Article (Discovery section)

Functional Bias and Demographic History Obscure Patterns of Selection among Single-Copy Genes in a Fungal Species Complex

Santiago Sánchez-Ramírez[1,2,*], Jean-Marc Moncalvo[1,3,4]

[1] Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto ON, M5S 3B2, Canada

* **Corresponding author**: E-mail: santiago.snchez@gmail.com, santiago.sanchez@mail.utoronto.ca

[2] Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto ON, M5S 2C6, Canada

[4] Centre for the Analysis of Genome Evolution and Function, University of Toronto, 25 Willcocks Street, Toronto ON, M5S 3B2, Canada

Running title: Polymorphism and divergence in *Amanita*

Abstract

Many different evolutionary processes may be responsible for explaining natural variation within genomes, some of which include natural selection at the molecular level and changes in population size. Fungi are highly adaptable organisms, and their relatively small genomes and short generation times make them pliable for evolutionary genomic studies. However, adaptation in wild populations has been relatively less documented compared to experimental or clinical studies. Here, we analyzed DNA sequences from 502 putative single-copy orthologous genes in 63 samples that represent seven

recently diverged North American *Amanita* (*jacksonii*-complex) lineages. For each gene and each species, we measured the genealogical sorting index (*gsi*) and infinite-site-based summary statistics, such as $\hat{\pi}$, $\hat{\theta}_w$, and D_{Taj} in coding and intron regions. MKT-based approaches and likelihood-ratio-test K_n/K_s models were used to measure natural selection in all coding sequences. Multi-locus (Extended) Bayesian Skyline Plots (eBSP) were used to model intraspecific demographic changes through time based on unlinked, putative neutral regions (introns). Most genes show evidence of long-term purifying selection, likely reflecting a functional bias implicit in single-copy genes. We find that two species have strongly negatively skewed Tajima's D, while three other have a positive skew, corresponding well with patterns of demographic expansion and contraction. Standard MKT analyses resulted in a high incidence of near-zero α with a tendency towards negative values. In contrast, α estimates based on the distribution of fitness effects (DFE), which accounts for demographic effects and slightly deleterious mutations, suggest a higher proportion of sites fixed by positive selection. The difference was more notorious in species with expansion signatures or with historically low population sizes, evidencing the concealing effects of specific demographic histories. Finally, we attempt to mitigate Gene Ontology term overrepresentation, highlighting the potential adaptive or ecological roles of some genes under positive selection.

Key words: *Amanita jacksonii*, comparative population genomics, ectomycorrhizal, exon-targeted sequencing, Extended Bayesian Skyline Plot, molecular evolution

Introduction

Understanding the nature of genomic variation can help illustrate how natural selection and demography balance out throughout the history of a species. The neutral theory of molecular evolution (Kimura 1986; Ohta 1992) postulates that most mutations contributing substantially to the

genetic pool of a population will be either neutral or nearly so. Behind this reasoning is the assumption that changes that alter the amino acid composition of a protein will tend to be deleterious and selected against, while those that are advantageous will be rare. In this framework, one can test whether DNA variation at a given locus fits the expectations under the neutral equilibrium model (Tajima 1989; Fu and Li 1993, 1999). These models primarily allow for explicit testing of neutrality departures, however different kinds of evolutionary processes, such as selective sweeps, recombination, and demographic changes, can mimic similar results (Bachtrog and Andolfatto 2006; Ramírez-Soriano et al. 2008). Nonetheless, selection is likely to affect specific regions through the fixation of linked neutral variation –a process known as hitchhiking or selective sweeps–, while demographic changes are more likely to have genome-wide effects (Andolfatto 2001; Nielsen 2001).

Natural selection at the molecular level is detectable in protein-coding DNA. This can be accomplished by looking at the relationship between functional (amino acid-changing or non-synonymous) and silent (neutral or synonymous) variation. A widely use test of selection is the McDonald-Kreitman test (MKT, McDonald and Kreitman 1991), which is based on the comparison between the ratio of functional (π_n) and silent (π_s) mutations within a species (polymorphism) and the ratio of fixed K_n and K_s mutations between species (divergence). If both polymorphism and divergence were solely driven by mutation and drift, one would expect their ratio to equal one. This model assumes that all K_s mutations are neutral and that K_n mutations are either strongly deleterious, neutral or strongly advantageous (Smith and Eyre-Walker 2002). In this framework, advantageous mutations are more likely to contribute to divergence due to their high fixation probability, while deleterious mutations are likely to be found at lower frequencies due to the effects of purifying selection. On the other hand, the vast majority of segregating mutations are likely to be neutral, as they will tend to persist in the population for longer periods until they eventually go to fixation by

