), *S. pombe* (Subtelny et al. 2014), and *Neurospora crassa* (Yu et al. 2015) (Supplementary Table S3 and Materials and Methods).

Interestingly, intron-containing genes tend to have higher mRNA expression and translation efficiency than that of intronless genes (Figure 9). Overall, our results provide independent evidence of diverse roles of fungal introns in transcription and translation.

DISCUSSION

Widespread intron loss in the fungal kingdom

This study has shown that intron loss was widespread in the fungal kingdom during evolution (Figure 1). The most extreme cases are microsporidian parasites, which have lost all, or nearly all introns. Microsporidian parasites have the smallest eukaryotic genomes and coding capacities known to date (Corradi 2015; Han and Weiss 2017). Intriguingly, Chytridiomycota and Mucaromycota, two other early-diverging phyla, are instead characterized by the retention of ancestral introns and maintain relatively high intron densities. *Gonapodya prolifera*, a chytrid fungus, has the highest intron density of all the fungi in our analysis (4.8 introns/kbp), 73% of the intron density of humans *Homo sapiens*.

As previously observed, ascomycetes have lost more introns than basidiomycetes, which is in agreement with a study of 99 eukaryotic species (Csuros et al. 2011). Many ascomycetes are unicellular fungi or yeasts, known to have low numbers of introns (in particular, the subphylum Saccharomycotina) (Byrne and Wolfe 2005; Neuvéglise et al. 2011; Hooks et al. 2014). Our analysis has further shown substantial intron loss in Pezizomycotina, notably the *Cyphellophora* and *Sporothrix* sp. (below 10% of the intron densities of the last fungal ancestor). These species are conidia producing fungi that have a yeast or yeast-like stage as part of their life cycle (Barros et al. 2011; Feng et al. 2012).

In contrast, many basidiomycetes are fruit-body producing fungi that are known to have relatively higher numbers of introns. However, prominent intron loss has also occurred in Ustilaginomycotina (Kämper et al. 2006; Stajich et al. 2007). Many of these smuts and allies have evolved into plant pathogens and evolved a yeast state as part of their life cycle (Rush and Aime 2013; Wang et al. 2015; Benevenuto et al. 2018; Kijpornyongpan et al. 2018).

However, not all yeasts or yeast-like fungi have low intron densities. For example, *Pneumocystis murina* and *Cryptococcus neoformans*, which are ascomycetous and basidiomycetous yeasts, have high intron densities (3.7 and 3.6 introns/kbp).

246 **‘Concerted evolution’ of introns and their host genes**

247 What is the function of introns? The role of most introns is unclear as they are mostly dispensable
248 (Niu 2008). To address this we chose eight ascomycetes and basidiomycetes with extensive intron loss
249 for in-depth analysis. These intron-poor species all have a yeast or yeast-like stage in their life cycle.
250 Our evolutionary and statistical approaches have shown that remaining introns are unlikely to be
251 conserved by chance (Figure 4, 5 and 6).

252 Several studies have shown that the 5' splice sites of intron-poor species are more conserved
253 than that of intron-rich species (Irimia et al. 2007; Skelly et al. 2009; Neuvéglise et al. 2011). In
254 addition, previous studies have shown that deleting most introns in *S. cerevisiae* does not significantly
255 compromise growth but starvation resistance (Parenteau et al. 2008; Parenteau et al. 2011; Parenteau
256 et al. 2019). These support our idea that introns are retained because of their useful functions.

257 Interestingly, *S. cerevisiae* and *U. maydis* have significantly higher proportions of duplicated
258 genes with introns (Figure 4). Functional divergence might have occurred in one of the paralogs (and
259 their introns) as shown in *S. cerevisiae* (Kellis et al. 2004; Pleiss et al. 2007; Parenteau et al. 2011;
260 Petibon et al. 2016; Parenteau and Abou Elela 2019). It would be interesting to see whether the
261 introns of duplicated genes in *U. maydis* have similar roles as that of *S. cerevisiae*.

262

263 **Regulatory roles of introns in transcription and translation**

264 Notably, most of the first introns are located near the translation initiation codons (Figure 7). Indeed,
265 intron loss near the 3' end of a gene was previously found to be prevalent in some protozoa and fungi,
266 probably due to reverse transcriptase-mediated intron loss (Fink 1987; Roy and Gilbert 2005; Russell
267 et al. 2005; Lee et al. 2010; Zhang et al. 2010; Franzén et al. 2013; Koonin et al. 2013; Zhu and Niu
268 2013a; Zhu and Niu 2013b; Irimia and Roy 2014). This location bias of introns has also been found in
269 intron-rich metazoa and plants (Lim et al. 2018).

270 Introns are also more abundant in ancient genes, in particular, ribosomal protein genes (Figure
271 8). This is in agreement with a previous study on seven saccharomycetous yeasts (Bon et al. 2003). In
272 addition, introns are more abundant in genes that have higher mRNA expression and translation
273 efficiency, irrespective of their cellular functions (Figure 9). This extends previous analyses of global
274 gene expression of *S. cerevisiae* (Juneau et al. 2006; Hoshida et al. 2017). In metazoa and plants,
275 introns may enhance transcription or translation, in part, through EJC (Wiegand et al. 2003; Diem et
276 al. 2007; Chazal et al. 2013; Le Hir et al. 2016). EJCs deposit at about 20-24 bases upstream of the
277 exon-exon junctions upon splicing, carrying over the ‘memory’ of splicing events to cytoplasmic
278 translation. However, *S. cerevisiae* has no EJCs, unlike complex eukaryotes or even the fission yeast
279 *S. pombe*. It remains unclear how intron enhances transcription and translation in *Saccharomycetes*
280 (Moabbi et al. 2012; Hoshida et al. 2017).

We propose that highly conserved intron positions are indicative of functional importance, e.g. the ncRNA gene *snR191* embedded in the intron of *NOG2* gene (Figure 6). This intron was previously found to be highly conserved in the family *Saccharomycetaceae* (Hooks et al. 2014; Hooks et al. 2016). Some other introns may harbor functional structured RNA elements, such as the introns of *RPL18A* and *RPS22B* pre-mRNAs that promote RNase III-mediated degradation, and the *GLC7* intron that modulates gene expression during salt stress (Danin-Kreiselman et al. 2003; Juneau et al. 2006; Parenteau et al. 2008; Hooks et al. 2016).

Concluding remarks

By encompassing an unprecedented number of species, from a single group of eukaryotes with a range of very different evolutionary histories, these results allow us to better understand commonalities of intron evolution. We find a remarkable trend towards intron number reduction across lineages, as well as highly predictable patterns of intron retention in intron-poor species at the level of gene function, specific gene, specific intron, and genic position. These results provide explanations for long-observed patterns, while revealing previously unknown patterns to be explained by future studies.

MATERIALS AND METHODS

Genome sequences and annotations

We retrieved 633 fungal genomes (FASTA and GTF files) from the Ensembl Fungi release 34 (Zerbino et al. 2018). In addition, the *L. starkeyi* and *Neolecta irregularis* genomes were retrieved from Ensembl Fungi 42 and NCBI Genome, respectively, whereas seven Ustilaginomycotina and two *Taphrinomycotina* genomes from JGI MycoCosm (Cissé et al. 2013; Grigoriev et al. 2014; Riley et al. 2016; Mondo et al. 2017; Nguyen et al. 2017; Kijpornyongpan et al. 2018). Detailed information can be found in Supplementary Table S1.

Redundant species were filtered by assembly level (ftp://ftp.ncbi.nlm.nih.gov/genomes/ASSEMBLY_REPORTS/assembly_summary_genbank.txt) (Kitts et al. 2016). Complete genomes were retained, otherwise the assemblies at the chromosome, scaffold, or contig levels. For redundant assemblies, only the assemblies with the highest numbers of CDS were retained. For outgroups, the genomes of *Homo sapiens* and the cellular slime mold *Fonticula alba* were downloaded from Ensembl 95 and Ensembl Protists 42, respectively.

The annotation of the UTR and UTR introns of *S. cerevisiae* was retrieved from YeastMine (Balakrishnan et al. 2012). The GO terms of *S. cerevisiae* were retrieved from the *Saccharomyces* Genome Database (Cherry et al. 2012).

316

317 **Taxonomic and phylogenetic trees**

318 We chose a subset of 263 fungi for the inference of ancestral introns then pruned an 1100 taxa tree
319 from concatenated analyses (J. Stajich, personal communication, December 24, 2018). *Homo sapiens*
320 and *Fonticula alba* were included as outgroups. For visualization, the tips and nodes were color-coded
321 by inferred intron densities using the R package ggtree v1.16.6 (Yu et al. 2017).

322 For phylogenetic independent contrasts analysis, we retrieved a phylogenetic tree from the
323 SILVA database release-111 (Yarza et al. 2017). The tip labels were replaced using AfterPhylo.pl and
324 the tree was pruned using filter_tree.py (Caporaso et al. 2010; Zhu 2014).

