

23

24 **ABSTRACT**

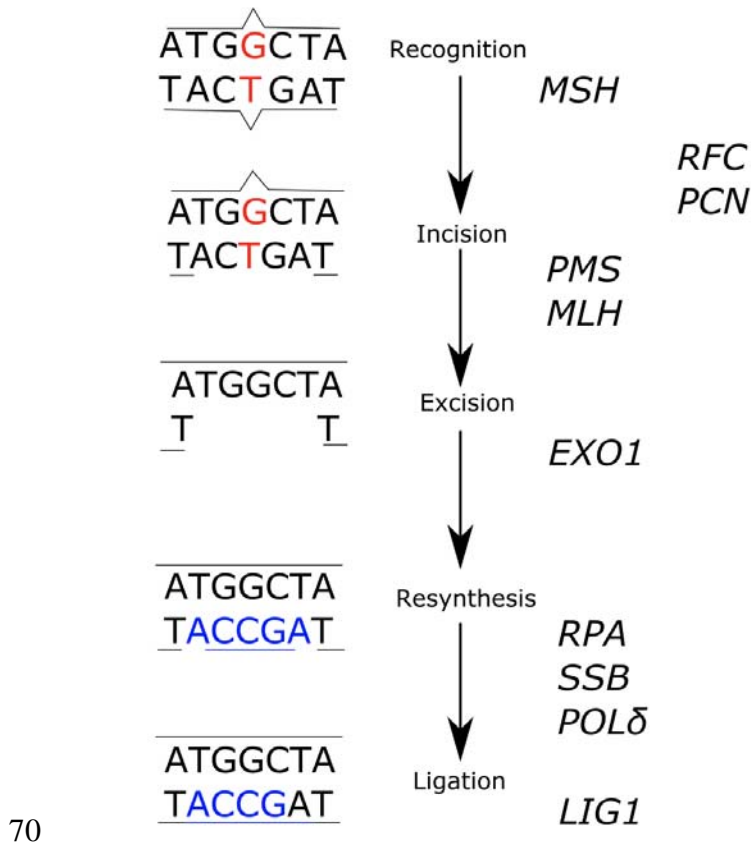
25 The DNA mismatch repair (MMR) pathway corrects mismatched bases produced during DNA
26 replication and is highly conserved across the tree of life, reflecting its fundamental importance
27 for genome integrity. Loss of function in one or a few MMR genes can lead to increased
28 mutation rates and microsatellite instability, as seen in some human cancers. While loss of MMR
29 genes has been documented in the context of human disease and in hypermutant strains of
30 pathogens, examples of entire species and species lineages that have experienced substantial
31 MMR gene loss are lacking. We examined the genomes of 1,107 species in the fungal phylum
32 Ascomycota for the presence of 52 genes known to be involved in the MMR pathway of fungi.
33 We found that the median ascomycete genome contained 49 / 52 MMR genes. In contrast, four
34 closely related species of obligate plant parasites from the powdery mildew genera *Erysiphe* and
35 *Blumeria*, have lost between 6 and 22 MMR genes, including *MLH3*, *EXO1*, and *DPB11*. The
36 lost genes span MMR functions, include genes that are conserved in all other ascomycetes, and
37 loss of function of any of these genes alone has been previously linked to increased mutation
38 rate. Consistent with the hypothesis that loss of these genes impairs MMR pathway function, we
39 found that powdery mildew genomes with high levels of MMR gene loss exhibit increased
40 numbers of monomer repeats, longer microsatellites, accelerated sequence evolution, elevated
41 mutational bias in the A|T direction, and decreased GC content. These results identify a striking
42 example of macroevolutionary loss of multiple MMR pathway genes in a eukaryotic lineage,
43 even though the mutational outcomes of these losses appear to resemble those associated with
44 detrimental MMR dysfunction in other organisms.

45 **Introduction**

46 An ensemble of DNA repair pathways and cell cycle checkpoints are responsible for detecting
47 and repairing DNA damage, ensuring faithful maintenance of the genome (Friedberg et al. 2005;
48 Giglia-Mari et al. 2010). Among DNA repair pathways, the DNA mismatch repair (MMR)
49 pathway is arguably one of the best characterized (Marti et al. 2002). The MMR pathway is
50 responsible for repairing bases that were incorrectly paired during DNA replication via five
51 steps: recognition, incision, removal, re-synthesis, and ligation (Fig. 1) (Hsieh & Zhang 2017;
52 Fukui 2010; Marti et al. 2002). The MMR pathway is highly conserved in both bacteria and
53 eukaryotes; cells that experience reduction or loss of function in this pathway have increased
54 levels of mutation, as seen in cancer and drug-resistant fungal pathogen strains (Fukui 2010;
55 Campbell et al. 2017; Billmyre et al. 2017, 2020; Dos Reis et al. 2019).

56
57 Although DNA repair genes are generally highly conserved, certain fungal lineages have been
58 reported to have a more limited repertoire of DNA repair genes, particularly within the phylum
59 Ascomycota. For example, budding yeasts (subphylum Saccharomycotina) and fission yeasts
60 (Taphrinomycotina) have fewer DNA repair genes than filamentous fungi (Pezizomycotina)
61 (Milo et al. 2019; Shen et al. 2020). Furthermore, DNA repair genes that were lost from budding
62 yeasts and fission yeasts are more likely to also be lost in filamentous fungi (Milo et al. 2019).
63 One lineage that has endured extensive losses in its repertoire of DNA repair genes is the
64 *Hanseniaspora* genus of budding yeasts (Steenwyk et al. 2019). *Hanseniaspora* species have
65 undergone punctuated sequence evolution and have accumulated large numbers of diverse types
66 of substitutions, including those associated with specific gene losses such as UV damage,
67 suggesting that DNA repair is impaired by the high levels of DNA repair gene loss. These

68 findings suggest that DNA repair genes are not universally conserved across fungi and that their
69 loss is compatible with long-term evolutionary survival and diversification of fungal lineages.



71 **Fig. 1. The DNA Mismatch Repair (MMR) pathway corrects mismatched bases produced**
72 **during DNA replication and prevents instability in microsatellites.** The pathway is comprised
73 of five conserved steps: recognition of mispaired bases, incision of the DNA strand, excision of
74 the incorrectly paired bases, resynthesis of the DNA strand, and ligation of the newly synthesized
75 segment to the DNA strand. *RFC* and *PCNA* genes are involved with multiple steps in the
76 pathway. We categorized *RFC* genes as being primarily responsible for resynthesis, though they
77 are also involved in loading proliferating cell nuclear antigen (Surtees & Alani 2004). We
78 categorize proliferating cell nuclear antigen (*PCN1*, *POL30*) as a recognition gene, though it has
79 been implicated in early and late steps in the MMR pathway (Lau et al. 2002; Schofield & Hsieh
80 2003; Umar et al. 1996; Surtees & Alani 2004). In addition, *RFC* and *PCNA* can interact with
81 *EXO1* to modulate excision directionality (Surtees & Alani 2004).

82

83 One well-established consequence of MMR dysfunction is mutation in microsatellite regions of
84 the genome. Microsatellites are repetitive tracts of DNA, with motifs 1-6 bp long repeated at
85 least five times (Beier et al. 2017). Microsatellites are typically highly polymorphic between

86 individuals and are commonly used as markers in population biology, forensics, paternity testing,
87 and tumor characterization (Richman 2015). Due to their repetitive nature, microsatellites are
88 prone to experiencing polymerase slippage, which is usually corrected by the MMR pathway
89 (Ellegren 2004; Richman 2015). If the MMR pathway does not recognize these errors, as is the
90 case in cancer, microsatellite instability (MSI) can occur (Campbell et al. 2017). MSI is defined
91 by a hypermutable phenotype resulting from a loss of function in the MMR pathway (Boland &
92 Goel 2010). Instability in microsatellites trends towards elongation in these regions, but
93 contraction can also occur (Ellegren 2004).

94

95 Beyond increased mutation in microsatellite regions, aberrant function of the MMR pathway is
96 associated with genome-wide signatures of genetic instabilities (Boland & Goel 2010; Billmyre
97 et al. 2017, 2020). MMR mutations have been implicated in the development of hypermutant and
98 ultrahypermutant human cancers, which constitute approximately 15% of human tumors and less
99 than 1% of tumors, respectively (Campbell et al. 2017). Interestingly, very few tumors with low
100 mutation rates contained mutations in the MMR pathway, whereas more than a third of
101 hypermutant tumors and virtually all the ultrahypermutants contained mutations in MMR genes
102 (Campbell et al. 2017). Hypermutant tumors had high levels of MSI suggesting their
103 hypermutant phenotype is due, at least in part, to MMR dysfunction (Campbell et al. 2017).