73 random genetic drift (Sella et al. 2009). However, the proportion of adaptive mutations can be under-
74 or overestimated under certain demographic scenarios combined with weak selection (Smith and
75 Eyre-Walker 2002; Eyre-Walker 2006; Messer and Petrov 2013a). For instance, in species with
76 reduced population sizes, slightly deleterious mutations are effectively neutral, thus they are at higher
77 frequency in the population resulting in an underestimation of the proportion of adaptive mutations.
78 Similarly, a recent population expansion will tend to decrease polymorphism, confounding sweep-
79 driven positive selection (Nielsen 2001; Nielsen et al. 2005; Wright and Gaut 2005). Furthermore,
80 another approach is to quantify selection directly as the ratio of substitution rates at K_n and K_s sites
81 (Kimura 1977; Huges and Nei 1988; Goldman and Yang 1994). In a neutral scenario, K_n/K_s or ω
82 should approximate 1, with values above indicating diversifying (positive) selection, and values
83 below indicating purifying (negative) selection (Nielsen 1997; Yang and Nielsen 2000; Nielsen
84 2001).
85
86 Detecting genome-wide patterns of selection and demography is now feasible through the growing
87 availability and accessibility of whole-genome data; the field now known as population genomics
88 (Charlesworth 2010). Most studies addressing questions such as genome-wide effects of selection
89 have usually focus on a number of model organisms (Hough et al. 2013; Ellegren 2014), most of
90 which have good reference genomes and well-documented annotations, while for the vast majority of
91 species, empirical evidence is still lacking (Cutter and Payseur 2013). Nonetheless, lowering
92 sequencing costs and more efficient bioinformatics tools are making these types of approaches
93 accessible to non-model organisms (Ekblom and Galindo 2011; Aguileta et al. 2010).
94
95 Fungi play multiple essential roles in the environment, mainly as decomposers of organic matter, but
96 also as pathogens, commensals, and mutualists. Not only due to their diverse ecology (Selbmann et

al. 2013), but also because of their genetic machinery (Anderson et al. 1992; Schoustra et al. 2007), fungi are regarded as highly adaptable organisms. Most evidence comes from experimental studies in yeasts (Suutari et al. 1990; Davies et al. 1995; Piper et al. 2001; Liu 2006; Dettman et al. 2007; Anderson et al. 2010; Gerstein et al. 2014), drug resistance (Kontoyiannis and Lewis 2002; Anderson 2005), and the evolution of virulence in pathogens (Bentrup and Russell 2001; Becher et al. 2010; Fisher et al. 2012). In spite of being great candidates (mostly due to relatively smaller genomes and short generation times) to study evolutionary genomics (Gladieux et al. 2014), adaptation in wild fungal populations and species has been scarcely explored. However, recent work in few fungal species has evidenced adaptive mechanisms molecularly and experimentally in the wild. For instance, in *Neurospora*, environmental factors such as temperature and latitude are contributing components of genome divergence and adaptation in wild populations (Ellison et al. 2011). Moreover, Aguileta et al. (2012) and Gladieux et al. (2013) highlight the importance of selection and molecular adaptation during host switches and host specialization in *Botrys* and *Microbotryum*. Finally, Branco et al. (2015) explores patterns of genome divergence in two populations of the ectomycorrhizal (EM) –root-associated symbionts– fungus *Suillus* adapted coastal and montane environments. These studies are few recent examples that evidence genome-wide adaptive mechanisms in wild fungal populations.

The genus *Amanita* comprises 500-1000 species (Tulloss 2005), most of which live as EM symbionts (Wolfe et al. 2012). The Caesar's mushrooms (sect. *Caesareae*) form a clade of edible EM *Amanita* distributed worldwide, and are particularly diverse in North America (Sánchez-Ramírez et al. 2015a). With its eight to seven species, the *A. jacksonii* complex is an example of one of two continental Plio-Pleistocene radiations, leading to variable demographic histories and geographic ranges (Sánchez-Ramírez et al. 2015b; Fig. 1). Here we sought to disentangle demographic from adaptive

processes in a multi-species comparative framework, mining a data set of 502 putative-single-copy genes produced by exon-target sequencing. We also explore functional enrichments of genes under selection and speculate about their role in adaptive and ecological processes.

Results

Targets, sequence data, and bioinformatics

Reciprocal BLAST hits between predicted proteomes in the draft genomes of *A. jacksonii* and *A. basii* identified 3,427 putative single-copy candidate genes. The total number of structural genes compared was of 8,511 in *A. jacksonii* and 5,878 in *A. basii*. From the set of 3,427, we selected genes based on a criterion of a stretch of 60 bp of conserved DNA and a minimum number of four probes per gene. Most genes exceeded the minimum number criterion by having 7 to 8 non-overlapping probes. This resulted in 3,887 probes hybridizing to 502 genes.

The total number of reads per sample ranged between 1,787,146 and 24,316,811, with a mean of 9,157,472. The percentage of on-target reads per sample ranged between 56.6% and 96.3%, with an average of 81%. The average number of on-target reads per sample varied between 3,078.6 and 45,760.5 per gene, with a mean of 14,720 across samples. All 502 genes were recovered for all 63 samples, except for samples RET 109-4 and FCME12243 where one gene was not recovered, and sample FCME5528 where two genes were not recovered.

The genes ranged in size from 1,136 to 13,720 bp, with an average length of 4,950 bp. Introns were smaller in size than exons [coding DNA (CDS)], with an average of 68.56 compared to 258.69 bp in CDS. Each gene had about 14 introns and 15 CDS on average. Eighty percent of the data corresponded to CDS, while only 20% corresponded to introns. In total, the data represented

2,483,742 nucleotide sites. We measured the amount of missing data in the alignments given that low mapping quality regions and/or regions with no coverage resulted in missing data (e.g. Ns). We find that both the amount of missing data and the number of variable sites (across all 7 species) is correlated with gene size (Supplementary Fig. S1).