325

326 **Orthology analysis**

327 For the inference of ancestral introns, orthologous genes were identified using HMMER v3.1b2
328 (Johnson et al. 2010). Profile hidden markov models (HMMs) were retrieved from the 1000 Fungal
329 Genomes Project (1KFG) and fuNOG (eggNOG v4.5) (Huerta-Cepas, Szklarczyk, et al. 2016;
330 Bewick et al. 2019). A HMM database was built using hmmpress. Homology sequences were detected
331 using hmmsearch. For species that have multiple hits per HMM, only the top hit was retained. To
332 remove false positives, hits with bit scores over 276.48 were retained. This threshold was estimated
333 from the distribution of bit scores (bimodal lognormal) using the R package cutoff v0.1.0 (Choisy
334 2015). Only the orthologs that captured over 80% (212/265) of the species were used in the
335 subsequent analyses (1445 orthologs).

336 Eight intron-poor species were selected for analysis of intron functions, including *S.*
337 *cerevisiae* and *C. dubliniensis* in *Saccharomycotina*, *C. europaea* and *S. schenckii* in *Pezizomycotina*,
338 and *U. maydis*, *P. hubeiensis*, *M. miltonrushii* and *M. sympodialis* in *Ustilaginomycotina*. The
339 orthologs of these intron-poor species were identified using proteinortho5 (using parameter -synteny)
340 (Lechner et al. 2011). A total of 1302 orthologs were identified. In contrast to the above approach, this
341 approach is less scalable but unrestricted by a predefined set of orthologs (HMMs).

342 Duplicated genes were identified using SkewGD v1 (Tian 2018). This pipeline includes
343 sequence clustering and ‘age’ estimation using K_s (the number of synonymous substitutions per
344 synonymous site) (Blanc and Wolfe 2004; Vanneste et al. 2013).

345

346 **Intron alignment**

347 For the inference of ancestral introns, protein sequences were aligned using Clustal Omega v1.2.4
348 (using parameter --hmm-in) (Sievers and Higgins 2018). Annotations of intron positions were
349 extracted from GTF/GFF files using ReSplicer (by calling the splice.extractAnnotations class) (Sêton
350 Bocco and Csürös 2016). The alignments were realigned using IntronAlignment (Csurös et al. 2007).

The orthologs of the intron-poor species were aligned using MUSCLE v3.8.31 (Edgar 2004). The protein sequences were realigned using ReSplicer and IntronAlignment as above. Intron positions were then re-annotated using ReSplicer, by calling a series of java classes splice.extractAnnotations, splice.collectStatistics, and splice.checkSites. Realignment was repeated using re-annotated intron positions.

Inference of ancestral introns

We inferred ancestral introns from 1445 orthologs of 263 fungal genomes using Malin (Csurös 2008). Firstly, we generated a table of intron presence/absence in the orthologs using Malin. It included 46,381 intron sites allowing a maximum of 53 ambiguous characters per site.

Failure to account for variation in intron loss rate across sites can lead to an underestimation in intron density of eukaryotic ancestors (Stajich et al. 2007), and previous experiments with rate variation models across sites in Malin showed that model fit was significantly impacted solely by variation in loss rate across intron sites (Csuros et al. 2011). Here, intron gain and loss rates were optimized in Malin using maximum likelihood using the constant rate and rate-variation models starting from the standard null model and running 1000 optimization rounds (likelihood convergence threshold = 0.001). For the constant rate model, each intron site has only a branch-specific gain and loss rate. In contrast, for the rate-variation model, intron sites additionally belong to one of two discrete rate loss categories.

Malin calculates gain/loss rates and intron density at the root by numerical optimization of the likelihood. For both the constant rate and rate-variation models, we used 100 bootstrap replicates of the intron table to assess uncertainty about inferred rate parameters and intron site history for every node. For model comparison, the likelihood-ratio test statistic calculated as $\Delta = -2 \times (L1 - L2)$, where $L1$ is the log-likelihood of the constant rate model ($L1 = -402882$) and $L2$ is the log-likelihood of the rate-variation model ($L2 = -397548$). The likelihood-ratio test statistic is 10,668, which was then compared to a χ^2 distribution with one degree of freedom. In this comparison, we obtained a p -value of 0.0. Therefore, we rejected the constant rate results and chose the more complex rate-variation model. Finally, we inferred ancestral densities by using Dollo parsimony (Farris 1977).

For all analyses, we scaled the number of inferred introns to intron density by multiplying by 0.37 and dividing by 322, where 0.37 and 322 are intron density and the number of introns in *Schizosaccharomyces pombe* in the orthologous dataset, respectively. *S. pombe* was used as a reference because it has a high-quality annotation and over an order of magnitude higher intron density than *S. cerevisiae* (Csuros et al. 2011; Lock et al. 2018).

386 **Phylogenetic independent contrasts analysis**

387 Three features (intron density, genome size, and the number of CDS) were first examined for
388 normality using different transformation functions in the R package bestNormalize v1.4.3 (Peterson
389 2018). Ordered quantile transformation was chosen. Phylogenetic independent contrasts analysis was
390 carried out using the R package caper v1.0.1 (Orme et al. 2018).

391

392 **Branch-site test**

393 The orthologous protein sequences were aligned using PRANK v.150803 (Löytynoja and Goldman
394 2008; Jeffares et al. 2015). The aligned protein sequences were converted to aligned DNA sequences
395 using PAL2NAL (Suyama et al. 2006). These aligned DNA sequences were used to build
396 phylogenetic trees using RaxML v8.2.9 (using parameters -f a -x 1181 -N 1000 -m GTRGAMMA)
397 (Stamatakis 2014). To identify positively selected genes, branch-site tests were performed using both
398 the aligned DNA sequences and phylogenetic trees using ETE toolkit v3.1.1 (ete-evol, a CodeML
399 wrapper) (Yang 2007; Huerta-Cepas, Serra, et al. 2016). The positive selection (bsA, alternative
400 hypothesis) and relaxation (bsA1, null hypothesis) evolutionary models were fit to the orthologous
401 dataset. This involved modeling each branch by recursively marking the remaining branches as the
402 foreground branches, and comparing them using likelihood-ratio tests (using parameters --models M0
403 bsA bsA1 --leaves --tests bsA,bsA1).

404

405 **snoRNA prediction**

406 The Stockholm alignment files of fungal snoRNA families were downloaded from
407 <http://www.bioinf.uni-leipzig.de/publications/supplements/17-001> (Canzler et al. 2018). These files
408 were used to build HMMs or covariance models using Infernal v1.1.2 (Nawrocki and Eddy 2013).
409 These models were used to detect the snoRNA genes encoded by introns. The functional elements in
410 the snoRNAs were predicted using snoscan v0.2b and the snoGPS web server (Lowe and Eddy 1999;
411 Schattner et al. 2005).

412

413 **Gene ontology analysis**

414 Functional annotation of *S. cerevisiae* genes was performed using the Bioconductor packages
415 clusterProfiler v3.0.5 and org.Sc.sgd.db v3.4.0 (Yu et al. 2012; Huber et al. 2015; Carlson 2017).
416 Redundant GO terms were removed using the simplify function with default settings. Genes were
417 grouped by GO terms and normalized/plotted using the R package massageR v0.7.2 (Stanstrup 2017).

418

419 **RNA-seq and ribosome profiling data analyses**

List of RNA-seq and ribosome profiling datasets used are available in Supplementary Table S3. The genome and annotation files of *Candida albicans* and *Schizosaccharomyces pombe* were downloaded from the *Candida* Genome Database assembly 22 and PomBase release 30, respectively (Skrzypek et al. 2017; Lock et al. 2018).

Reads were aligned to ncRNAs using STAR v2.5.2b (Dobin et al. 2013). Unmapped reads were then aligned to the genome with transcript annotation. Uniquely mapped reads were counted using featureCounts v1.5.0-p3 (Liao et al. 2014).

For RNA-seq, count data were normalized to Reads Per Kilobase per Million (RPKM) mapped reads. $RPKM = \text{read_counts}/(\text{gene_length}/1000)/(\text{total_read_counts}/10^6)$. For ribosome profiling, count data were used to calculate translation efficiency. Translation efficiency = $(\text{ribosome_footprints}/\text{total_footprint_counts})/(\text{RNA-seq_read_counts}/\text{total RNA-seq_read_counts})$.

We detected the *S. cerevisiae* orthologs in other species using proteinortho5 (using parameter -synteny) (Lechner et al. 2011). We found 3063, 2506, and 2541 *S. cerevisiae* orthologs in *C. albicans*, *S. pombe*, and *N. crassa*, respectively. The orthologs were grouped by introns presence/absence and GO terms. These data were then mapped to the RNA-seq and ribosome profiling results by gene.

Statistical analysis

Statistical analysis and plotting were performed using R ≥ 3.4 (R Core Team 2018). Fisher's exact test, the chi-square test, Welch two sample t-test and Spearman's rank correlation were calculated using the base R system. Computation of binomial confidence intervals using Bayesian inference was performed using the binom package (Dorai-Raj 2014). All *p*-values obtained from multiple testing were adjusted using the Bonferroni correction to avoid false positives (Armstrong 2014), unless otherwise mentioned. Plots were constructed using ggplot2 v2.2.1 (Wickham 2016), unless otherwise stated.

Data availability

Code and data for this study are available at https://github.com/Brookesloci/fungi_intron_paper_2020/.

ACKNOWLEDGMENTS

CMB and CSL were funded by the University of Otago. CSL was a recipient of a Dr. Sulaiman Daud 125th Jubilee Postgraduate Scholarship. SWR and BNW were supported by the National Science Foundation award #1616878.