104 Hypermutation has also been observed in fungal pathogen strains that have lost MMR pathway
105 genes, potentially driving within-host adaptation and the evolution of drug resistance. For
106 example, Rhodes *et al.* (2017) found that hypermutation caused by mutations in three MMR
107 pathway genes, including *MSH2*, resulted in a rapid increase in the mutation rate of the human
108 pathogenic fungus *Cryptococcus neoformans*, contributing to infection relapse. Similarly,

109 Billmyre *et al.* (2017) sequenced multiple strains of the human pathogenic fungus *Cryptococcus*
110 *deuterogattii* (phylum Basidiomycota) and found that a group of strains with mutations in the
111 *MSH2* gene experienced higher rates of mutation when compared with closely related strains
112 harboring an intact *MSH2* gene. Hypermutation in *C. deuterogattii* mediated rapid evolution of
113 antifungal drug resistance (Billmyre *et al.* 2017, 2020).

114
115 In contrast to MMR gene loss in the microevolutionary context of genetic or infectious disease,
116 the extent of MMR gene loss across lineages spanning multiple species remains understudied. To
117 determine the macroevolutionary impact of MMR gene conservation and loss, we characterized
118 patterns of MMR gene presence and absence in the fungal phylum Ascomycota (Fig. 2). We
119 found that the MMR pathway was highly conserved across Ascomycota, with the median species
120 having 49 / 52 MMR genes present. However, we found that *Blumeria graminis* and species in
121 the powdery mildew genus *Erysiphe* (subphylum Pezizomycotina, class Leotimycetes), a group
122 of obligate plant parasites, had many fewer MMR genes and a faster rate of sequence evolution
123 than their relatives and most other fungal taxa. Specifically, *Erysiphe necator* has lost 10 MMR
124 genes, *Erysiphe pisi* has lost 22 MMR genes, *Erysiphe pulchra* has lost 9 MMR genes, and
125 *Blumeria graminis* has lost 6 MMR genes (Fig. 3). In contrast, species closely related to
126 *Erysiphe* and *Blumeria* have lost only 1 – 3 MMR genes, consistent with the high degree of
127 MMR gene conservation in the rest of the phylum. Evolutionary genomic analyses revealed that
128 MMR gene losses in *Erysiphe* and *Blumeria* (hereafter referred to as high loss taxa or HLT) are
129 associated with a proliferation of monomeric runs and elongation of microsatellites of all motif
130 lengths, both of which are hallmarks of MMR pathway dysfunction. Reflecting these losses,
131 *Erysiphe* and *Blumeria* genomes also have more pronounced mutational biases and accelerated

132 rates of mutation. These results suggest that obligate plant parasites in the genera *Blumeria* and
133 *Erysiphe* have diversified while lacking otherwise highly conserved MMR genes.

134

135 **METHODS**

136 **Curation of the set of DNA mismatch repair pathway genes**

137 To investigate the presence and absence of MMR genes across the fungal phylum Ascomycota,
138 we curated a dataset of MMR genes from the genomes of three fungal model organisms
139 representing the three different subphyla: *Saccharomyces cerevisiae* (subphylum
140 Saccharomycotina), *Neurospora crassa* (Pezizomycotina), and *Schizosaccharomyces pombe*
141 (Taphrinomycotina). We used three sources to curate genes that are part of the MMR pathway:
142 the Kyoto Encyclopedia of Genes and Genomes ([KEGG, genome.jp/kegg/](http://KEGG.genome.jp/kegg/); Kanehisa & Goto,
143 [2000](#)), the *Schizosaccharomyces pombe* database (PomBase, pombase.org/; Lock et al., 2019;
144 The Gene Ontology Consortium, 2019), and the *Saccharomyces* Genome Database (SGD,
145 yeastgenome.org/; Cherry et al., 2012). All genes in the KEGG diagram of the MMR pathway
146 for each species were included and the gene ontology (GO) term “mismatch repair” was used to
147 search for the genes on SGD and Pombase (Ashburner et al. 2000). We used both
148 computationally and manually curated genes from SGD. We began curating our set of MMR
149 genes in *S. cerevisiae*, with a total of 30 MMR genes identified with KEGG and SGD. Next, we
150 searched KEGG and Pombase for genes in *S. pombe* that had not been annotated as part of the
151 MMR pathway in *S. cerevisiae* (n = 15). We concluded by searching for *N. crassa* MMR genes
152 in KEGG which had not already been categorized as MMR genes in the other two species (n =7).
153 KEGG listed two sequences for the *N. crassa* gene *LIG1*; however, since our sequence similarity

154 search analyses with both sequences yielded identical patterns of loss, we present them as one
 155 gene. This approach yielded a total of 52 genes associated with MMR (Table 1).

156

157 **Table 1. List and function of MMR genes used in this study.**

	<i>S. cerevisiae</i>	<i>S. pombe</i>	<i>N. crassa</i>
Recognition	<i>HSM3, MSH1, MSH2, MSH3, MSH4, MSH5, MSH6, NHP6A, POL30, RAD34, RAD4</i>	<i>PCN1, RHP41, RHP42</i>	---*
Incision	<i>MLH1, MLH2, MLH3, PMS1, RNH201</i>	<i>HNT3, MYH1,</i>	<i>PMS2</i>
Removal	<i>DIN7, EXO1</i>	<i>UVE1</i>	---*
Re-synthesis	<i>DPB11, POL31, POL32, POL3, RFA1, RFA2, RFC1, RFC2, RFC3, RFC4, RFC5</i>	<i>CDC1, CDC27, CDC6, CDM1, SSB1, SSB2, SSB3</i>	<i>POLD1, POLD2, POLD3, POLD4, RPA3</i>
Ligation	<i>CDC9</i>	<i>ADL1, CDC17</i>	<i>LIG1</i>

158 Genes sourced from Chang et al., 2001; Chauleau et al., 2015; Fukui, 2010; KEGG; Kaur et al., 1999; Marti et al.,
 159 2003; Reyes et al., 2015

160 *no *N. crassa* genes were selected for that function.

161

162

163 **MMR gene conservation analysis**

164 To examine the conservation of MMR genes across Ascomycota, we implemented a sensitive
 165 probabilistic modeling approach using profile Hidden Markov Models (pHMMs) (Johnson et al.
 166 2010) of MMR genes across the genomes of 1,107 species (Shen et al. 2020). To construct
 167 pHMMs, we first searched for putative homologs of MMR genes in the fungal RefSeq protein
 168 database using the blastp function of BLAST+, v2.8.1, with a bitscore threshold of 50 and an e-
 169 value cutoff of 1×10^{-3} (Pearson 2013). We retrieved the top 100 hits using SAMTOOLS, v1.6 (Li
 170 et al. 2009) with the ‘faidx’ function. We used MAFFT, v7.402 (Katoh et al. 2002), with the
 171 ‘genafpair’ and ‘reordered’ parameters, a maximum of 1000 cycles of iterative refinement, the
 172 BLOSUM62 matrix, a gap opening penalty of 1.0, and the retree parameter set to 1, to align the
 173 sequences following previously established protocol (Steenwyk, Shen, et al. 2019). We then used
 174 the aligned amino acid sequences as input to the ‘hmmbuild’ function in HMMER-3.1B2 to

175 construct each pHMM. We ran the pHMMs of the 52 proteins against all 1,107 proteomes using
176 the ‘hmmsearch’ function. For a gene to be considered present, we set a bitscore threshold of at
177 least 50 and an e-value threshold of less than 1×10^{-6} . We used the tblastn function of BLAST+,
178 v2.8.1 with a bitscore threshold of 50, e-value cutoff of 1×10^{-6} , and 50% minimum query
179 coverage to verify MMR gene absence using the protein sequence of the gene in question and the
180 1,107 Ascomycota genomes. We used the Interactive Tree of Life (iTOL), v4 (Letunic & Bork
181 2019) to visualize the conservation of MMR genes on the Ascomycota phylogeny and to map
182 losses on it.