Genealogical sorting and reciprocal monophyly

In order to measure how well genes supported previous species delimitations (Sánchez-Ramírez et al. 2015b) and to assess the relative contribution of phylogenetic conflict (e.g. due to introgression and/or ancestral polymorphism) to polymorphism patterns, we constructed Bayesian trees and calculated the genealogical sorting index (*gsi*, Cummings et al. 2008). The *gsi* measures the degree of co-ancestry within labeled groups and cannot distinguish between introgression and ancestral polymorphism (i.e. incomplete lineage sorting). However, it provides an overall view of how consistent gene trees are with species divergences (the index goes from 1 to 0). Measures of *gsi* indicate that the vast majority of gene trees support the monophyly of at least six species, which all have median *gsi* values above 0.8 (Fig. S2 and S3 in Online Supplementary Data). In contrast, species *A.* sp-jack5 and *A.* sp-jack6, delimited as distinct species in Sánchez-Ramírez et al. 2015b, had median *gsi* values between 0.6 and 0.4 suggesting some degree of phylogenomic conflict between them. When considered as a single unit the *gsi* value increased to 0.82 (Fig. S3 in Online Supplementary Data). For this reason, data from *A.* sp-jack5 and *A.* sp-jack6 were hereafter regarded as *A.* sp-jack5/6 and considered in the following analyses as a single species.

DNA polymorphism

The highest $\hat{\pi}_s$ was found in *A.* sp-F11 with a mean of 0.0119, while the lowest mean, 0.0029, was found in *A.* sp-jack3 –a 3 order magnitude difference (Table 1). As expected due to selective

169 constrain, $\hat{\pi}_n$ was lower in comparison with mean values ranging from 0.0005 in *A.* sp-jack3 to
170 0.0013 in *A.* sp-F11 and *A.* sp-jack5/6 (Table 1). The mean $\hat{\theta}_s$ ranged from 0.0098 in *A.* sp-jack5/6 to
171 0.0027 in *A.* sp-jack3, while the mean $\hat{\theta}_n$ ranged from 0.0012 in *A. jacksonii* to 0.0005 in *A.* sp-jack3
172 (Table 1). On average, higher polymorphism was observed in synonymous sites compared to intronic
173 sites (Table 1). We find that on average, two species (*A. jacksonii* and *A.* sp-jack5/6) had a negatively
174 skewed Tajima's D (*D*), in particular *A. jacksonii*; two other species (*A.* sp-jack2 and *A.* sp-jack3) had
175 average *D* values close to 0; and three other (*A.* sp-jack1, *A.* sp-T31, and *A.* sp-F11) had positively
176 skewed D_{Taj}values (Fig. 2a). D_n was on average lower than D_s. In contrast, Fu and Way's H (Fu and
177 Way 2000), which measures the skewness of the derived (polarized by ourgroup) site-frequency
178 spectrum (SFS), was lower in synonymous sites than in non-synonymous sites (Table 1). Due to
179 potential bias in frequency-based estimators of polymorphism, such as Tajima's *D*, caused by
180 ancestral polymorphism and/or gene flow (Städler et al. 2009; Cutter and Choi 2010; Cutter et al.
181 2012), we looked at the relationship between *gsi* values and *D*. With the expectation that a strong
182 effect would show a linear relationship between low *gsi* values and *D*, we found that genes with
183 significant amounts of within-species phylogenetic conflict do not skew estimates compared to those
184 that fully support monophyly within species (*gsi* = 1; Fig. S4 Online Supplementary Data).
185
186 *Divergence and rates of adaptive evolution*
187 Divergence (*K*) consistently varied between species in the order of 10% for synonymous sites and 2%
188 for non-synonymous sites (Table 1). The ratios between $\hat{\pi}_n$and $\hat{\pi}_s$ ($\hat{\pi}_n/\hat{\pi}_s$) and $K_n$and K_s (K_n/K_s)
189 were quite close to each other, both within and between species (Fig. 2b). This resulted in neutrality
190 index (NI) values that ranged on average from 0.977 in *A.* sp-F11 to 1.614 in *A.* sp-jack3, and α
191 values close to zero with a tendency of negative values (Fig. 2c). Negative α values can be caused

Table 1. Polymorphism, divergence, and codon usage bias estimations for 502 genes in the *A. jacksonii* complex. N, sample size; *S*, segregating sites; π, nucleotide diversity; θ, Watterson's theta; *H*, Fay and Wu's *H*; NI, neutrality index; CAI, codon adaptation index. Sub-indices 'n', 's', and 'int' represent non-synonymous, synonymous, and intronic sites.