455 REFERENCES

- 456 Andrews S. 2016. FastQC: a quality control tool for high throughput sequence data. Available from:
457 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- 458 Armstrong RA. 2014. When to use the Bonferroni correction. *Ophthalmic Physiol. Opt.* 34:502–508.
- 459 Balakrishnan R, Park J, Karra K, Hitz BC, Binkley G, Hong EL, Sullivan J, Micklem G, Cherry JM.
460 2012. YeastMine--an integrated data warehouse for *Saccharomyces cerevisiae* data as a
461 multipurpose tool-kit. *Database* 2012:bar062.
- 462 Barros MB d. L, M B d, de Almeida Paes R, Schubach AO. 2011. *Sporothrix schenckii* and
463 Sporotrichosis. *Clin. Microbiol. Rev.* 24:633–654.
- 464 Benevenuto J, Teixeira-Silva NS, Kuramae EE, Croll D, Monteiro-Vitorello CB. 2018. Comparative
465 genomics of smut pathogens: Insights from orphans and positively selected genes into host
466 specialization. *Front. Microbiol.* 9:660.
- 467 Bewick AJ, Hofmeister BT, Powers RA, Mondo SJ, Grigoriev IV, James TY, Stajich JE, Schmitz RJ.
468 2019. Diversity of cytosine methylation across the fungal tree of life. *Nat Ecol Evol* 3:479–490.
- 469 Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age
470 distributions of duplicate genes. *Plant Cell* 16:1667–1678.
- 471 Bon E, Casaregola S, Blandin G, Llorente B, Neuvéglise C, Munsterkötter M, Guldener U, Mewes
472 H-W, Van Helden J, Dujon B, et al. 2003. Molecular evolution of eukaryotic genomes:
473 hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res.* 31:1121–1135.
- 474 Bonnet A, Grosso AR, Elkaoutari A, Coleno E, Presle A, Sridhara SC, Janbon G, Géli V, de Almeida
475 SF, Palancade B. 2017. Introns protect eukaryotic genomes from transcription-associated genetic
476 instability. *Mol. Cell* 67:608–621.e6.
- 477 Byrne KP, Wolfe KH. 2005. The Yeast Gene Order Browser: combining curated homology and
478 syntenic context reveals gene fate in polyploid species. *Genome Res.* 15:1456–1461.
- 479 Canzler S, Stadler PF, Schor J. 2018. The fungal snoRNAome. *RNA* 24:342–360.
- 480 Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG,
481 Goodrich JK, Gordon JI, et al. 2010. QIIME allows analysis of high-throughput community
482 sequencing data. *Nat. Methods* 7:335–336.
- 483 Carlson M. 2017. org.Sc.sgd.db: Genome wide annotation for Yeast. Available from:
484 <http://bioconductor.org/packages/org.Sc.sgd.db/>
- 485 Chan LG, Cohen JL, de Moura Bell JMLN. 2018. Conversion of agricultural streams and
486 food-processing by-products to value-added compounds using filamentous fungi. *Annu. Rev.*
487 *Food Sci. Technol.* 9:503–523.
- 488 Chazal P-E, Daguenet E, Wendling C, Ulryck N, Tomasetto C, Sargueil B, Le Hir H. 2013. EJC core
489 component MLN51 interacts with eIF3 and activates translation. *Proc. Natl. Acad. Sci. U. S. A.*

110:5903–5908.

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, et al. 2012. *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* 40:D700–D705.

Choisy M. 2015. cutoff: Identify a cutoff value from bimodal data. Available from: <https://github.com/choisy/cutoff>

Chorev M, Carmel L. 2012. The function of introns. *Front. Genet.* 3:55.

Cissé OH, Almeida JMGCF, Fonseca A, Kumar AA, Salojärvi J, Overmyer K, Hauser PM, Pagni M. 2013. Genome sequencing of the plant pathogen *Taphrina deformans*, the causal agent of peach leaf curl. *MBio* 4:e00055–13.

Comeron JM, Kreitman M. 2000. The correlation between intron length and recombination in *Drosophila*. Dynamic equilibrium between mutational and selective forces. *Genetics* 156:1175–1190.

Corradi N. 2015. Microsporidia: Eukaryotic intracellular parasites shaped by gene loss and horizontal gene transfers. *Annu. Rev. Microbiol.* 69:167–183.

Csurös M. 2008. Malin: maximum likelihood analysis of intron evolution in eukaryotes. *Bioinformatics* 24:1538–1539.

Csurös M, Holey JA, Rogozin IB. 2007. In search of lost introns. *Bioinformatics* 23:i87–i96.

Csuros M, Rogozin IB, Koonin EV. 2011. A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS Comput. Biol.* 7:e1002150.

Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, Becnel JJ, Didier ES, Fan L, Heiman DI, Levin JZ, et al. 2012. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Res.* 22:2478–2488.

Danin-Kreiselman M, Lee CY, Chanfreau G. 2003. RNase III-mediated degradation of unspliced pre-mRNAs and lariat introns. *Mol. Cell* 11:1279–1289.

De Conti L, Baralle M, Buratti E. 2013. Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* 4:49–60.

Desjardins CA, Sanscrainte ND, Goldberg JM, Heiman D, Young S, Zeng Q, Madhani HD, Becnel JJ, Cuomo CA. 2015. Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. *Nat. Commun.* 6:7121.

Diem MD, Chan CC, Younis I, Dreyfuss G. 2007. PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat. Struct. Mol. Biol.* 14:1173–1179.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

Dorai-Raj S. 2014. binom: Binomial confidence intervals for several parameterizations. Available

525 from: <https://cran.r-project.org/web/packages/binom/index.html>

526 Dornburg A, Townsend JP, Wang Z. 2017. Maximizing power in phylogenetics and phylogenomics: A
527 perspective illuminated by fungal big data. *Adv. Genet.* 100:1–47.

528 Duret L. 2001. Why do genes have introns? Recombination might add a new piece to the puzzle.
529 *Trends Genet.* 17:172–175.

530 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
531 *Nucleic Acids Res.* 32:1792–1797.

532 Farris JS. 1977. Phylogenetic Analysis Under Dollo’s Law. *Syst. Biol.* 26:77–88.

533 Feng P, Lu Q, Najafzadeh MJ, van den Ende AHGG, Sun J, Li R, Xi L, Vicente VA, Lai W, Lu C, et
534 al. 2012. *Cyphellophora* and its relatives in *Phialophora*: biodiversity and possible role in human
535 infection. *Fungal Divers.* 65:17–45.

536 Fink GR. 1987. Pseudogenes in yeast? *Cell* 49:5–6.

537 Franzén O, Jerlström-Hultqvist J, Einarsson E, Ankarklev J, Ferella M, Andersson B, Svärd SG. 2013.
538 Transcriptome profiling of *Giardia intestinalis* using strand-specific RNA-seq. *PLoS Comput.*
539 *Biol.* 9:e1003000.

540 Gallegos JE, Rose AB. 2015. The enduring mystery of intron-mediated enhancement. *Plant Sci.*
541 237:8–15.

542 Gilbert W. 1978. Why genes in pieces? *Nature* 271:501–501.

543 Grau-Bové X, Torruella G, Donachie S, Suga H, Leonard G, Richards TA, Ruiz-Trillo I. 2017.
544 Dynamics of genomic innovation in the unicellular ancestry of animals. *Elife* [Internet] 6.
545 Available from: <http://dx.doi.org/10.7554/eLife.26036>

546 Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, Riley R, Salamov A, Zhao X,
547 Korzeniewski F, et al. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic*
548 *Acids Res.* 42:D699–D704.

549 Grützmann K, Szafranski K, Pohl M, Voigt K, Petzold A, Schuster S. 2014. Fungal alternative
550 splicing is associated with multicellular complexity and virulence: a genome-wide multi-species
551 study. *DNA Res.* 21:27–39.

552 Han B, Weiss LM. 2017. Microsporidia: Obligate intracellular pathogens within the fungal kingdom.
553 *Microbiol Spectr* 5:FUNK – 0018–2016.

554 Hellens RP, Brown CM, Chisnall MAW, Waterhouse PM, Macknight RC. 2016. The emerging world
555 of small ORFs. *Trends Plant Sci.* 21:317–328.

556 Heyer EE, Moore MJ. 2016. Redefining the translational status of 80S monosomes. *Cell*
557 164:757–769.

558 Hoepfner MP, Poole AM. 2012. Comparative genomics of eukaryotic small nucleolar RNAs reveals
559 deep evolutionary ancestry amidst ongoing intragenomic mobility. *BMC Evol. Biol.* 12:183.

560 Hooks KB, Delneri D, Griffiths-Jones S. 2014. Intron evolution in *Saccharomycetaceae*. *Genome*
561 *Biol. Evol.* 6:2543–2556.

562 Hooks KB, Naseeb S, Parker S, Griffiths-Jones S, Delneri D. 2016. Novel intronic RNA structures
563 contribute to maintenance of phenotype in *Saccharomyces cerevisiae*. *Genetics* 203:1469–1481.

564 Hoshida H, Kondo M, Kobayashi T, Yarimizu T, Akada R. 2017. 5'-UTR introns enhance protein
565 expression in the yeast *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 101:241–251.