183

184 **Microsatellite identification and characterization**

185 To identify microsatellites and evaluate their numbers and lengths between genomes with
186 substantial MMR gene loss against those with higher levels of MMR gene conservation, we used
187 the Microsatellite Identification tool (MISA), v2.0 (Beier et al. 2017). Specifically, we compared
188 the microsatellites of two groups of taxa, each of which contained four species. The group of
189 high loss taxa (HLT) contains the powdery mildews *Blumeria graminis*, *Erysiphe necator*,
190 *Erysiphe pisi*, and *Erysiphe pulchra*, which show high levels of MMR gene loss relative to other
191 ascomycetes. The group of low loss taxa (LLT) contains four closely related species with low
192 levels of MMR gene loss, similar to patterns seen across the rest of the phylum: *Articulospora*
193 *tetracladia*, *Ascocoryne sarcooides*, *Cairneyella variabilis*, and *Phialocephala scopiformis*. The
194 length minimums used for MISA to identify a microsatellite are as follows: 1 base pair (bp)
195 motifs must repeat 12 times, 2 bp motifs must repeat 6 times, 3-6 bp motifs must repeat 5 times.
196 All values used are MISA defaults, except the monomeric parameter, which was increased from
197 the default value of 10 repeats to 12 (Beier et al. 2017; Temnykh et al. 2001). A 2-way ANOVA
198 test was performed to test for significance in the number of microsatellites controlled by genome

199 size of each motif length between HLT and LLT. If the 2-way ANOVA rejected the null
200 hypothesis ($\alpha = 0.05$), pairwise comparisons were made with the Tukey Honest Significant
201 Differences (HSD) test. We performed the statistical analysis using R, v3.4.1 ([https://www.r-](https://www.r-project.org/)
202 [project.org/](https://www.r-project.org/)) and made the figures using ggplot2, v3.1.0 (Wickerham 2016), and ggpubfigs,
203 v1.0.0 (Steenwyk 2020).

204

205 **Estimation of mutational bias and rate of sequence evolution**

206 To characterize the mutational spectra and estimate the rate of sequence evolution between HLT
207 and LLT, we first identified and aligned orthologous sequences across all eight genomes.
208 Orthologous single-copy protein sequences from genes present in all eight genomes ($n = 823$)
209 were identified using the BUSCO, v4.0.4 (Waterhouse et al. 2018) pipeline and the OrthoDB,
210 v10, Ascomycota database (Creation date: 2019-11-20) (Kriventseva et al. 2019). We hereafter
211 refer to the 823 single-copy genes as BUSCO genes. BUSCO genes were aligned using MAFFT,
212 v7.402 (Kato et al. 2002), using the same settings described above. Codon-based alignments
213 were generated by threading the corresponding nucleotide sequences onto the protein alignment
214 using 'thread_dna' function in PhyKIT, v0.1 (Steenwyk et al. 2020).

215

216 To examine patterns of substitutions, we used codon-based alignments to identify nucleotides
217 that differed in a given taxon of interest compared to *C. variabilis*, which was the sister taxon to
218 a clade comprised of the other seven genomes of interest in the Ascomycota phylogeny. More
219 specifically, we compared the character states for a species of interest to *C. variabilis* for each
220 site of each alignment, tracking codon position information (i.e., first, second, or third codon
221 position). We also determined if the substitution was a transition or transversion and examined

222 substitution directionality (e.g., A|T to G|C or G|C to A|T) using *C. variabilis* as the outgroup.
223 These analyses were completed using custom python scripts that utilize functions in Biopython,
224 v1.70 (Cock et al. 2009).
225
226 Finally, we used the codon alignments to compare the rate of sequence evolution between HLT
227 and LLT. Specifically, we measured the ratio of the rate of nonsynonymous substitutions to the
228 rate of synonymous substitutions (dN/dS or ω) along the species phylogeny for each gene using
229 the CODEML function in PAML, v4.9 (Yang 2007). For each test, the null hypothesis (H_0) was
230 that all branches had the same ω value (model = 0); the alternative hypothesis (H_A) was that all
231 HLT branches, including the branch of their most recent common ancestor, had one ω value and
232 all other branches had a distinct ω value (model = 2). To determine if the alternative hypothesis
233 was a better fit than the null hypothesis ($\alpha = 0.05$) we used a likelihood ratio test.

234

235 **Results**

236 **MMR genes are highly conserved across the fungal phylum Ascomycota**

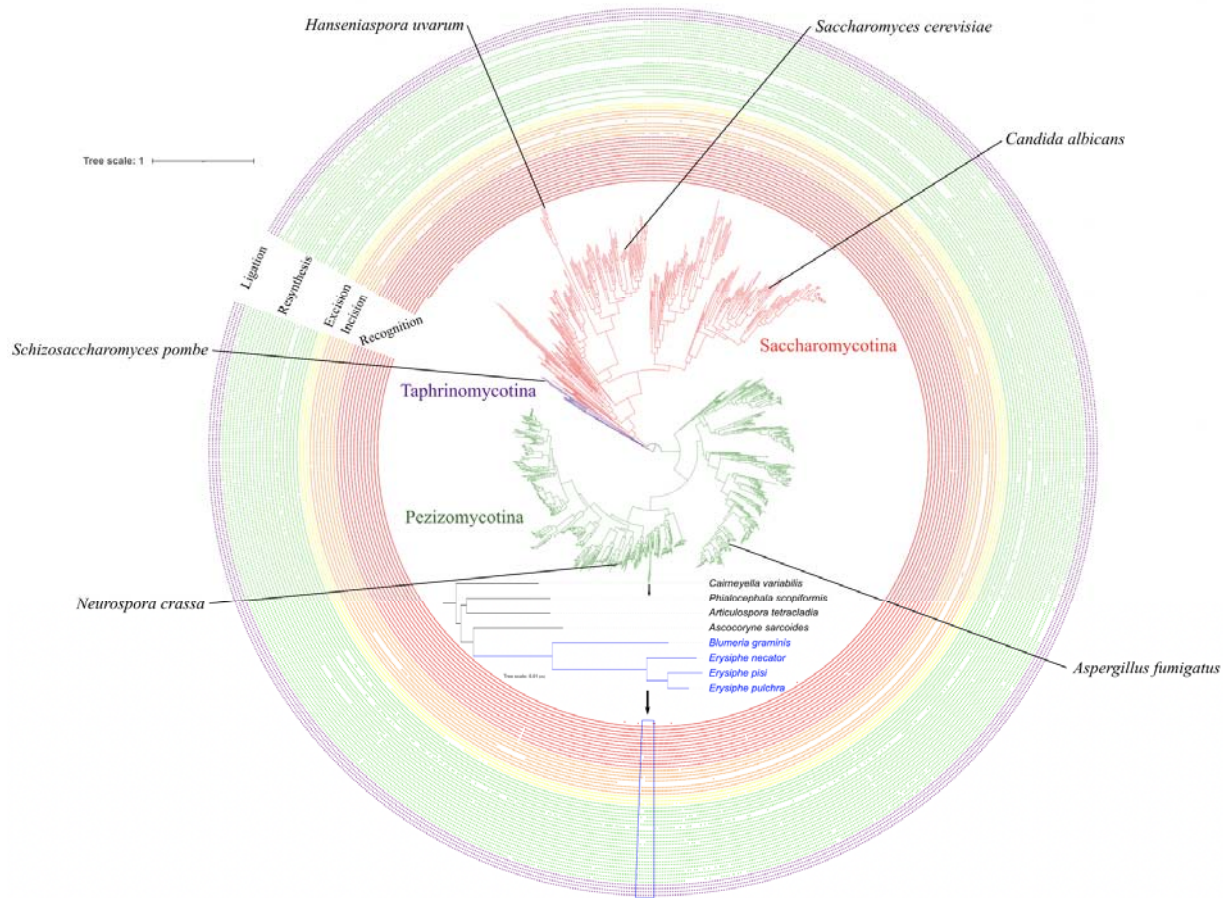
237 By examining the presence of 52 MMR genes using a combination of sequence similarity search
238 algorithms across the genomes of 1,107 fungal species, we found that the MMR pathway is
239 highly conserved across Ascomycota (a median of 49 / 52 MMR genes per species; Fig. 2; File
240 S1). Sixteen genes were present in all species; these included five recognition genes (*MSH1*,
241 *MSH2*, *MSH3*, *MSH4*, and *MSH6*), one incision gene (*MLH1*), one removal gene (*DIN7*), five
242 resynthesis genes (*CDC6*, *RFC2*, *RFC3*, *RFC4*, and *RFC5*), and all four ligation genes (*ADL1*,
243 *CDC17*, *CDC9*, and *LIG1*). Few genes experienced extensive loss. Of the 11 most commonly
244 lost genes, which were lost in >5% of species, two (*MYH1* and *UVE1*) were lost in the common

245 ancestor of Saccharomycotina, in addition to losses observed in other taxa. The remaining nine
246 genes are unevenly distributed across functions; seven are involved in DNA resynthesis (*CDC27*,
247 *CDM1*, *POL32*, *POLD3*, *POLD4*, *RPA3*, and *SSB3*), one is involved in mismatch recognition
248 (*HSM3*), one is involved in incision (*HNT3*). These findings suggest that genes in the MMR
249 pathway are well conserved across Ascomycota.