	A. jacksonii	*A.* sp-jack56	*A.* sp-jack2	*A.* sp-jack3	*A.* sp-jack1	*A.* sp-T31	*A.* sp-F11
N	44	26	18	14	8	8	8
S_n	16 (3–39)	17 (4–39)	12 (1–29)	5 (0–14)	5 (0–12)	4 (0–11)	9 (1–23)
S_s	25 (6–62)	29 (12–52)	22 (5–52)	7 (1–20)	8 (1–18)	7 (1–18)	19 (4–42)
S_{int}	17 (2–48)	19 (3–45)	15 (1–41)	5 (0–15)	5 (0–15)	5 (0–14)	13 (1–35)
$\hat{\pi}_n$	0.0008 (0.0001–0.0021)	0.0013 (0–0.0025)	0.001 (0–0.0024)	0.0005 (0.0002–0.0013)	0.0007 (0.0002–0.0015)	0.0006 (0.0003–0.0017)	0.0013 (0.0001–0.0032)
$\hat{\pi}_s$	0.0059 (0.0013–0.0157)	0.0088 (0.0041–0.0154)	0.0081 (0.0021–0.0172)	0.0029 (0.0003–0.0077)	0.0051 (0.0012–0.011)	0.0047 (0.0007–0.0109)	0.0119 (0.0034–0.0232)
$\hat{\pi}_{int}$	0.0034 (0.0005–0.0097)	0.0049 (0.0017–0.0099)	0.0048 (0.0006–0.0121)	0.0018 (0–0.0055)	0.0028 (0–0.0077)	0.0027 (0–0.0074)	0.0067 (0.0014–0.0148)
$\hat{\theta}_n$	0.0012 (0.0003–0.0027)	0.0015 (0.0005–0.0029)	0.0011 (0.0002–0.0023)	0.0005 (0–0.0012)	0.0006 (0–0.0013)	0.0005 (0–0.0013)	0.0011 (0.0002–0.0026)
$\hat{\theta}_s$	0.0073 (0.0022–0.0166)	0.0098 (0.0049–0.0166)	0.0079 (0.0023–0.0161)	0.0027 (0.0003–0.0067)	0.0044 (0.001–0.0094)	0.0038 (0.0006–0.0087)	0.0096 (0.0029–0.019)
$\hat{\theta}_{int}$	0.0042 (0.0009–0.0106)	0.0054 (0.0021–0.0099)	0.0046 (0.0008–0.0111)	0.0016 (0–0.0044)	0.0024 (0–0.0065)	0.0022 (0–0.006)	0.0054 (0.0012–0.0118)
H_n	-0.1198 (-0.6092–0.1929)	-0.0635 (-0.2935–0.1128)	-0.0274 (-0.1759–0.1143)	-0.0589 (-0.3115–0.9992)	0.0137 (-0.1256–1.2172)	-0.0114 (-0.2371–1.2172)	0.0246 (-0.078–0.5348)
H_s	-2.2802 (-5.4425–1.1004)	-0.1197 (-0.3405–0.0735)	-0.0444 (-0.1897–0.0858)	-0.1074 (-0.2942–0.9992)	0.0006 (-0.131–0.0789)	-0.02 (-0.1177–0.0774)	-0.0163 (-0.0539–0.0326)
K_n	0.0215 (0.0077–0.0415)	0.0203 (0.007–0.0404)	0.0216 (0.0079–0.04)	0.0221 (0.0081–0.0404)	0.0201 (0.0076–0.0378)	0.0213 (0.008–0.039)	0.0214 (0.0082–0.0402)
K_s	0.1394 (0.0884–0.1876)	0.1369 (0.0892–0.1887)	0.1439 (0.0927–0.1878)	0.1533 (0.103–0.2056)	0.1508 (0.1011–0.2037)	0.1624 (0.1075–0.2168)	0.1541 (0.1009–0.203)
NI	1.1798 (0.3653–2.6753)	1.185 (0.4161–2.2767)	1.0788 (0.2228–2.6991)	1.614 (0–4.6016)	1.3401 (0–3.9734)	1.271 (0–3.5996)	0.9777 (0.2139–2.2241)
CAI	0.419 (0.2787–0.5453)	0.4167 (0.2794–0.5331)	0.4138 (0.2725–0.5288)	0.4045 (0.2619–0.5255)	0.3973 (0.2674–0.5242)	0.3991 (0.2577–0.5174)	0.4105 (0.2641–0.5274)

196 by strong purifying selection, resulting in $\hat{\pi}_n/\hat{\pi}_s > K_n/K_s$ as deleterious variant will not contribute to

197 divergence, but will be found at lower frequencies as polymorphism (Nielsen 2001; Egea et al. 2008).
198 To explore this possibility, we performed likelihood-ratio-test (LRT) of codon-based substitution
199 models (Yang and Nielsen 2002) between species and found that most genes (278, 55%) fitted a
200 model that supports purifying selection (the nearly-neutral model, M1a, with parameters $\omega = 1$ and
201 $\omega < 1$), while only 10% (48) were supported by a diversifying selection model (the positive selection
202 model, M2a, with parameters $\omega = 1$, $\omega < 1$, and $\omega > 1$). For genes fitting the nearly-neutral and
203 diversifying selection models, the mean ω across sites was 0.23 ± 0.09 and 0.27 ± 0.12, respectively.
204 The rest (176, 35%) did not significantly (0.05 significance level) fit either model, so were deemed to
205 be under stronger purifying selection, given that $\omega \ll 1$ (mean ω 0.18 ± 0.1). Moreover, genes with
206 longer than expected coalescence times (i.e. ancestral polymorphism or introgression) can also lead
207 to a $\hat{\pi}_n/\hat{\pi}_s > K_n/K_s$ effect by increasing the frequency of weakly deleterious variants, or variants

208 under balancing selection (Charlesworth 2009), causing negative α values. However, we find no
209 relationship between *gsi* and α (Fig. S5 Online Supplementary Data). A third possibility is that weak
210 selection at synonymous coding (e.g. codon usage bias) is reducing synonymous relative to non-
211 synonymous diversity, however we find no such pattern (Table 1 and Fig. S6 Online Supplementary
212 Data).
213
214 The MKT is susceptible to violations of the standard neutral model. In particular, demographic
215 change and the segregation of slightly deleterious mutations can bias α estimates unworldly or
216 downwardly (Eyre-Walker 2006; Messer and Petrov 2013a). One way that we accounted for this bias
217 was by excluding low-frequency variants in the standard MKT. The other way was by calculating α
218 and ω using the distribution of fitness effects (DFE: α_{dfe} and ω_α) (Eyre-Walker and Keightley 2007,