566 Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L,
567 Girke T, et al. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat.*
568 *Methods* 12:115–121.

569 Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, analysis, and visualization of
570 phylogenomic data. *Mol. Biol. Evol.* 33:1635–1638.

571 Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR,
572 Sunagawa S, Kuhn M, et al. 2016. eggNOG 4.5: a hierarchical orthology framework with
573 improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids*
574 *Res.* 44:D286–D293.

575 Huff JT, Zilberman D, Roy SW. 2016. Mechanism for DNA transposons to generate introns on
576 genomic scales. *Nature* 538:533–536.

577 Irimia M, Penny D, Roy SW. 2007. Coevolution of genomic intron number and splice sites. *Trends*
578 *Genet.* 23:321–325.

579 Irimia M, Roy SW. 2014. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb.*
580 *Perspect. Biol.* 6:a016071.

581 Jeffares DC, Tomiczek B, Sojo V, dos Reis M. 2015. A beginners guide to estimating the
582 non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. *Methods*
583 *Mol. Biol.* 1201:65–90.

584 Jo B-S, Choi SS. 2015. Introns: The functional benefits of introns in genomes. *Genomics Inform.*
585 13:112–118.

586 Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM
587 search procedure. *BMC Bioinformatics* 11:431.

588 Juneau K, Miranda M, Hillenmeyer ME, Nislow C, Davis RW. 2006. Introns regulate RNA and
589 protein abundance in yeast. *Genetics* 174:511–518.

590 Kämper J, Kahmann R, Bölker M, Ma L-J, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE,
591 Müller O, et al. 2006. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago*
592 *maydis*. *Nature* 444:97–101.

593 Keeling PJ, Corradi N, Morrison HG, Haag KL, Ebert D, Weiss LM, Akiyoshi DE, Tzipori S. 2010.
594 The reduced genome of the parasitic microsporidian *Enterocytozoon bieneusi* lacks genes for

core carbon metabolism. *Genome Biol. Evol.* 2:304–309.

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

Kempken F. 2013. Alternative splicing in ascomycetes. *Appl. Microbiol. Biotechnol.* 97:4235–4241.

Kijpornyongpan T, Mondo SJ, Barry K, Sandor L, Lee J, Lipzen A, Pangilinan J, LaButti K, Hainaut M, Henrissat B, et al. 2018. Broad genomic sampling reveals a smut pathogenic ancestry of the fungal clade *Ustilaginomycotina*. *Mol. Biol. Evol.* 35:1840–1854.

Kitts PA, Church DM, Thibaud-Nissen F, Choi J, Hem V, Sapojnikov V, Smith RG, Tatusova T, Xiang C, Zherikov A, et al. 2016. Assembly: a resource for assembled genomes at NCBI. *Nucleic Acids Res.* 44:D73–D80.

Koonin EV, Csuros M, Rogozin IB. 2013. Whence genes in pieces: reconstruction of the exon-intron gene structures of the last eukaryotic common ancestor and other ancestral eukaryotes. *Wiley Interdiscip. Rev. RNA* 4:93–105.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34:1812–1819.

Laxa M. 2016. Intron-mediated enhancement: A tool for heterologous gene expression in plants? *Front. Plant Sci.* 7:1977.

Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12:124.

Lee RCH, Gill EE, Roy SW, Fast NM. 2010. Constrained intron structures in a microsporidian. *Mol. Biol. Evol.* 27:1979–1982.

Le Hir H, Saulière J, Wang Z. 2016. The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.* 17:41–54.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930.

Lim CS, T. Wardell SJ, Kleffmann T, Brown CM. 2018. The exon–intron gene structure upstream of the initiation codon predicts translation efficiency. *Nucleic Acids Res.* 46:4575–4591.

Lock A, Rutherford K, Harris MA, Wood V. 2018. PomBase: The scientific resource for fission yeast. *Methods Mol. Biol.* 1757:49–68.

Logsdon JM Jr, Tyshenko MG, Dixon C, D-Jafari J, Walker VK, Palmer JD. 1995. Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. *Proc. Natl. Acad. Sci. U. S. A.* 92:8507–8511.

Loh Y-H, Brenner S, Venkatesh B. 2008. Investigation of loss and gain of introns in the compact genomes of pufferfishes (*Fugu* and *Tetraodon*). *Mol. Biol. Evol.* 25:526–535.

Lowe TM, Eddy SR. 1999. A computational screen for methylation guide snoRNAs in yeast. *Science*

630 283:1168–1171.

631 Löytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence
632 alignment and evolutionary analysis. *Science* 320:1632–1635.

633 Luo Y, Li S. 2007. Genome-wide analyses of retrogenes derived from the human box H/ACA
634 snoRNAs. *Nucleic Acids Res.* 35:559–571.

635 Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302:1401–1404.

636 Mekouar M, Blanc-Lenfle I, Ozanne C, Da Silva C, Cruaud C, Wincker P, Gaillardin C, Neuvéglise
637 C. 2010. Detection and analysis of alternative splicing in *Yarrowia lipolytica* reveal structural
638 constraints facilitating nonsense-mediated decay of intron-retaining transcripts. *Genome Biol.*
639 11:R65.

640 Mikhailov KV, Simdyanov TG, Aleoshin VV. 2017. Genomic survey of a hyperparasitic
641 microsporidian *Amphiamblys* sp. (Metchnikovellidae). *Genome Biol. Evol.* 9:454–467.

642 Moabbi AM, Agarwal N, El Kaderi B, Ansari A. 2012. Role for gene looping in intron-mediated
643 enhancement of transcription. *Proc. Natl. Acad. Sci. U. S. A.* 109:8505–8510.

644 Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K, Haridas S, Kuo A, Salamov
645 A, Ahrendt SR, et al. 2017. Widespread adenine N6-methylation of active genes in fungi. *Nat.*
646 *Genet.* 49:964–968.

647 Mourier T, Jeffares DC. 2003. Eukaryotic intron loss. *Science* 300:1393.

648 Muzzey D, Sherlock G, Weissman JS. 2014. Extensive and coordinated control of allele-specific
649 expression by both transcription and translation in *Candida albicans*. *Genome Res.* 24:963–973.

650 Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*
651 29:2933–2935.

652 Ndikumana S, Pelin A, Williot A, Sanders JL, Kent M, Corradi N. 2017. Genome analysis of
653 *Pseudoloma neurophila*: A microsporidian parasite of zebrafish (*Danio Rerio*). *J. Eukaryot.*
654 *Microbiol.* 64:18–30.

655 Neuvéglise C, Marck C, Gaillardin C. 2011. The intronome of budding yeasts. *C. R. Biol.*
656 334:662–670.

657 Nguyen TA, Cissé OH, Yun Wong J, Zheng P, Hewitt D, Nowrousian M, Stajich JE, Jedd G. 2017.
658 Innovation and constraint leading to complex multicellularity in the Ascomycota. *Nat. Commun.*
659 8:14444.

660 Niu D-K. 2007. Protecting exons from deleterious R-loops: a potential advantage of having introns.
661 *Biol. Direct* 2:11.

662 Niu D-K. 2008. Exon definition as a potential negative force against intron losses in evolution. *Biol.*
663 *Direct* 3:46.

664 Niu D-K, Yang Y-F. 2011. Why eukaryotic cells use introns to enhance gene expression: Splicing

665 reduces transcription-associated mutagenesis by inhibiting topoisomerase I cutting activity. *Biol.*
666 *Direct* 6:24.

667 Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W. 2018. caper: Comparative
668 Analyses of Phylogenetics and Evolution in R. Available from:
669 <https://cran.r-project.org/web/packages/caper/index.html>

670 Parenteau J, Abou Elela S. 2019. Introns: Good Day Junk Is Bad Day Treasure. *Trends Genet.*
671 35:923–934.

672 Parenteau J, Durand M, Morin G, Gagnon J, Lucier J-F, Wellinger RJ, Chabot B, Elela SA. 2011.
673 Introns within ribosomal protein genes regulate the production and function of yeast ribosomes.
674 *Cell* 147:320–331.

675 Parenteau J, Durand M, Véronneau S, Lacombe A-A, Morin G, Guérin V, Cecez B, Gervais-Bird J,
676 Koh C-S, Brunelle D, et al. 2008. Deletion of many yeast introns reveals a minority of genes that
677 require splicing for function. *Mol. Biol. Cell* 19:1932–1941.

678 Parenteau J, Maignon L, Berthoumieux M, Catala M, Gagnon V, Abou Elela S. 2019. Introns are
679 mediators of cell response to starvation. *Nature* 565:612–617.

680 Patthy L. 2003. Modular assembly of genes and the evolution of new functions. *Genetica*
681 118:217–231.

682 Peterson RA. 2018. Estimating normalization transformations with bestNormalize. Available from:
683 <https://github.com/petersonR/bestNormalize>

684 Petibon C, Parenteau J, Catala M, Elela SA. 2016. Introns regulate the production of ribosomal
685 proteins by modulating splicing of duplicated ribosomal protein genes. *Nucleic Acids Res.*
686 44:3878–3891.

687 Peyretailade E, Parisot N, Polonais V, Terrat S, Denonfoux J, Dugat-Bony E, Wawrzyniak I,
688 Biderre-Petit C, Mahul A, Rimour S, et al. 2012. Annotation of microsporidian genomes using
689 transcriptional signals. *Nat. Commun.* 3:1137.