250

251 A comparison of our results with those reported in Milo et al. (2019) revealed similar patterns of
252 gene presence and absence. For example, Milo et al. (2019) found that *MYH1* was absent from
253 much of Pezizomycotina, which is consistent with our results. However, we did identify a few
254 differences (inferred losses by Milo et al. (2019) vs. inferred presence in our analyses), which
255 suggest that our pipeline is more conservative in classifying gene losses. We surmise that these
256 differences stem from differences in the gene detection pipelines employed and the divergent
257 objectives of the two studies; Milo et al. (2019) aimed to identify orthologs via a reciprocal best
258 BLAST hit approach, whereas we aimed to identify homologs using pHMMs with absences
259 verified using TBLASTN.

260



261

262 **Fig. 2. Conservation of mismatch repair (MMR) pathway genes across the fungal phylum**
263 **Ascomycota.** MMR genes are generally highly conserved across the phylum. A few model
264 organisms and species of particular interest to medicine and agriculture are labeled as well as a
265 representative species of the faster-evolving *Hanseniaspora* lineage. Genes are colored
266 according to their function; red is recognition, orange is incision, yellow is excision, green is
267 resynthesis, and purple is ligation. Branches are colored by subphylum; budding yeasts /
268 Saccharomycotina (n = 332 species) are in red, fission yeasts / Taphrinomycotina (n = 14
269 species) are in purple, and filamentous fungi / Pezizomycotina (n = 761 species) are in green.
270 Taxon names have been omitted from the phylogeny for visualization purposes; the phylogenetic
271 tree with taxon names can be found in Figure S1 and Shen et al. (2020). The inset phylogenetic
272 tree shows the HLT (in blue) and LLT (in black), with the blue box beneath highlighting them.
273

274 **Extensive loss of MMR genes in a lineage of powdery mildews**

275 Although MMR genes are highly conserved across Ascomycota, we found that a lineage of
276 obligate plant parasite powdery mildews have among the fewest MMR genes of the 1,107
277 Ascomycota species examined. *Erysiphe necator* has lost 10 MMR genes, *Erysiphe pisi* 22, and

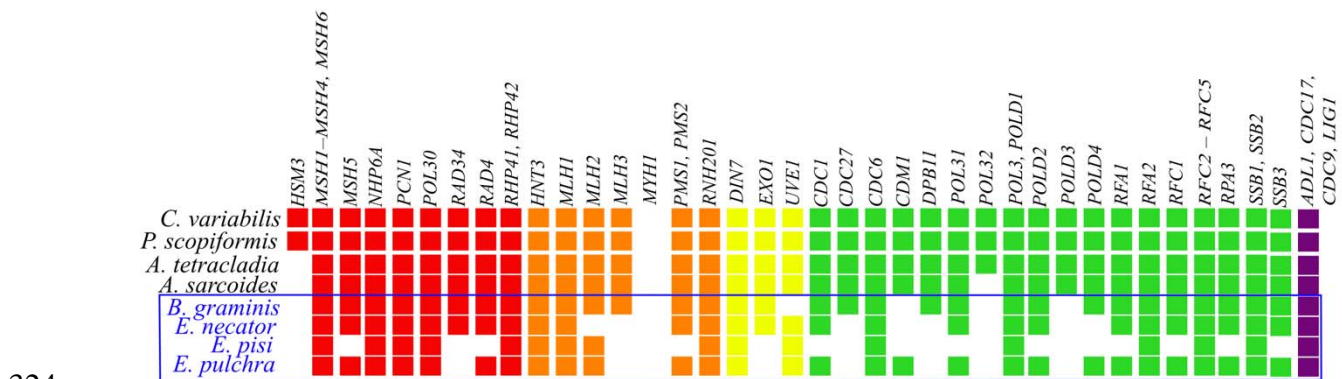
278 *Erysiphe pulchra* 9 (Fig. 3). *E. necator* has been previously documented to have a high rate of
279 genome evolution (Milo et al. 2019) and genomic instability (Jones et al. 2014). *Blumeria*
280 *graminis*, which is sister to the *Erysiphe* genus, has lost 6 MMR genes; previous studies reported
281 extensive gene loss in diverse pathways in this species, generally in genes thought to be
282 unnecessary for their biotrophic lifestyle (Spanu et al. 2010). In contrast, the closely related
283 species *Cairneyella variabilis* and *Phialocephala scopiformis* only lack *MYH1*, an adenine DNA
284 glycosylase that is lost in most filamentous fungi (Chang et al. 2001). In addition to *MYH1*,
285 closely related species *Articulospora tetracladia* and *Ascocoryne sarcoides* lack *HSM3*, which
286 has been lost in almost all Pezizomycotina genomes. *A. sarcoides* is also lacking *POL32*, a DNA
287 polymerase δ subunit, which is part of a larger complex that participates in multiple DNA repair
288 pathways, including nucleotide excision repair and base excision repair (Gerik et al. 1998). Much
289 like the rest of the phylum, genes associated with resynthesis are lost more frequently, but
290 *Erysiphe* and *Blumeria* have lost genes associated with all MMR functions (Table 1) except
291 ligation. In addition, seven of the observed MMR gene losses occur nowhere else in
292 Ascomycota: *EXO1* (excision), *MLH2* (incision), *MLH3* (incision), *MSH5* (recognition), *PMS1*
293 (incision), *PMS2* (incision), and *RFC1* (resynthesis). Taken together, these results suggest that
294 HLT may have a partially functional MMR pathway.

295
296 While select *Erysiphe* taxa have lost more genes than any other species, there are other species
297 with moderate to high levels of MMR gene loss across the phylum. A total of 186 species have
298 lost 6 or more genes across Ascomycota: 4 species in subphylum Taphrinomycotina, 161 in
299 Saccharomycotina, and 21 in Pezizomycotina. The disproportionate number of
300 Saccharomycotina and Taphrinomycotina species is consistent with our knowledge that

301 organisms in these lineages have, on average, a smaller number of DNA repair genes compared
302 to Pezizomycotina (Milo et al. 2019). MMR gene loss in certain Saccharomycotina lineages,
303 such as in some species from the genera *Hanseniaspora* (Steenwyk, Opulente, et al. 2019),
304 *Tetrapisispora*, and *Dipodascus*, is comparable to the loss observed in HLT; however, only 18
305 other species in Ascomycota showed MMR gene loss to the same degree as any *Erysiphe*
306 species. In general, species with elevated levels of gene loss primarily lost genes noted as
307 commonly lost earlier in this paper (see “MMR genes are highly conserved across the fungal
308 phylum Ascomycota”), with occasional losses in other genes.

309
310 There was a striking discrepancy between the presence and absence of MMR genes inferred by
311 pHMM versus TBLASTN in HLT that was not observed in other species. When measured by
312 pHMM, *E. pulchra* lost 42 MMR genes, as opposed to 9 when confirmed with TBLASTN, and
313 *E. necator* and *E. pisi* lost 51 MMR genes, as opposed to 10 and 22, respectively. *B. graminis*
314 lost 9 MMR genes by pHMM and 6 when absences were verified with TBLASTN. In the closely
315 related species *C. variabilis*, *P. scopiformis*, *A. tetracladia*, and *A. sarcoides*, genes deemed
316 absent by HMM also did not have a match for TBLASTN. This is unexpected, given that
317 pHMMs are more sensitive in sequence similarity searches and typically outperform TBLASTN
318 when detecting genes on an evolutionary timescale (Yoon 2009). Given that the inputs in the
319 pHMM-based searches were gene annotation predictions and the inputs in the TBLASTN-based
320 searches were genome assemblies, it is possible that the genes missed by pHMMs are genes that
321 have not been annotated. Many of the TBLASTN hits are not located near a start codon,
322 suggesting that they may have been pseudogenized.

323



324
 325 **Fig. 3: The powdery mildews *Erysiphe* and *Blumeria* have lost many more mismatch repair**
 326 **(MMR) pathway genes than closely related species.** High loss taxa (HLT; shown in blue font)
 327 have lost 6 – 22 MMR genes, while low loss taxa (LLT; shown in black font) and most other
 328 species in the rest of the Ascomycota phylum have lost 1 – 3 genes. Note that the losses of
 329 *EXO1*, *MLH2*, *MLH3*, *MSH5*, *PMS1*, *PMS2*, and *RFC1* are uniquely observed in HLT. Genes
 330 are colored according to their function; red is recognition, orange is incision, yellow is excision,
 331 green is resynthesis, and purple is ligation.
 332

333 High MMR gene loss taxa show increased number and length of microsatellites

334 Examination of microsatellites revealed microsatellite expansions in HLT in comparison to LLT.
 335 Specifically, we found statistically significant increases in the number and length of
 336 microsatellites in HLT compared to LLT (Fig. 4A, Tables 2, 3, S1, and S2). Overall, after
 337 controlling for genome size, HLT had significantly more microsatellites than LLT ($F = 34.83$; p
 338 < 0.001 ; ANOVA; Tables 3 and S1). This effect was driven by the very large increase in the
 339 number of homopolymer runs in *Erysiphe* and *Blumeria* (Fig. 4C) ($p < 0.001$; Tukey HSD; Table
 340 S1). There was no statistically significant difference between the groups in the numbers of
 341 microsatellites with a 2-6 bp motif length (Table S1). HLT showed significantly higher average
 342 microsatellite lengths at every motif size than LLT (Fig. 4B) ($p < 0.01$ for 1 bp, $p < 0.001$ for all
 343 other motif lengths; Wilcoxon rank sum test; Table S2). HLT have an increased number of
 344 monomeric runs (after controlling for genome size) and an increase in length of microsatellites of
 345 all motif lengths, suggesting that the MMR pathway's function is compromised in these species.