2009). Notably, the proportion of adaptive fixations, α_{dfe}, resulted in higher estimations based on the standard MKT test (Fig 3A). We also note higher between species variation, both with respect to α_{dfe} and ω_{α}(Fig. 3). Mean α_{dfe} values ranged from 63% in *A. jacksonii* to 11% in *A.* sp-jack1 (Fig. 3A). Given that results based on the DFE indicate an underestimation of α, it would be conservative to consider genes with $\alpha > 0$ under positive selection. The number of genes with at least one adaptive mutation per 1,000 bp (i.e. $\alpha > 0.1\%$) were 286 in *A.* sp-F11, 280 in *A.* sp-jack2, 240 in *A. jacksonii*, 216 in *A.* sp-jack5/6, 213 in *A.* sp-jack1, 203 in *A.* sp-T31, and 151 in *A.* sp-jack3. Besides α_{dfe} and ω_{α}, the DFE can also be used to estimate the proportions of deleterious mutations under different selection strengths (scaled by the effective population size, $N_e s$). The DFE estimates up to four different categories, namely $N_e s < 1$, $1 < N_e s < 10$, $10 < N_e s < 100$, and $N_e s > 100$. Most species had a higher proportion of mutations with very strong effects (Fig. 4; e.g. $N_e s > 100$), with proportions between 0.4 and 0.8 (except *A.* sp-T31, which had a median lower than 0.4). *Amanita jacksonii* and *A.* sp-T31 had the highest proportion of mutations with intermediate effects ($10 < N_e s < 100$), while categories with $N_e s < 10$ had proportions of less than 0.2 for all species (Fig. 4).

Historical population sizes

We inferred the demographic history of each species using extended Bayesian skyline model (eBSP, Heled and Drummond 2008) implemented in BEAST v1.8.2 (Drummond et al. 2012). We opted for reducing our data set in demographic history estimation for several reasons: (1) it has been shown that 16 informative genes or more should be sufficient to estimate accurate population size histories (Heled and Drummond 2008); (2) some of the genes we sampled are contiguous and possibly linked, biasing the assumption of free recombination between loci; and (3) to improve Bayesian parameter mixing and convergence times, which increases with more data (Suchard and Rambaut 2009).

The eBSP analyses supported at least one demographic change within the 95% posterior density distribution (PDD) in all seven species. The species with the highest mean number of demographic changes was *A.* sp-F11 with 1.37 [1, 2 95% PDD], while the lowest was in *A.* sp-jack1 with 0.3 [0, 1 95% PDD]. Demographic trends of population size expansions were observed in *A. jacksonii*, *A.* sp-jack5/6, *A.* sp-jack2, and *A.* sp-F11, and to some extent in *A.* sp-T31, while *A.* sp-jack3 and *A.* sp-jack1 had a rather constant demographic trend (Fig. 5). The most abrupt demographic expansions were found in *A. jacksonii* and *A.* sp-jack5/6 with a near 6 to 7-fold increase in population size within the last ca. 2 Myr (Fig. 5). Similar estimates were found in the two-phase demographic model in the DFE analysis.

Gene Ontology

Due to functional constrain, single-copy genes typically fall within specific functional groups, for instance, in binding and catalytic activities (Aguileta et al. 2008; Han et al. 2014). We find a similar functional bias (Fig. 6A), where a large proportion included proteins involved with energetic and metabolic processes (ATP binding, GO:0005524) and protein-protein interactions (protein binding, GO:0005515, Fig. 6). To less extent proteins related directly or indirectly with gene expression regulation (i.e. nucleotide [GO:0000166], nucleic acid [GO:0003676], and DNA [GO:0003677] binding; zinc ion binding [GO:0008270]; regulation of transcription, DNA-dependent [GO:0006355]), transport and signaling (i.e. protein phosphorylation [GO:0006468], phosphorylation [GO:0016310], protein kinase activity [GO:0004672, GO:0004712], trans-membrane transport [GO:0055085], protein transport [GO:0015031]), metabolic processes (i.e. oxidation-reduction processes [GO:0055114], oxidoreductase activity [GO:0016491], carbohydrate metabolic process [GO:0005975], hydrolase activity [GO:0016787]), and homeostasis (calcium ion binding

267 [GO:0005509]) were also found in the mix. Nonetheless, we accounted for overrepresented Gene
268 Ontology (GO) terms by adding up the positive selection signal (e.g. α and the proportion of sites
269 with $\omega > 1$) across genes and species and then scaling by the total number of unique GO terms (see
270 Materials and Methods). Some of the GO terms found in genes under strong directional selection
271 included proteins related to transcription activities, cellular transport, and metabolic activities (Fig.
272 6B). Interestingly, GO term composition of genes under diversifying selection varied from that of
273 genes under directional selection (e.g. $\alpha > 0$), finding stronger selection in proteins with functions
274 related to biosynthetic and other metabolic processes (Fig. 6b). From the 474 genes across all species
275 were found with $\alpha > 0$, 17 were found exclusively in *A.* sp-F11, followed by 10 in *A.* sp-T31 and
276 sp_jack5/6, 8 in *A. jacksonii* and *A.* sp-jack2, 7 in *A.* sp-jack1, 6 in *A.* sp-jack5/6. Information about
277 their GO and annotation can be found in Table S1 Online Supplementary Data.
278
279 **Discussion**
280 Interpreting DNA variation from an evolutionary perspective has been a long-standing goal for
281 population geneticists, both doing empirical and theoretical work. The neutral equilibrium theory of
282 molecular evolution (Kimura 1977, 1986) provides a null framework to test hypotheses about natural
283 selection and other processes acting on the genome. While the core of population genomic research
284 has developed around coalescent-based theory in a single –or at least a pair– of species (Hough et al.
285 2013; Charlesworth 2010), there are clear benefits in integrating comparative and phylogenetic data,
286 in multi-species multi-locus assessments (Cutter 2013).
287
288 *The exon-targeted sequence-capture approach*
289 Before and early into the "whole-genome sequencing" era, most population genomics studies relied
290 on PCR and Sanger sequencing of individual genes for data production. Some classic studies