690 Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C. 2007. Rapid, transcript-specific changes in
691 splicing in response to environmental stress. *Mol. Cell* 27:928–937.

692 Priest SJ, Yadav V, Heitman J. 2020. Advances in understanding the evolution of fungal genome
693 architecture. *F1000Res.* 9:776.

694 R Core Team. 2018. R: A language and environment for statistical computing. Vienna, Austria: R
695 Foundation for Statistical Computing Available from: <https://www.R-project.org/>

696 Riley R, Haridas S, Wolfe KH, Lopes MR, Hittinger CT, Göker M, Salamov AA, Wisecaver JH, Long
697 TM, Calvey CH, et al. 2016. Comparative genomics of biotechnologically important yeasts.
698 *Proc. Natl. Acad. Sci. U. S. A.* 113:9882–9887.

699 Rose AB. 2018. Introns as Gene Regulators: A Brick on the Accelerator. *Front. Genet.* 9:672.

700 Roy SW. 2004. The origin of recent introns: transposons? *Genome Biol.* 5:251.

701 Roy SW, Gilbert W. 2005. The pattern of intron loss. *Proc. Natl. Acad. Sci. U. S. A.* 102:713–718.

702 Rush TA, Aime MC. 2013. The genus *Meira*: phylogenetic placement and description of a new
703 species. *Antonie Van Leeuwenhoek* 103:1097–1106.

704 Russell AG, Shutt TE, Watkins RF, Gray MW. 2005. An ancient spliceosomal intron in the ribosomal
705 protein L7a gene (Rpl7a) of *Giardia lamblia*. *BMC Evol. Biol.* 5:45.

706 Sapountzis P, Gruntjes T, Otani S, Estevez J, da Costa RR, Plunkett G 3rd, Perna NT, Poulsen M.
707 2015. The enterobacterium *Trabulsiella odontotermitis* presents novel adaptations related to its
708 association with fungus-growing termites. *Appl. Environ. Microbiol.* 81:6577–6588.

709 Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for
710 the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33:W686–W689.

711 Schmitz J, Zemmann A, Churakov G, Kuhl H, Grützner F, Reinhardt R, Brosius J. 2008. Retroposed
712 SNOfall--a mammalian-wide comparison of platypus snoRNAs. *Genome Res.* 18:1005–1010.

713 Sêton Bocco S, Csűrös M. 2016. Splice sites seldom slide: Intron evolution in oomycetes. *Genome*
714 *Biol. Evol.* 8:2340–2350.

715 Shaul O. 2017. How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* 91:145–155.

716 Shi Y. 2017. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat.*
717 *Rev. Mol. Cell Biol.* 18:655–670.

718 Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many protein
719 sequences. *Protein Sci.* 27:135–145.

720 Skelly DA, Ronald J, Connelly CF, Akey JM. 2009. Population genomics of intron splicing in 38
721 *Saccharomyces cerevisiae* genome sequences. *Genome Biol. Evol.* 1:466–478.

722 Skrzypek MS, Binkley J, Binkley G, Miyasato SR, Simison M, Sherlock G. 2017. The Candida
723 Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization
724 of high throughput sequencing data. *Nucleic Acids Res.* 45:D592–D596.

725 Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T,
726 Salamov A, Carpenter ML, et al. 2008. The *Trichoplax* genome and the nature of placozoans.
727 *Nature* 454:955–960.

728 Stajich JE, Dietrich FS, Roy SW. 2007. Comparative genomic analysis of fungal genomes reveals
729 intron-rich ancestors. *Genome Biol.* 8:R223.

730 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
731 phylogenies. *Bioinformatics* 30:1312–1313.

732 Stanstrup J. 2017. massageR: Functions for data massaging in R. Available from:
733 https://gitlab.com/R_packages/massageR

734 Stoltzfus A. 2004. Molecular evolution: Introns fall into place. *Curr. Biol.* 14:R351–R352.

735 Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. 2014. Poly(A)-tail profiling reveals an
736 embryonic switch in translational control. *Nature* 508:66–71.

737 Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments
738 into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.

739 Sverdlov AV, Rogozin IB, Babenko VN, Koonin EV. 2005. Conservation versus parallel gains in
740 intron evolution. *Nucleic Acids Res.* 33:1741–1748.

741 Tian L. 2018. SkewGD: Summation of Ks pairs for Exploration of Whole Genome Duplications.
742 Available from: https://github.com/LongTianPy/SkewGD_v1

743 Tonegawa S, Maxam AM, Tizard R, Bernard O, Gilbert W. 1978. Sequence of a mouse germ-line
744 gene for a variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci. U. S. A.*
745 75:1485–1489.

746 Uhse S, Pflug FG, Stirnberg A, Ehrlinger K, von Haeseler A, Djamei A. 2018. In vivo insertion pool
747 sequencing identifies virulence factors in a complex fungal-host interaction. *PLoS Biol.*
748 16:e2005129.

749 Vanneste K, Van de Peer Y, Maere S. 2013. Inference of genome duplications from age distributions
750 revisited. *Mol. Biol. Evol.* 30:177–190.

751 Wang B, Liang X, Gleason ML, Hsiang T, Zhang R, Sun G. 2020. A chromosome-scale assembly of
752 the smallest Dothideomycete genome reveals a unique genome compaction mechanism in
753 filamentous fungi. *BMC Genomics* 21:321.

754 Wang Q-M, Begerow D, Groenewald M, Liu X-Z, Theelen B, Bai F-Y, Boekhout T. 2015. Multigene
755 phylogeny and taxonomic revision of yeasts and related fungi in the *Ustilaginomycotina*. *Stud.*
756 *Mycol.* 81:55–83.

757 Weber MJ. 2006. Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet.*
758 2:e205.

759 Wheeler ML, Limon JJ, Underhill DM. 2017. Immunity to commensal fungi: Detente and disease.
760 *Annu. Rev. Pathol.* 12:359–385.

761 Whelan TA, Lee NT, Lee RCH, Fast NM. 2019. Microsporidian Introns Retained against a
762 Background of Genome Reduction: Characterization of an Unusual Set of Introns. *Genome Biol.*
763 *Evol.* 11:263–269.

764 Wickham H. 2016. ggplot2: Elegant graphics for data analysis. Springer

765 Wiegand HL, Lu S, Cullen BR. 2003. Exon junction complexes mediate the enhancing effect of
766 splicing on mRNA expression. *Proc. Natl. Acad. Sci. U. S. A.* 100:11327–11332.

767 Witten JT, Ule J. 2011. Understanding splicing regulation through RNA splicing maps. *Trends Genet.*
768 27:89–97.

769 Worden AZ, Lee J-H, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E,

- 770 Everett MV, et al. 2009. Green evolution and dynamic adaptations revealed by genomes of the
771 marine picoeukaryotes *Micromonas*. *Science* 324:268–272.
- 772 Wu B, Macielog AI, Hao W. 2017. Origin and spread of spliceosomal introns: Insights from the
773 fungal clade *Zymoseptoria*. *Genome Biol. Evol.* 9:2658–2667.
- 774 Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*
775 24:1586–1591.
- 776 Yarza P, Yilmaz P, Panzer K, Glöckner FO, Reich M. 2017. A phylogenetic framework for the
777 kingdom *Fungi* based on 18S rRNA gene sequences. *Mar. Genomics* 36:33–39.
- 778 Yenerall P, Krupa B, Zhou L. 2011. Mechanisms of intron gain and loss in *Drosophila*. *BMC Evol.*
779 *Biol.* 11:364.
- 780 Yu C-H, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS, Liu Y. 2015. Codon usage influences the local
781 rate of translation elongation to regulate co-translational protein folding. *Mol. Cell* 59:744–754.
- 782 Yu G, Smith DK, Zhu H, Guan Y, Lam TT. 2017. ggtree : an r package for visualization and
783 annotation of phylogenetic trees with their covariates and other associated data. McNerny G,
784 editor. *Methods Ecol. Evol.* 8:28–36.
- 785 Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological
786 themes among gene clusters. *OMICS* 16:284–287.
- 787 Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A,
788 Girón CG, et al. 2018. Ensembl 2018. *Nucleic Acids Res.* 46:D754–D761.
- 789 Zhang L-Y, Yang Y-F, Niu D-K. 2010. Evaluation of models of the mechanisms underlying intron loss
790 and gain in *Aspergillus* fungi. *J. Mol. Evol.* 71:364–373.
- 791 Zhang Y, Sachs MS. 2015. Control of mRNA stability in fungi by NMD, EJC and CBC factors
792 through 3'UTR introns. *Genetics* 200:1133–1148.
- 793 Zhu Q. 2014. AfterPhylo: a Perl script for manipulating trees after phylogenetic reconstruction.
794 Available from: <https://github.com/qiyunzhu/AfterPhylo>
- 795 Zhu T, Niu D-K. 2013a. Frequency of intron loss correlates with processed pseudogene abundance: a
796 novel strategy to test the reverse transcriptase model of intron loss. *BMC Biol.* 11:23.
- 797 Zhu T, Niu D-K. 2013b. Mechanisms of intron loss and gain in the fission yeast
798 *Schizosaccharomyces*. *PLoS One* 8:e61683.