346

347 **Table 2. MMR gene loss is associated with an increased number of microsatellites**

Species	Group	Genome Size (Mb)	1 bp	2 bp	3 bp	4 bp	5 bp	6 bp
<i>A. tetracladia</i>	LLT	41.8	306	725	1028	525	288	174
<i>A. sarcoides</i>	LLT	34.3	213	677	601	365	164	75
<i>C. variabilis</i>	LLT	50.7	420	374	336	54	47	69
<i>P. scopiformis</i>	LLT	48.9	325	475	840	301	159	71
<i>B. graminis</i>	HLT	124.5	6379	1047	1752	659	389	190
<i>E. necator</i>	HLT	52.5	4616	2829	1135	1172	290	145
<i>E. pisi</i>	HLT	69.3	4463	1841	1303	521	80	125
<i>E. pulchra</i>	HLT	63.5	2681	969	990	522	89	87

348 Low loss taxa (LLT): *Articulospora tetracladia*, *Ascocoryne sarcoides*, *Cairneyella variabilis*, *Phialocephala*
 349 *scopiformis*

350 High loss taxa (HLT): *Blumeria graminis*, *Erysiphe necator*, *Erysiphe pisi*, *Erysiphe pulchra*

351

352 **Table 3. Statistical analysis shows significant differences in number of microsatellites**
 353 **between high loss taxa (HLT) and low loss taxa (LLT)**

	^a Df	^b Sum Sq	^c Mean Sq	^d F-value	p-value
Group (HLT vs LLT)	1	1.610 x 10 ⁻⁹	1.610 x 10 ⁻⁹	19.40	9.12 x 10 ⁻⁵
Motif length	5	5.849 x 10 ⁻⁹	1.170 x 10 ⁻⁹	14.09	1.19 x 10 ⁻⁷
^e Group x Motif length	5	4.632 x 10 ⁻⁹	9.264 x 10 ⁻⁹	11.16	1.52 x 10 ⁻⁶
Residuals	36	2.989 x 10 ⁻⁹	8.300 x 10 ⁻¹¹	N/A	N/A

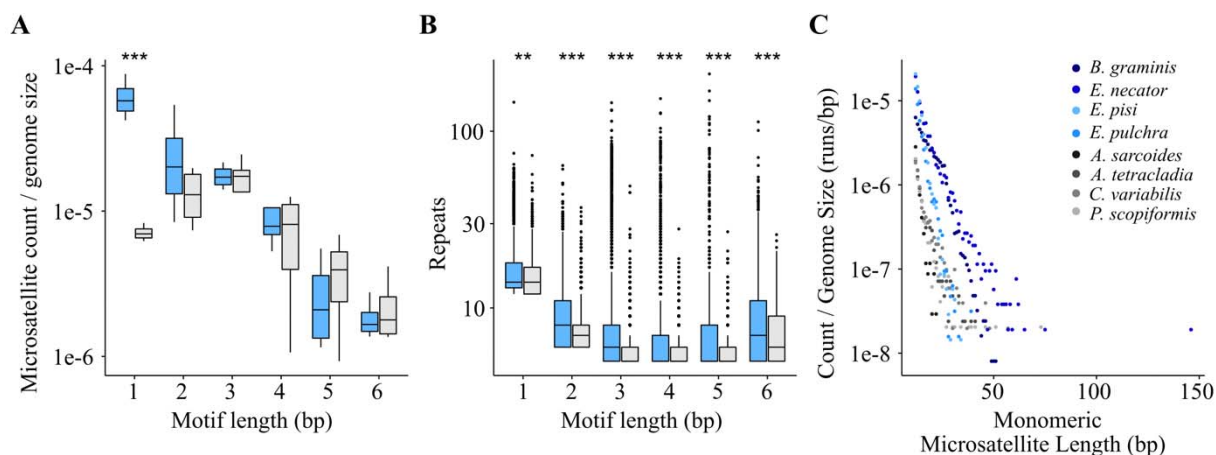
354 ^aDegrees of freedom

355 ^bSum of the squares of the deviations of all the observations from their mean

356 ^cSample variance: sum of squares divided by degrees of freedom

357 ^dTest statistic from ANOVA analysis

358 ^eInteraction between group and motif length



359

360 **Fig. 4: Genomes of high loss taxa (HLT; blue bars) show a proliferation of monomers and**
 361 **an increase in their microsatellite lengths compared to low loss taxa (LLT; grey bars).** (A)

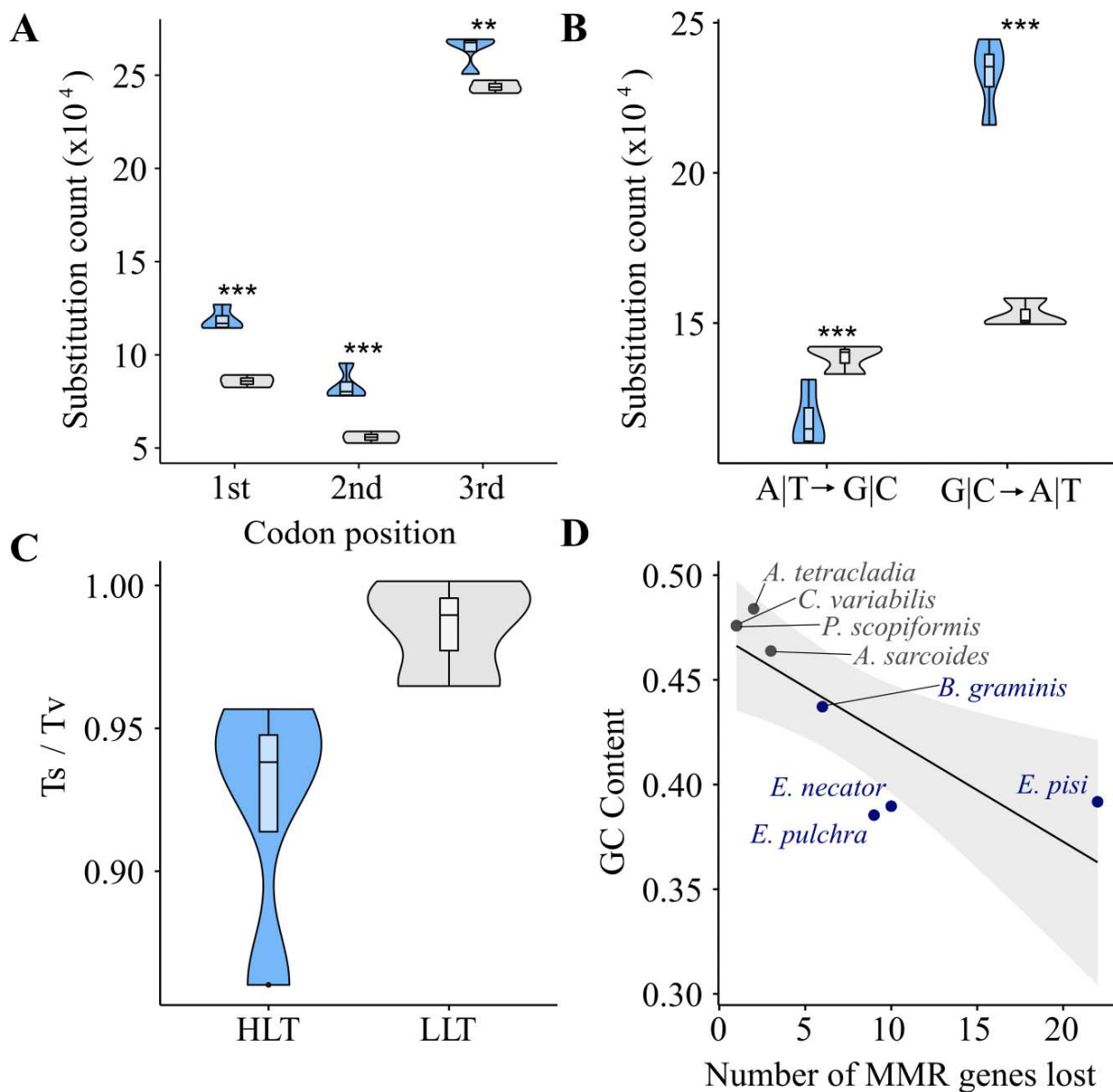
362 Examination of microsatellites in HLT and LLT (gray bars) showed a significant increase in the
 363 number of monomeric runs in HLT ($p < 0.001$; ANOVA, Tukey HSD; Tables 3 and S1). (B)