included data ranging from tens to few hundreds of genes, in what were then called "whole-genome" approaches. Exon-targeted sequencing takes advantage of the rapid and massive data production of next-generation sequencing, while focusing on specific genes of interest. Although any class of gene can be potentially targeted, we opted for single-copy genes, attempting to avoid potential spurious DNA hybridizations with paralogues, which can bias diversity estimates. However, targeting single-copy genes has the disadvantage of characterizing genome-wide diversity based on a single "class" of genes, which may not necessarily represent a heterogeneous sample from the genome. Nonetheless, the degree of variation and distribution of the data within each species, in combination with our results, suggests that our sample of 502 genes maybe a good representation genome-wide variation. Furthermore, this method has been proven successful when DNA material is only available in low quantities (Hancock-Hanser et al. 2013), or from degraded samples (Templeton et al. 2013). All of our samples came from dried museum specimens and FTA plant saver cards (Dentinger et al. 2010), which would not reach quantity/quality standards for other popular downstream genomic applications such as whole-genome resequencing or reduced-library representation approaches (e.g. RAD-tag, GBS). In particular, reduced-library representation approaches are practical for sampling genome-wide SNPs, however the shortness of the fragments makes it impossible to conduct codon-based divergence and polymorphism assessments. Therefore, exon-targeting methods are a practical and efficient strategy for scoring many loci and complete structural genes in non-model organisms.

Effects of selection and demography on genomic variation

Studies in flies (Smith and Eyre-Walker 2002; Bierne and Eyre-Walker 2004; Sella et al. 2009), plants (Slotte et al. 2010), and humans (Fay et al. 2001) have suggested that the rate of adaptive substitution can reach high proportions (20–40%), even in the presence of high levels of selective constrain or purifying selection (Williamson et al. 2014; Slotte et al. 2010). The rate of adaptive

315 evolution is thought to be even higher in microorganisms, such as bacteria (Charlesworth and Eyre-
316 Walker 2006). In contrast, in these *Amanita* species, α seems to be consistently underestimated (Fig.
317 2C). A number of different factors can explain potentially low levels of adaptive evolution. Among
318 them, we will discuss how specific evolutionary scenarios such as purifying selection, relaxed
319 selective constrain, and demographic history can contribute to this pattern, highlighting evidence in
320 favor or against such factors.
321
322 Strong selective constraint is expected in protein coding DNA, largely because the majority of amino
323 acid changing mutations are likely to be deleterious (Ohta 1992; Andolfatto 2005). One possible
324 explanation for values of $\alpha \sim 0$ is purifying selection. If non-synonymous substitutions are not being
325 fixed between species, it is likely that they were deleterious early in the evolution of a lineage, and
326 thus purged from the population (Fay and Wu 2003), typically leading to a low ratio of K_n/K_s. Both
327 polymorphism (Table 1) and divergence (Fig. 2B) data overall support high selective constraint in the
328 *A. jacksonii* complex. Similarly, codon LRTs indicate that 90% of the genes favor a model in which
329 the majority of sites are under long-term purifying selection. Most species also resulted with high
330 proportions of new (deleterious) mutations under strong selective effects (e.g. $N_e s > 10$; Fig. 4). We
331 believe that this strong signal of high purifying selection could, in part, be a reflection of a functional
332 bias in single-copy genes (Fig. 6). Many single-copy genes are thought to be under strong selective
333 constrain for duplication, which is also linked to functional constrain (De Smet et al. 2013). Other
334 studies in plants and fungi have characterized single-copy genes under conserved cellular functions,
335 such as binding and metabolism (Aguileta et al. 2008; Han et al. 2014).
336
337 Moreover, mixed genome-wide signals of positive and negative selection are not uncommon
338 (Williamson et al. 2014; Fay et al. 2001), and our results also show evidence of positive selection.

While we find little support for extensive diversifying selection, as only 10% of genes supported model M2a (site class with $\omega > 1$), more genes were found under directional selection within and across species. On average each species had about half the total number of genes (ca. 230 from 502) with $\alpha > 0$, and about 94% (474) of all 502 genes had at least one species with $\alpha > 0$. However, only 25 genes where found with $\alpha > 0$ for all seven species, which is a number closer to the number of genes found under diversifying selection. The diversifying selection model requires selection in favor of allelic diversity, which may rarely happen in genes under functional constraint (Yang et al. 2000). Nonetheless, molecular adaptation is possible, even when $K_n/K_s < 1$ (Kryazhimskiy and Plotkin 2008).

The fate of adaptive and slightly deleterious mutations is highly affected by N_e (Ohta 1992; Eyre-Walker and Keightley 2007; Charlesworth 2009). Standard MKT approaches are generally insensitive to the presence of slightly deleterious mutations and/or demographic changes, which can bias α estimates (Smith and Eyre-Walker 2002; Eyre-Walker and Keightley 2009; Messer and Petrov 2013a). Methods such as the DFE model can both account for slightly deleterious (non-synonymous) mutations that tend to contribute more to polymorphism than to divergence, as well as demographic changes (Eyre-Walker and Keightley 2007, 2009). Our results show that α (standard MKT) is largely underestimated across all species (Fig. 3). This is more apparent in species that have undergone recent population expansions (Fig. 5). Slightly deleterious polymorphisms are more likely to result in underestimation of α in species with low N_e or that have undergone a population contraction (Smith and Eyre-Walker 2002; Charlesworth 2009; Eyre-Walker and Keightley 2009). We note that effectively neutral non-synonymous mutations with $N_e s \approx 1$ are less than 20% in all species (Fig. 4), however, α (standard MKT) is strongly underestimated in species with historically low N_e (Fig. 5).