799

800

801

802

803

804

Table 1. Intron densities of the ancestral and current states of fungal clades. See also Figure 1 and 2.

Clade	Number of Species	Ancestral state ^a	Current state ^b	
			Mean	Median
Cryptomycota (<i>Rozella allomyces</i>)	1	NA	2.70	2.70
Microsporidia	15	0.06	0.01	0.00
Chytridiomycota	2	5.92	4.25	4.25
Blastocladiomycota (<i>Allomyces macrogynus</i>)	1	NA	1.43	1.43
Entomophthoromycotina (<i>Conidiobolus coronatus</i>)	1	NA	1.70	1.70
Mucaromycota	5	6.05	2.67	2.20
Pucciniomycotina	8	4.20	3.09	3.03
Ustilaginomycotina	20	2.93	0.56	0.16
Agaricomycotina	39	4.67	3.57	3.79
Taphrinomycotina	13	4.51	1.45	1.24
Saccharomycotina	36	1.68	0.05	0.01
Pezizomycotina	122	1.06	0.52	0.52

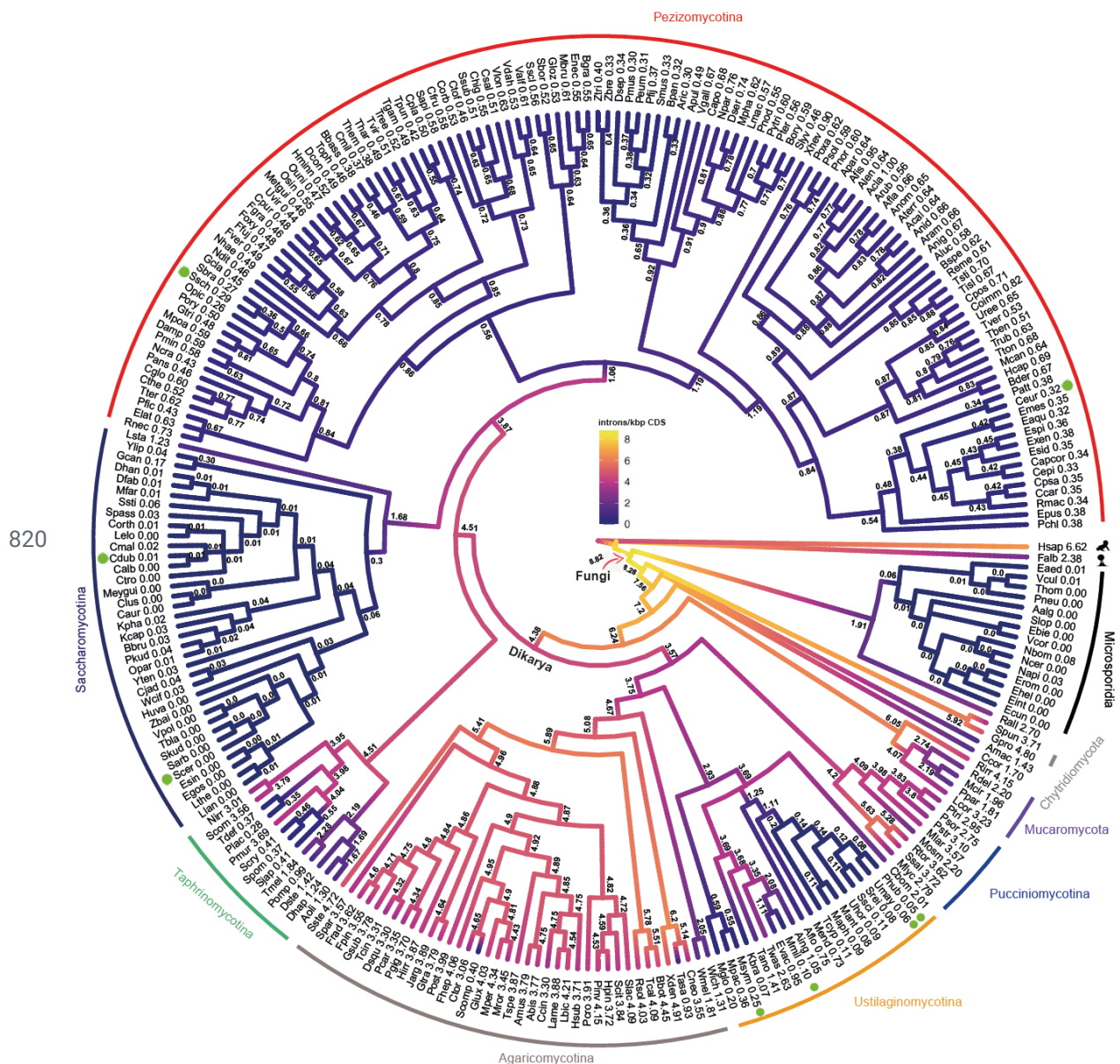
^a Obtained from the inference of intron gain/loss.

^b Arithmetic mean or median inferred introns/kbp of the species within a clade.

Introns/kbp, the number of introns per one kbp of protein-coding sequence; NA, not applicable.

818 FIGURES

819



821

822 **Figure 1. Widespread loss of introns during the evolution of *Fungi*.** Ancestral introns were
823 inferred from 1445 orthologs in 263 fungal species using Malin (Csuros, 2008), a Markov model with
824 rates across sites and branch-specific gain and loss rates. Branches are color-coded with intron
825 densities from the median posterior distribution for each node. A list of full names and intron densities
826 are available in Supplementary Table S2. See also related Table 1 and Figure 2. Introns/kbp, the
827 number of introns per one kbp of protein-coding sequence. Green filled circles denote eight intron
828 poor species selected for additional analysis.

829

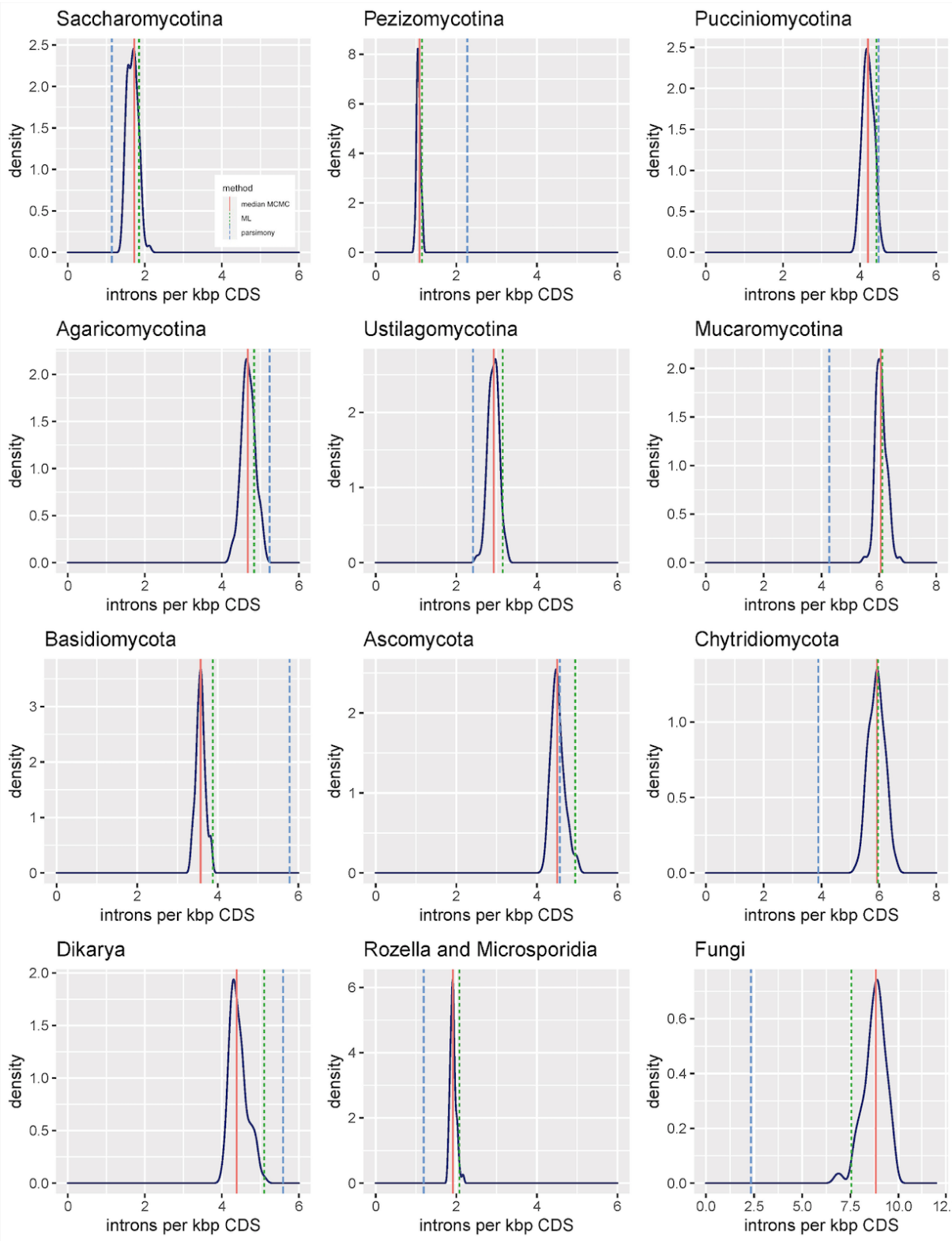


Figure 2. Intron densities of the fungal ancestral states derived from a Monte-Carlo approximation of 100 bootstrap distributions. Dotted lines denote the ancestral intron densities inferred from Dollo parsimony (blue) and maximum likelihood (ML, green) models. See also related Figure 1 and Table 1. Introns/kbp, the number of introns per one kbp of protein-coding sequence.