364 Microsatellites of each motif length are significantly longer in HLT ($p < 0.01$ for 1 bp, $p < 0.001$

365 for all other motif lengths; Wilcoxon rank sum test; Table S2). (C) Monomeric runs are longer
366 and more numerous in HLT than LLT (Tables S1 and S2).
367

368 **High loss taxa show mutational biases**

369 By examining patterns of substitutions among HLT and LLT we found that HLT displayed
370 stronger mutational biases associated with impaired DNA repair pathway function in comparison
371 to LLT. For example, significantly more substitutions were observed at all codon positions in
372 HLT vs. LLT (Fig. 5A) ($p < 0.01$; Tukey HSD; Table S3) and a significant bias towards
373 substitutions in the A|T direction (Fig. 5B) ($p < 0.001$; Tukey HSD; Table S4). HLT also had a
374 lower ratio of transitions to transversions (0.92 ± 0.04) than LLT (0.99 ± 0.02), though this is not
375 statistically significant (Fig. 5C) ($p = 0.06$; Wilcoxon rank sum exact test; Table S5).
376 Additionally, HLT had lower GC content (HLT: $40.10 \pm 0.02\%$ vs. LLT: $47.49 \pm 0.01\%$). Linear
377 regression revealed a significant decrease ($F = 11.7$, $p = 0.01$; Table S6) in GC content of the
378 genomes as the number of MMR genes lost increases (Fig. 5D).



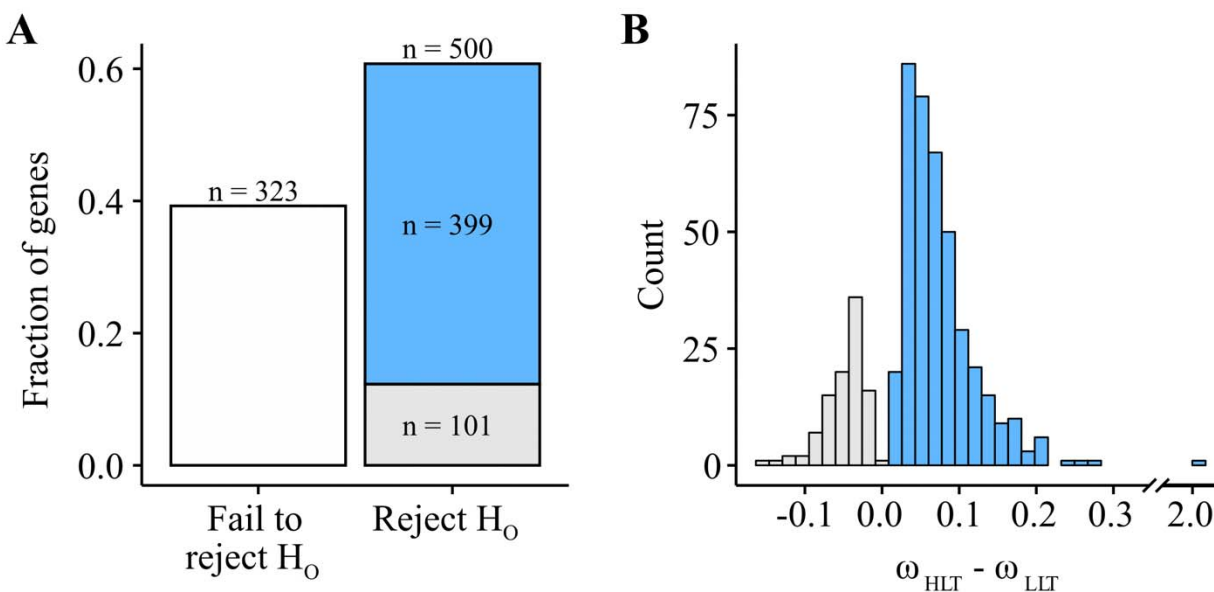
379

380 **Fig. 5: High loss taxa (HLT) show diverse types of mutational bias compared to low loss**
 381 **taxa (LLT).** (A) HLT (blue bars / fonts) show increased counts in base substitution at every
 382 codon position when compared to LLT (gray bars / fonts) ($p < 0.01$; ANOVA, Tukey HSD;
 383 Table S3). (B) HLT show significant mutational bias towards mutation in the A|T direction,
 384 while this trend is not significant in LLT ($p < 0.001$; $p = 0.27$; ANOVA, Tukey HSD; Table S4).
 385 (C) HLT show a decreased ratio of transitions to transversions when compared to LLT, though
 386 this difference is not statistically significant ($p = 0.06$; Table S5). (D) Genome GC content
 387 decreases with increasing MMR gene loss (adjusted $R^2 = 0.6045$; $p = 0.01$; linear regression;
 388 Table S6).

389

390 **High loss taxa have experienced accelerated rates of sequence evolution**

391 To test if the rate of evolution of HLT differed from that of LLT, we performed ω -based branch
392 tests. Our null hypothesis was that all branches of the phylogeny for our selected eight species
393 had the same rate of evolution, while our alternate hypothesis posited that HLT branches,
394 including the branch of their most recent common ancestor, experienced a different rate of
395 sequence evolution than LLT branches. We found that 60.75% of genes rejected the null
396 hypothesis ($\alpha = 0.05$; $n = 500$) and 39.25% failed to reject the null ($n = 323$) (Fig. 6A; File S2).
397 Of the genes which rejected the null hypothesis, 79.80% ($n = 399$) experienced higher rates of
398 substitution in HLT, which constitutes 48.48% of all genes tested (Fig. 6A). Among the genes
399 that rejected the null hypothesis, the difference between the ω values for the HLT (median $\omega =$
400 0.0899) and the LLT (median $\omega = 0.0567$) showed accelerated rates of substitution for HLT
401 branches (Fig. 6B). These results suggest that MMR gene loss is associated with a genome-wide
402 signature of accelerated mutation rates.



403
404 **Fig. 6: Powdery mildew high loss taxa (HLT) show accelerated rates of evolution.** (A) Most
405 (60.75%; $n = 500$) BUSCO genes reject the null hypothesis that HLT branches experience the
406 same rate of substitution as LLT branches. 48.48% of BUSCO genes support a higher rate of
407 evolution for HLT ($n = 399$; in blue), 12.27% support a higher rate of evolution for LLT ($n =$

408 101; in grey), and 39.25% (n = 323) fail to reject the null hypothesis that the rate of substitution
409 is uniform across HLT and LLT branches (in white). Among genes that support the alternative
410 hypothesis, 79.80% (n = 399) support that *Erysiphe* and *Blumeria* evolve more quickly than
411 LLT. (B) Among genes which reject the null hypothesis, the distribution of differences between
412 ω values for HLT and LLT branches show elevated substitution rates in HLT.

413
414

415 **DISCUSSION**

416
417 Using sequence similarity searches, we examined the conservation of the MMR pathway across
418 1,107 ascomycete species. The near universal conservation of the vast majority of MMR genes
419 across the phylum confirms this pathway's known critical role for DNA maintenance (Fukui
420 2010; Schofield & Hsieh 2003; Kunkel & Erie 2005). However, we also discovered that a
421 lineage of *Erysiphe* and *Blumeria* powdery mildews, named HLT, have experienced significant
422 MMR gene loss (Fig. 3). HLT exhibit increases in monomer count, microsatellite length,
423 mutational biases, and rate of evolution (Figs. 4, 5, and 6), suggesting that the function of their
424 MMR pathway may be impaired. While DNA repair mechanisms are present in all eukaryotes
425 and are highly conserved, there is mounting evidence of exceptions to this rule in the fungal
426 kingdom (Steenwyk, Opulente, et al. 2019; Steenwyk 2021). The increased MMR gene absence
427 observed in HLT correlates with changes in the microsatellite compositions of their genomes.
428 The significant difference between HLT and LLT in the number of monomeric runs is consistent
429 with mutational patterns present in human cancers and MMR deficient yeast (Arzimanoglou et
430 al. 1998; Lang et al. 2013). Monomeric runs are the most prone to replication fork slippage and
431 are used to diagnose MSI in tumors (Richman 2015). In addition to an increase in the number of
432 monomeric runs (Fig. 4A), HLT showed significantly longer microsatellites for each motif
433 length than LLT (Fig. 4B), which suggests impaired MMR function and increased replication
434 fork slippage.