Untangling the effects of demography and selection on patterns of polymorphism is difficult (Andolfatto 2001). In particular, because certain demographic scenarios, such as population expansions, can result in similar polymorphism footprints. Nevertheless, demographic events are more likely to affect the whole genome rather than individual loci in the case of selection (Stajich and Hahn 2005). We found marked differences in Tajima's *D* between species indicating different levels of skewness in the site frequency spectrum (SFS), in particular, the negative skew found in *A. jacksonii* (Fig. 2a). In this species, a previous study has evidenced an increase in population size linked to northward population expansion (Sánchez-Ramírez et al. 2015b). Here we corroborate with a larger genomic sample that this expansion signature (Fig. 5) is likely to be legitimate, but also note that it may be exacerbated by the effects of positive selection (Fig. 3). Overall, specific patterns of demographic history are consistent with Tajima's *D* values. For instance, species with positive Tajima's *D* values tended to have more constant and smaller population sizes. In comparison, species with negative *D* had expansion signatures and larger population sizes (Fig. 2A, 5). Sánchez-Ramírez et al. (2015b) found expansion signatures in *A.* sp-jack2 and *A.* sp-jack5/6, which are also supported by our results at a higher genomic scale (Fig. 5). Similarly, we find consistent demographic scenarios in *A.* sp-jack3, *A.* sp-jack1, and *A.* sp-T31, as reported in Sánchez-Ramírez et al. (2015b). We find particularly interesting contrasting demographic and molecular evolution patterns in two species: *A. jacksonii* and *A.* sp-F11. On one side, they have opposite site frequency skews. While *A. jacksonii* has tendency for low and high frequency variants (a negative *D* and *H*; Fig. 2a and Table 1), *A.* sp-F11 only has a tendency for intermediate frequency variants (positive *D*; Fig. 2a). This alone could be explained by differences in their demographic history (Fig. 5), where the former has suffered a bottleneck previous to a dramatic expansion northward, while the latter has had a more stable range

386 and demography, with comparatively high N_e (Fig. 1 and 5). While these differences are clear, both
387 species have the highest α_{dfe} values in the complex (Fig. 3), suggesting high proportions of adaptive
388 evolution. One possibility for this apparent contradiction is that, in *A. jacksonii*, population size
389 expansion could have led to the fixation of slightly deleterious mutations, contributing
390 disproportionally to divergence and overestimating α_{dfe} (Eyre-Walker and Keightley 2009). Another
391 explanation is that glacial cycles during the Pleistocene could have acted as fluctuating episodes of
392 selection, which can lead to high levels of adaptive evolution (Huerta-Sanchez et al. 2008; Bell 2009;
393 Gossmann et al. 2014). This group most likely diversified within the last 5 Myr (Fig. 1B; Sánchez-
394 Ramírez et al. 2015b), during a period that is characterized by a series of abrupt temperature
395 oscillations that have deeply affected the genetic diversity of many populations and species (Hewitt
396 1996, 2000, 2004; Hoffmann and Sgrò 2011)
397
398 *Functional bias, adaptive genes, and the EM habit*
399 The genes that were assayed here were not selected for their specific functions. In fact, they were
400 selected due to their single-copy nature, which facilities bait homology and specificity in exon-
401 targeting procedures, and lowers the rate of alignment artifacts. Nonetheless, single-copy genes are
402 known to be functionally constrained (De Smet et al. 2013), which can introduce noise when
403 assessing functional enrichments of genes under selection. In order to reduce the effect of
404 overrepresentation of GO term due to this functional bias, we devised a measure to scale down the
405 amount of selection in overrepresented genes (see Material and Methods). Differences in abundance
406 of GO terms in the overall data set (Fig. 6A) and GO terms found in genes under selection (Fig. 6A,
407 B), suggest that the overrepresentation was mitigated. Overall, protein functions of genes under
408 selection seem to be divided into three general categories; those related to (1) gene
409 transcription/expression activities, (2) signalling and transport, and (3) metabolic oxidative processes

(Fig. 6). Many of these functions coincide with divergent regions in the genome of another EM fungus (*S. brevipes*) that are also thought to be under selection (Branco et al. 2015). We also noted some functions related to carbohydrate metabolism, specifically with catalytic and hydrolytic activities (Fig. 6A, B). These may be compelling because all EM fungi depend on external carbohydrate sources that are generally supplied by their hosts (Nehls et al. 2007). Throughout the evolution of the EM symbiosis, many species have suffered from a convergent loss of decay mechanisms, while retaining moderate gene repertoires with lignocellulitic capabilities (Kohler et al. 2015). Adaptation in proteins related to carbohydrate metabolism, such as hydrolase activity, could allow for a more efficient breakdown of complex polysaccharides during periods of stress and/or low photosynthetic activity. Also worth mentioning, proteins with telomeric template reverse transcriptase activity were found under diversifying selection (Fig. 6C). Studies suggest that the dynamic nature of telomeric and subtelomeric regions are likely to contribute to rapid evolution of fungi during biotic interactions (Aguileta et al. 2009).