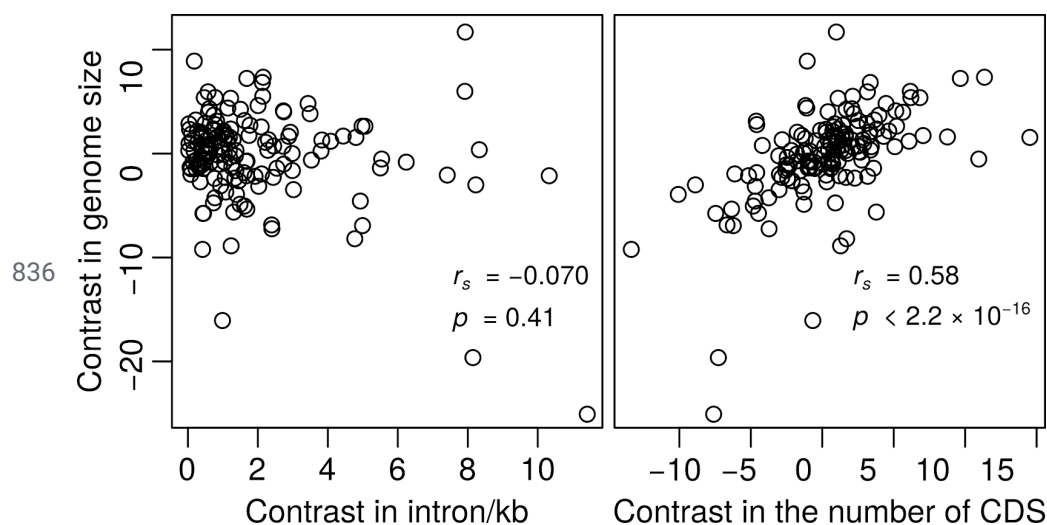
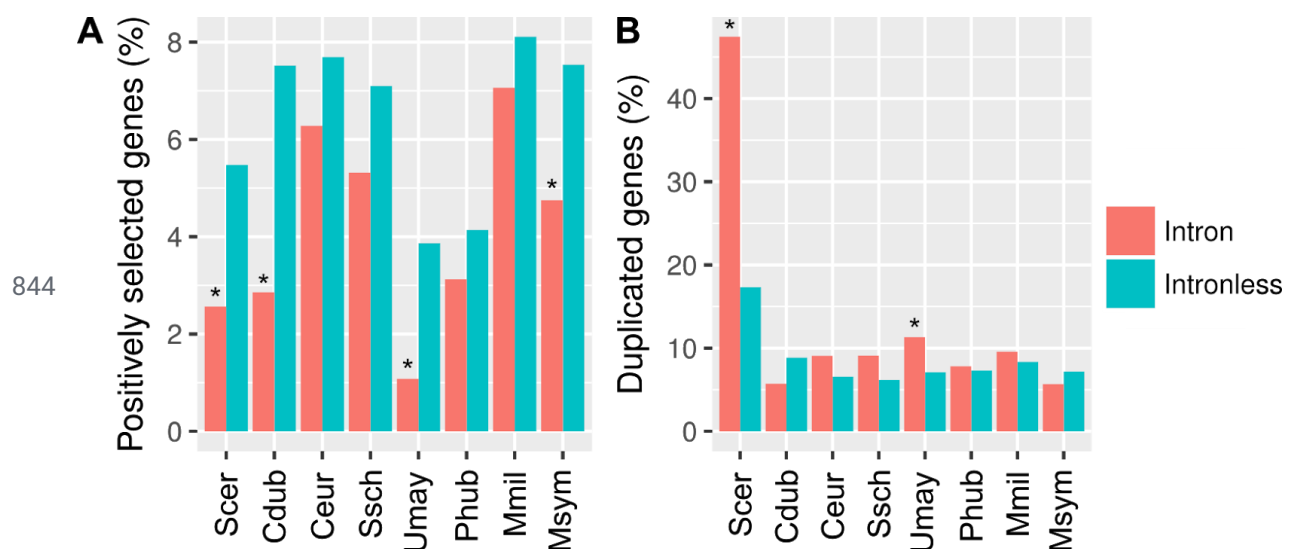


Figure 3. Intron density weakly correlates with genome size. Phylogenetic independent contrasts analysis of genome size versus intron density and the number of protein-coding genes. CDS, coding sequence; introns/kbp, the number of introns per one kbp of protein-coding sequence; r_s , Spearman's rho.

842

843



845

Figure 4. Features of intron-containing genes in intron-poor fungal species. Proportion of intronless and intron-containing genes that have undergone (A) positive selection and (B) gene duplication. *, $p < 0.05$ (Fisher's exact test, adjusted using the Bonferroni correction); Cdub, *Candida dubliniensis*; Ceur, *Cyphellophora europaea*; Mmil, *Meira miltonrushii*; Msym, *Malassezia sympodiali*; Phub, *Pseudozyma hubeiensis*; Scer, *Saccharomyces cerevisiae*; Ssch, *Sporothrix schenckii*; Umay, *Ustilago maydis*.

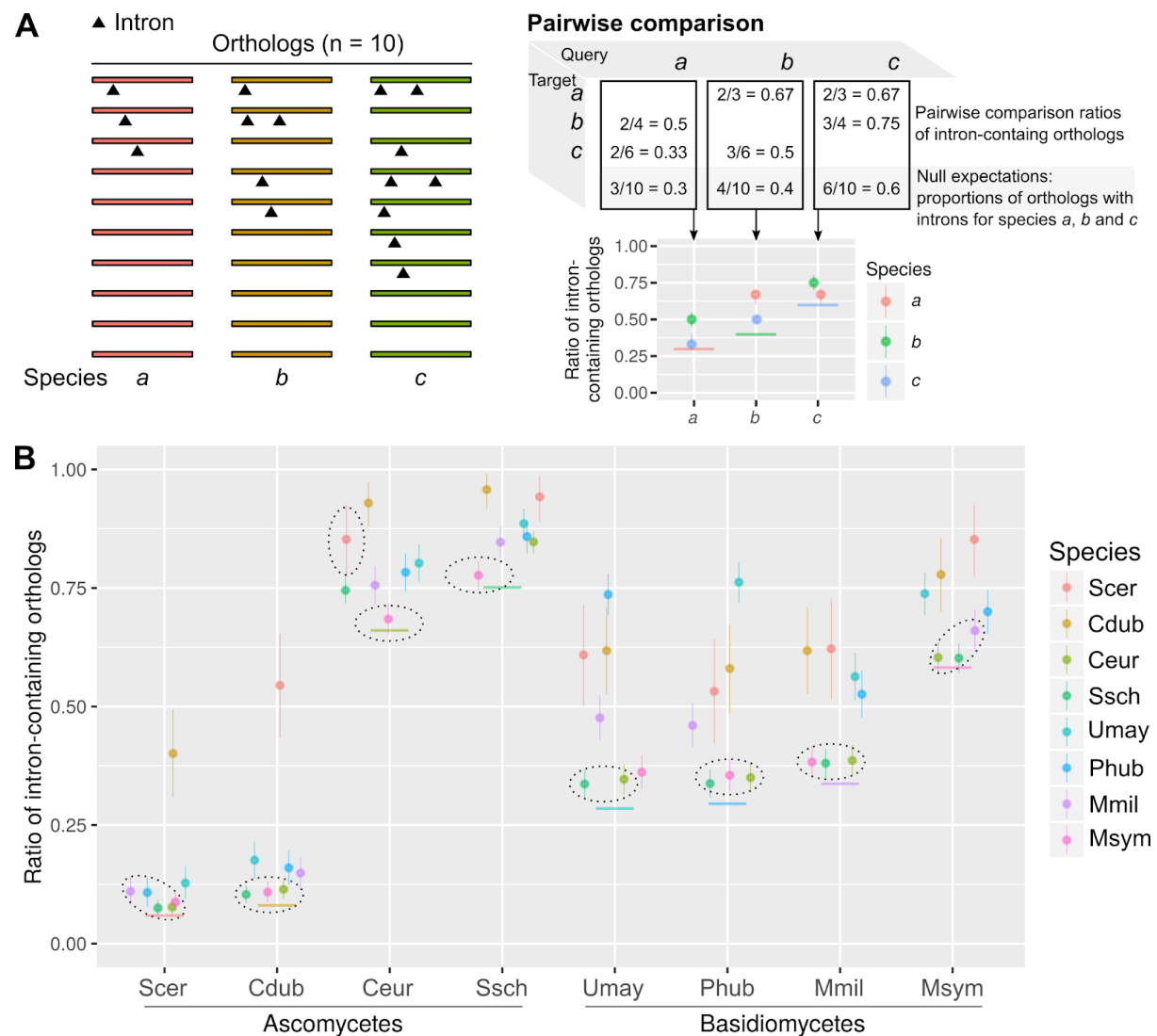
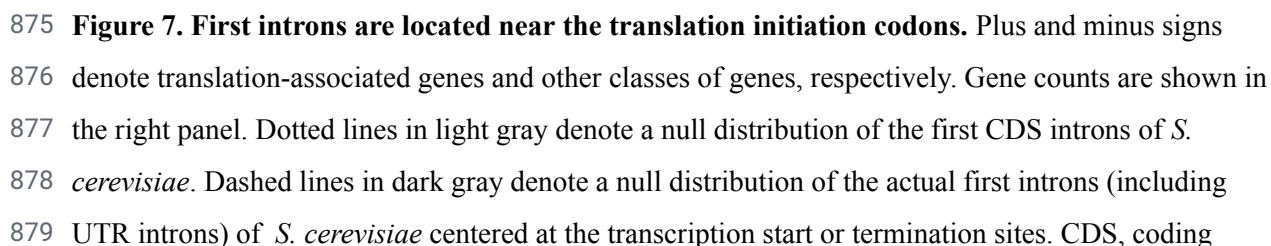