435
436 Examination of HLT genomes revealed mutational signatures suggesting that the MMR pathway
437 has been impaired by the observed gene losses. Patterns of substitutions suggest the loss of
438 MMR genes leads to increased substitution rates (Fig. 5A) and lower GC content (Fig. 5D).
439 More specifically, the prominent A|T bias of substitutions in the HLT is likely driven by the
440 known A|T bias of mutations previously observed in bacteria and eukaryotes, including *S.*
441 *cerevisiae* (Hershberg & Petrov 2010; Keightley et al. 2009; Zhu et al. 2014; Lynch 2010).
442 Furthermore, GC content decreases in proportion to the number of MMR genes lost in the HLT
443 and LLT, which was also observed among *Hanseniaspora*, a lineage of budding yeasts that have
444 lost diverse DNA repair genes (Steenwyk, Opulente, et al. 2019). There is no significant
445 difference between the transition to transversion (Ts/Tv) ratios of the HLT (0.92 ± 0.04) and
446 LLT (0.99 ± 0.02), though the trend follows what we would expect for HLT having less efficient
447 DNA repair. Both lineages exhibit near-neutral Ts/Tv ratios (Zhu et al. 2014; Lynch et al. 2008).
448 Examination of ω values suggests that faster rates of sequence evolution in HLT compared to the
449 LLT may be associated with MMR gene loss. Long branches, which reflect more substitutions
450 per site, have been previously reported elsewhere for *E. necator* (Milo et al. 2019), providing
451 independent support to our findings.

452
453 Loss of function in the MMR pathway may be advantageous in certain environments or under
454 certain lifestyles. For example, strains of human pathogens with impaired MMR function are
455 found in environments where antifungal drugs are present. Some of these strains have evolved
456 drug resistance, so the elevated mutation rate generated by MMR gene loss may be adaptive
457 under certain stressful situations (Billmyre et al. 2017, 2020; Rhodes et al. 2017). Loss in

458 separate DNA repair pathways is also present in other parasites and may contribute to elevated
459 mutation rates. Organisms with parasitic lifestyles tend to evolve more rapidly than free-living
460 organisms; while these mechanisms are unknown, previous work has identified that the loss of
461 the classical nonhomologous end joining (C-NHEJ) pathway is common in this lifestyle and may
462 even be a contributing factor (Nenarokova et al. 2019). Previous studies of genome structure in
463 *E. necator* have found genome expansion largely driven by transposable elements and suggest
464 that genome instability, particularly in copy number variants, can mediate rapid evolution of
465 fungicidal resistance (Jones et al. 2014). The evolution of fungicide resistance in powdery
466 mildews has massive implications for agriculture; major crops are impacted by these pathogens
467 and some are able to quickly evolve resistance to antifungal chemicals, with resistance evolution
468 accelerated by increased use (Jones et al. 2014; Vielba-Fernández et al. 2020). More broadly,
469 genome instability among HLT taxa reflects their parasitic lifestyle, which is associated with
470 gene loss and plastic genomic architecture (Schmidt & Panstruga 2011). Gene loss in primary
471 and secondary metabolism, enzymes acting on carbohydrates, and transporters, has been
472 documented in *Blumeria graminis*, as well as massive expansion in retrotransposons and genome
473 size, reflecting extreme genomic changes associated with its parasitic lifestyle (Spanu et al.
474 2010). Given their extreme genomic changes and importance to agriculture, *Blumeria* and
475 *Erysiphe* may be novel models to study the outcome and evolutionary trajectory of sustained loss
476 of MMR pathways.

477

478

479 **Funding**

480 M.A.P. was partially supported by the Vanderbilt Undergraduate Summer Research Program and
481 the Goldberg Family Immersion Fund. J.L.S. and A.R. were supported by the Howard Hughes
482 Medical Institute through the James H. Gilliam Fellowships for Advanced Study Program.
483 A.R.'s laboratory received additional support from the Burroughs Wellcome Fund, the National
484 Science Foundation (DEB-1442113), and the National Institutes of Health/National Institute of
485 Allergy and Infectious Diseases (R56AI146096). X.-X.S. was supported by the National Natural
486 Science Foundation of China (no. 32071665) and the Young Scholar 1000 Talents Plan.

487

488 **Acknowledgements**

489 We thank Qianhui Zheng for performing exploratory analysis on the HLT and LLT genomes.

490

491 **Data availability statement**

492 Supporting statistical analysis, the Ascomycota phylogeny with species names, and 2
493 supplementary data files (MMR gene presence/absence matrix and ω output) are available via
494 figshare at <https://doi.org/10.6084/m9.figshare.14410994.v1>. The data supporting the phylogeny
495 of Ascomycota are available at <https://doi.org/10.6084/m9.figshare.12751736.v4>.

496

497 **References**

498 Arzimanoglou II, Gilbert F, Barber HRK. 1998. Microsatellite instability in human solid tumors.
499 Cancer. 82:1808–1820. doi: [https://doi.org/10.1002/\(SICI\)1097-](https://doi.org/10.1002/(SICI)1097-0142(19980515)82:10<1808::AID-CNCR2>3.0.CO;2-J)
500 [0142\(19980515\)82:10<1808::AID-CNCR2>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0142(19980515)82:10<1808::AID-CNCR2>3.0.CO;2-J).
501 Ashburner M et al. 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. 25:25–
502 29. doi: 10.1038/75556.

503 Beier S, Thiel T, Münch T, Scholz U, Mascher M. 2017. MISA-web: a web server for
504 microsatellite prediction. *Bioinformatics*. 33:2583–2585. doi: 10.1093/bioinformatics/btx198.

505 Billmyre RB, Applen Clancey S, Li LX, Doering TL, Heitman J. 2020. 5-fluorocytosine
506 resistance is associated with hypermutation and alterations in capsule biosynthesis
507 in *Cryptococcus*. *Nat. Commun*. 11:127. doi: 10.1038/s41467-019-13890-z.

508 Billmyre RB, Clancey SA, Heitman J. 2017. Natural mismatch repair mutations mediate
509 phenotypic diversity and drug resistance in *Cryptococcus deuterogattii* Rokas, A, editor. *Elife*.
510 6:e28802. doi: 10.7554/eLife.28802.

511 Boland CR, Goel A. 2010. Microsatellite Instability in Colorectal Cancer. *Gastroenterology*.
512 138:2073-2087.e3. doi: 10.1053/j.gastro.2009.12.064.

513 Campbell BB et al. 2017. Comprehensive Analysis of Hypermutation in Human Cancer. *Cell*.
514 171:1042-1056.e10. doi: 10.1016/j.cell.2017.09.048.

515 Chang D-Y, Gu Y, Lu A-L. 2001. Fission yeast (*Schizosaccharomyces pombe*) cells defective in
516 the MutY-homologous glycosylase activity have a mutator phenotype and are sensitive to
517 hydrogen peroxide. *Mol. Genet. Genomics*. 266:336–342. doi: 10.1007/s004380100567.

518 Cherry JM et al. 2012. *Saccharomyces* Genome Database: the genomics resource of budding
519 yeast. *Nucleic Acids Res*. 40:D700–D705. doi: 10.1093/nar/gkr1029.

520 Cock PJA et al. 2009. Biopython: freely available Python tools for computational molecular
521 biology and bioinformatics. *Bioinformatics*. 25:1422–1423. doi: 10.1093/bioinformatics/btp163.

522 Ellegren H. 2004. Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet*.
523 5:435–445. doi: 10.1038/nrg1348.

524 Friedberg EC et al. 2005. *DNA Repair and Mutagenesis*. 2nd ed. doi: 10.1128/9781555816704.

- 525 Fukui K. 2010. DNA Mismatch Repair in Eukaryotes and Bacteria. *J. Nucleic Acids*. 2010. doi:
526 10.4061/2010/260512.
- 527 Gerik KJ, Li X, Pautz A, Burgers PMJ. 1998. Characterization of the Two Small Subunits of
528 *Saccharomyces cerevisiae* DNA Polymerase δ^* . *J. Biol. Chem.* 273:19747–19755. doi:
529 10.1074/jbc.273.31.19747.
- 530 Giglia-Mari G, Zotter A, Vermeulen W. 2010. DNA damage response. *Cold Spring Harb.*
531 *Perspect. Biol.* 3:a000745–a000745. doi: 10.1101/cshperspect.a000745.
- 532 Hershberg R, Petrov DA. 2010. Evidence That Mutation Is Universally Biased towards AT in
533 Bacteria. *PLOS Genet.* 6:e1001115. doi: 10.1371/journal.pgen.1001115.
- 534 Hsieh P, Zhang Y. 2017. The Devil is in the details for DNA mismatch repair. *Proc. Natl. Acad.*
535 *Sci.* 114:3552 LP – 3554. doi: 10.1073/pnas.1702747114.
- 536 Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative
537 HMM search procedure. *BMC Bioinformatics.* 11:431. doi: 10.1186/1471-2105-11-431.
- 538 Jones L et al. 2014. Adaptive genomic structural variation in the grape powdery mildew
539 pathogen, *Erysiphe necator*. *BMC Genomics.* 15:1081. doi: 10.1186/1471-2164-15-1081.
- 540 Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids*
541 *Res.* 28:27–30. doi: 10.1093/nar/28.1.27.
- 542 Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple
543 sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066. doi:
544 10.1093/nar/gkf436.
- 545 Keightley PD et al. 2009. Analysis of the genome sequences of three *Drosophila melanogaster*
546 spontaneous mutation accumulation lines. *Genome Res.* 19:1195–1201. doi:
547 10.1101/gr.091231.109.

548 Kriventseva E V et al. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal,
549 protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs.
550 *Nucleic Acids Res.* 47:D807–D811. doi: 10.1093/nar/gky1053.

551 Kunkel TA, Erie DA. 2005. DNA MISMATCH REPAIR. *Annu. Rev. Biochem.* 74:681–710.
552 doi: 10.1146/annurev.biochem.74.082803.133243.

553 Lang GI, Parsons L, Gammie AE. 2013. Mutation Rates, Spectra, and Genome-Wide
554 Distribution of Spontaneous Mutations in Mismatch Repair Deficient Yeast. *G3*
555 *Genes|Genomes|Genetics.* 3:1453 LP – 1465. doi: 10.1534/g3.113.006429.

556 Lau A, Blitzblau H, Bell SP. 2002. Cell-cycle control of the establishment of mating-type
557 silencing in *S. cerevisiae*. *Genes Dev.* 16:2935–2945. doi: 10.1101/gad.764102.

558 Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
559 developments. *Nucleic Acids Res.* 47:W256–W259. doi: 10.1093/nar/gkz239.

560 Li H et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 25:2078–
561 2079. doi: 10.1093/bioinformatics/btp352.

562 Lock A et al. 2019. PomBase 2018: user-driven reimplementaion of the fission yeast database
563 provides rapid and intuitive access to diverse, interconnected information. *Nucleic Acids Res.*
564 47:D821–D827. doi: 10.1093/nar/gky961.

565 Lynch M et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast.
566 *Proc. Natl. Acad. Sci.* 105:9272 LP – 9277. doi: 10.1073/pnas.0803466105.

567 Lynch M. 2010. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl.*
568 *Acad. Sci. U. S. A.* 107:961–968. doi: 10.1073/pnas.0912629107.

569 Marti TM, Kunz C, Fleck O. 2002. DNA mismatch repair and mutation avoidance pathways. *J.*
570 *Cell. Physiol.* 191:28–41. doi: <https://doi.org/10.1002/jcp.10077>.

- 571 Milo S, Harari-Misgav R, Hazkani-Covo E, Covo S. 2019. Limited DNA Repair Gene
572 Repertoire in Ascomycete Yeast Revealed by Comparative Genomics. *Genome Biol. Evol.*
573 11:3409–3423. doi: 10.1093/gbe/evz242.
- 574 Nenarokova A et al. 2019. Causes and Effects of Loss of Classical Nonhomologous End Joining
575 Pathway in Parasitic Eukaryotes Heitman Greg Ouellette, Marc, JM, editor. *MBio.* 10:e01541-
576 19. doi: 10.1128/mBio.01541-19.
- 577 Pearson WR. 2013. An Introduction to Sequence Similarity (“Homology”) Searching. *Curr.*
578 *Protoc. Bioinforma.* 42:3.1.1-3.1.8. doi: 10.1002/0471250953.bi0301s42.
- 579 Dos Reis TF et al. 2019. The *Aspergillus fumigatus* Mismatch Repair MSH2 Homolog Is
580 Important for Virulence and Azole Resistance. *mSphere.* 4:e00416-19. doi:
581 10.1128/mSphere.00416-19.
- 582 Rhodes J et al. 2017. A Population Genomics Approach to Assessing the Genetic Basis of
583 Within-Host Microevolution Underlying Recurrent Cryptococcal Meningitis Infection. *G3*
584 *Genes|Genomes|Genetics.* 7:1165 LP – 1176. doi: 10.1534/g3.116.037499.
- 585 Richman, S. (2015). Deficient mismatch repair: Read all about it (Review). *Int J Oncol*, 47(4),
586 1189–1202. doi: 10.3892/ijo.2015.3119.
- 587 Schmidt SM, Panstruga R. 2011. Pathogenomics of fungal plant parasites: what have we learnt
588 about pathogenesis? *Curr. Opin. Plant Biol.* 14:392–399. doi: 10.1016/j.pbi.2011.03.006.
- 589 Schofield MJ, Hsieh P. 2003. DNA Mismatch Repair: Molecular Mechanisms and Biological
590 Function. *Annu. Rev. Microbiol.* 57:579–608. doi: 10.1146/annurev.micro.57.030502.090847.
- 591 Shen X-X et al. 2020. Genome-scale phylogeny and contrasting modes of genome evolution in
592 the fungal phylum Ascomycota. *Sci. Adv.* 6:eabd0079. doi: 10.1126/sciadv.abd0079.
- 593 Spanu PD et al. 2010. Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal

- 594 Tradeoffs in Extreme Parasitism. *Science* (80-.). 330:1543 LP – 1546. doi:
595 10.1126/science.1194573.
- 596 Steenwyk JL. 2021. Evolutionary Divergence in DNA Damage Responses among Fungi. *MBio*.
597 12:e03348-20. doi: 10.1128/mBio.03348-20.
- 598 Steenwyk JL, Opulente DA, et al. 2019. Extensive loss of cell-cycle and DNA repair genes in an
599 ancient lineage of bipolar budding yeasts. *PLOS Biol*. 17:e3000255.
600 <https://doi.org/10.1371/journal.pbio.3000255>.
- 601 Steenwyk JL. 2020. JLSteenwyk/ggpubfigs: first release of ggpubfigs. doi:
602 10.5281/ZENODO.4126988.
- 603 Steenwyk JL et al. 2020. PhyKIT: a UNIX shell toolkit for processing and analyzing
604 phylogenomic data. *bioRxiv*. 2020.10.27.358143. doi: 10.1101/2020.10.27.358143.
- 605 Steenwyk JL, Shen X-X, Lind AL, Goldman GH, Rokas A. 2019. A Robust Phylogenomic Time
606 Tree for Biotechnologically and Medically Important Fungi in the Genera *Aspergillus* and
607 *Penicillium*; Boyle, JP, editor. *MBio*. 10:e00925-19. doi: 10.1128/mBio.00925-19.
- 608 Surtees JA, Alani E. 2004. Replication Factors License Exonuclease I in Mismatch Repair. *Mol*.
609 *Cell*. 15:164–166. doi: 10.1016/j.molcel.2004.07.004.
- 610 Temnykh S et al. 2001. Computational and Experimental Analysis of Microsatellites in Rice
611 (*Oryza sativa* L.): Frequency, Length Variation, Transposon Associations, and Genetic Marker
612 Potential. *Genome Res*. 11:1441–1452. doi: [10.1101/gr.184001](https://doi.org/10.1101/gr.184001).
- 613 The Gene Ontology Consortium. 2019. The Gene Ontology Resource: 20 years and still GOing
614 strong. *Nucleic Acids Res*. 47:D330–D338. doi: 10.1093/nar/gky1055.
- 615 Umar A et al. 1996. Requirement for PCNA in DNA Mismatch Repair at a Step Preceding DNA
616 Resynthesis. *Cell*. 87:65–73. doi: 10.1016/S0092-8674(00)81323-9.

617 Vielba-Fernández A et al. 2020. Fungicide Resistance in Powdery Mildew Fungi. *Microorg.* . 8.
618 doi: 10.3390/microorganisms8091431.

619 Waterhouse RM et al. 2018. BUSCO Applications from Quality Assessments to Gene Prediction
620 and Phylogenomics. *Mol. Biol. Evol.* 35:543–548. doi: 10.1093/molbev/msx319.

621 Wickerham H. 2016. ggplot2: Elegant Graphics for Data Analysis. <https://ggplot2.tidyverse.org>.

622 Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.*
623 24:1586–1591. doi: 10.1093/molbev/msm088.

624 Yoon B-J. 2009. Hidden Markov Models and their Applications in Biological Sequence
625 Analysis. *Curr. Genomics.* 10:402–415. doi: 10.2174/138920209789177575.

626 Zhu YO, Siegal ML, Hall DW, Petrov DA. 2014. Precise estimates of mutation rate and
627 spectrum in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 111:E2310–E2318. doi:
628 10.1073/pnas.1323011111.

629