Figure 5. Orthologous genes harbor introns concordantly. (A) Schematic example of a pairwise comparison of intron-containing orthologs among three species. **(B)** The ratios of intron-containing orthologs in a pairwise comparison in contrast to null expectations (solid horizontal colored lines). Non-significant chi-square tests (dotted circles, Bonferroni adjusted p -value > 0.01) suggest that introns can be retained in any genes. As a result, 34 of 56 comparisons between species are statistically significant. The binomial confidence intervals (95%) were estimated from these ratios using Bayesian inference with 1000 iterations (vertical colored lines). Cdub, *Candida dubliniensis*; Ceur, *Cyphellophora europaea*; Mmil, *Meira miltonrushii*; Msym, *Malassezia sympodiali*; Phub, *Pseudozyma hubeiensis*; Scer, *Saccharomyces cerevisiae*; Ssch, *Sporothrix schenckii*; Umay, *Ustilago maydis*.

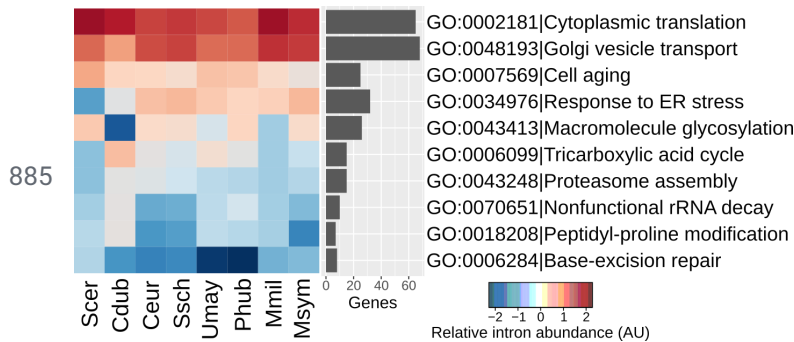
NOG2		5'SS	ASE <<<< >>>>	stem-loop 1 ASE	H box
Scer YNR053C	D L V P T V A	guauguaaacuuuuu...90..	gucagggu...40..	gucguguacuguuuuuggaagaga...114..	
Cdub CAx41289	D L V P T V A	guauguuuaaaaaaa...103..	gucagggu...60..	gcucguuacacuuuuugaagaaua...152..	
Ceru ETN383B7	D L V P T I A	guaaugugccuucuu...31..	gucagggc...34..	gucguuacacgcuuuuguuaauuu...82..	
Ssch ERS99730	D L V P S G V A	guaagcugcucucagu...135..	gucaggaa...41..	uugcguaacaacgcuuuuggaagau...101..	
Umay KIS71048	D L V P T V T	guaaugguuuuaccu...75..	gucaguaa...45..	aacuguuacgaauvguugaagaaga...154..	
Phubi GAC97322	D L V P T V T	guaauguuuucccaac...83..	gucaguaa...45..	aacuguuacgaauvguugaagaaga...156..	
Mym 6037	D L V P T V T	gugaguuuuuuuuuu...43..	gucaggac...23..	gcuguuacgaauvguugaagaaua...138..	
Msmj CCO98745	D L V P T V T	gucaguuuuuuccccc...32..	gucagggg...37..	cacuguuacgaauvguugaagaaga...158..	

Figure 6. Introns of *RPL7B* and *NOG2* have conserved positions. The introns of *RPL7B* and *NOG2* encode box C/D and box H/ACA snoRNAs (snR59 and snR191 in *S. cerevisiae*, respectively). The predictions of stem-loop 2 and antisense element (ASE) of the *M. miltonrushii* box H/ACA snoRNA are of low confidence. 5' SS and 3' SS denote 5' and 3' splice-sites, respectively. Cdub, *Candida dubliniensis*; Ceur, *Cyphellophora europaea*; Mmil, *Meira miltonrushii*; Msym, *Malassezia sympodiali*; Phub, *Pseudozyma hubeiensis*; Scer, *Saccharomyces cerevisiae*; Ssch, *Sporothrix schenckii*; Umay, *Ustilago maydis*.



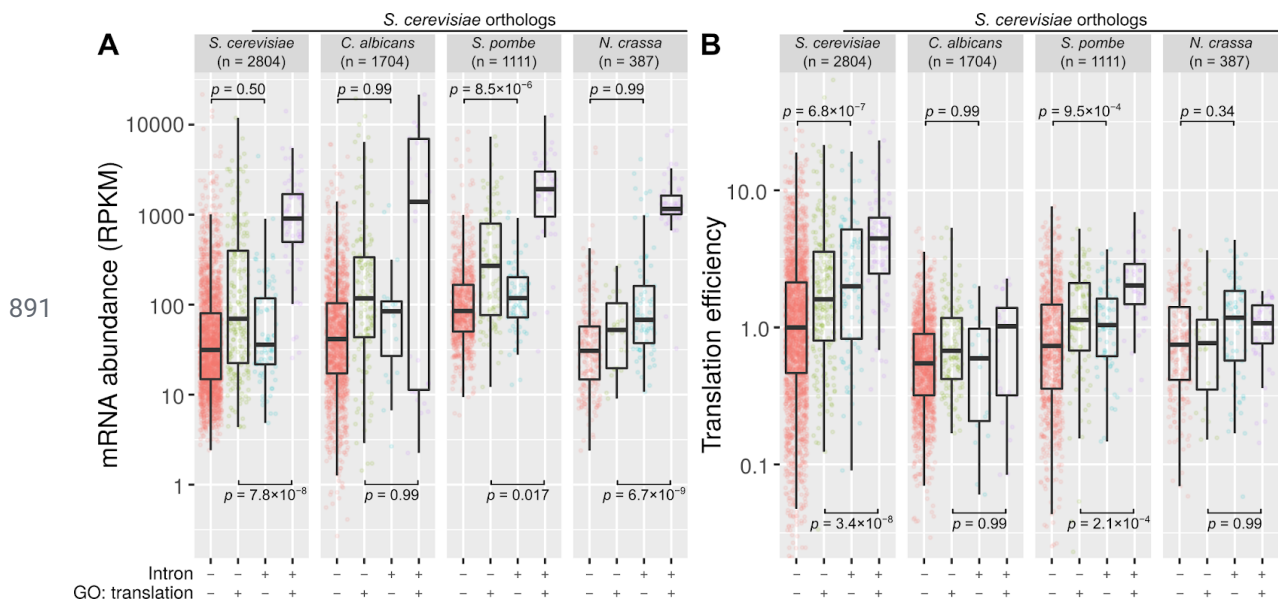
sequence; Cdub, *Candida dubliniensis*; Ceur, *Cyphellophora europaea*; GO, gene ontology; Mmil, *Meira miltonrushii*; Msym, *Malassezia sympodiali*; Phub, *Pseudozyma hubeiensis*; Scer, *Saccharomyces cerevisiae*; Ssch, *Sporothrix schenckii*; Umay, *Ustilago maydis*; UTR, untranslated regions.

884



886 **Figure 8. Introns are more abundant in specific classes of genes.** Cdub, *Candida dubliniensis*;
887 Ceur, *Cyphellophora europaea*; Mmil, *Meira miltonrushii*; Msym, *Malassezia sympodiali*; Phub,
888 *Pseudozyma hubeiensis*; Scer, *Saccharomyces cerevisiae*; Ssch, *Sporothrix schenckii*; Umay, *Ustilago*
889 *maydis*.

890



892 **Figure 9. Intron-containing genes have higher mRNA expression and translation efficiency. (A)**
893 RNA-seq and **(B)** ribosome profiling results of *S. cerevisiae* orthologs. Translation efficiency was
894 determined by the ratio of ribosome-protected fragments and mRNA read counts normalized to
895 respective library sizes. *C. albicans*, *Candida albicans*; *N. crassa*, *Neurospora crassa*; *p*, the *p*-values
896 of Welch two sample t-tests were adjusted with Bonferroni correction; RPKM, Reads Per Kilobase
897 per Million mapped reads; *S. cerevisiae*, *Saccharomyces cerevisiae*; *S. pombe*, *Schizosaccharomyces*
898 *pombe*.