Material and Methods

Samples and whole-genome data

Sixty-three dried specimens, representing eight species (Sánchez-Ramírez et al. 2015b) were processed from collections made in various locations in North America, all deposited in fungal herbaria (Table S2). DNA was extracted from ca. 30-50 mg of dried fungal tissue (gill) and extracted using a modified standard CTAB/proteinase K/chorophrom:isoamilic alcohol protocol (van der Nest et al. 2014) . In addition, whole-genome 454 sequences were produced by the Duke Center for Genomic and Computational Biology for a North American species outside the *A. jacksonii* complex –*A. basii*. These sequences were mapped onto the *A. jacksonii* TRTC168611 draft genome (van der

433 Nest et al. 2014), with the purpose of building a whole-genome draft assembly of an outgroup
434 species.
435
436 Identification of single-copy orthologs, probe design, and gene selection
437 Structural genes were predicted in *A. jacksonii* and *A. basii* using AUGUSTUS v3.0.3 (Stanke et al.
438 2006). All CDS in each species were reciprocally aligned against each other using command-line
439 (standalone) BLASTn, creating two files in tabular format (-outfmt 7). Next, a custom perl script
440 (rbh.pl, available from: https://sites.google.com/site/santiagosnchezrmirez/home/software/perl) was
441 used to find reciprocal single-copy hits, putting each pair in a single FASTA file. Each single-copy
442 pair was aligned using MUSCLE v3.6 (Edgar 2004). Probes for target hybridization were designed
443 based on conserved (identical) CDS regions, 60bp long, across both species. Only genes with four or
444 more non-overlapping regions were chosen. Two PCR primers (5'-
445 TAATACGACTCACTATAGGG-3' and 5'-CTATAGTGTCACCTAAATC-3') were added to the 5'
446 and 3' ends of the 60bp probe target, for a total of 100bp per oligonucleotide probe.
447
448 *Sequence capture, library preparation, and sequencing*
449 Custom oligos probes were synthesized by LCScience
450 (http://www.lcsciences.com/applications/genomics/oligomix/) using DNA microchip technology
451 (Gao et al. 2004). The synthesized probes were then PCR amplified as needed for downstream
452 applications. Probes were later transcribed into RNA using MegaScript T7 Kit (Invitrogen, Carlsbad,
453 CA) and Biotin-dUTP to produce labeled oligos. The DNA was sheared and a standard Illumina
454 TruSeq V2 DNA library kit was prepared, multiplexing all 46 samples. Sequence captures (probe-
455 target hybridization) were performed with 200ng of library, 500ng of biotinylated probes, and
456 Steptavidin beads (Invitrogen, Dynabeads M-270, CAT65305). The capture procedure was repeated

457 at least twice. The details of the protocols can be found in Supplementary file 1. All target captures
458 were directly sequenced in a single Illumina Hi-Seq 2500 lane.
459
460 *Bioinformatics*
461 *Read alignments, sequence phasing*, and *quality control*. Sequences were de-multiplexed into
462 individual fastQ files and each one mapped onto the targets' whole-gene sequences from *A. jacksonii*
463 using the Burrows-Wheeler Aligner's (BWA v0.7.12) algorithm BWA-MEM (Li and Durbin 2010).
464 The resulting SAM files were then filtered with *view* in SAMtools v0.1.19 (Li et al. 2009) for off
465 target reads, low quality mapping (< 30), and PCR duplicates, and converted to BAM files. The
466 reference sequences were indexed and the BAM files sorted. We used HapCompass (Aguiar and
467 Istrail 2012), which uses a read-based graph algorithm, to generate haploid sequences from BAM and
468 VCF files. The VCF file was produced by "piping" *mpileup*, bcftools' *view*, calling the "-cg" flags
469 for genotype likelihood computing, and *varFilter* in vcfutils.pl, only keeping SNPs with Fred quality
470 ≥ 20. The SAM files produced by HapCompass were then converted to BAM and sorted. Consensus
471 fastQ files were produced by "piping" *mpileup*, bcftools' *view*, and *vcf2fq* in vcfutils.pl, in a similar
472 way as above. A custom perl script (OrderFromSamtools.pl) was used to convert fastQ files to fastA
473 format and to produce by-gene alignments from by-sample files. A second perl script
474 (getPhasedGenome.pl) parsed those fastA files to produce "a" and "b" allelic variants per haploid
475 sequence. All missing data was scored as Ns, while unsorted polymorphisms were either left as
476 IUPAC ambiguity codes or scored as Ns, depending on the analysis.
477
478 *Phylogenomics and lineage sorting*. We used MrBayes v3.2 (Ronquist et al. 2012) to generate
479 individual gene trees based on the general time reversible (GTR) model and among-site rate
480 heterogeneity modeled as a gamma distribution (+G). For each gene, we ran two parallel runs each

with eight Metropolis-coupled Markov Chain Monte Carlo (MC^3) chains and 1 million generations. We sampled every 100^{th} state and discarded the 10% initial states as burnin. We assessed convergence by making sure that the average standard deviation of split frequencies was ≤ 0.01. Posterior trees were summarized onto a majority-rule consensus tree compatible with clades with frequencies ≥ 0.5. The summary trees were imported into R using *ape* (Paradis et al. 2004) where the genealogical sorting index (*gsi*, Cummings et al. 2008) was calculated. The *gsi* measures, in a range from 0 to 1, the degree of exclusive ancestry in labeled terminal groups, where 1 signifies group monophyly. These groups can represent any type of biological association, such as species or populations.

Analyses

DNA polymorphism and divergence. Within species all sites were bi-allelic. Coding regions were based on the annotation of *A. jacksonii* (van der Nest et al. 2014). In coding regions, sites were separated into synonymous (4-fold, 3-fold, and 2-fold degenerate sites) and nonsynonymous (0-fold degenerate sites) using the method of Nei and Gojobori (1986). Polymorphism was measured as the average number of pairwise nucleotide differences per site between any two random DNA sequences or $\hat{\pi}$ (Nei and Li 1979; Nei 1987), or the Watterson's estimator $\hat{\theta}_w$, which is measured as the proportion of segregating sites S in a sequence of length L divided by the $(n-1)$th harmonic mean number. Both $\hat{\pi}$ and $\hat{\theta}_w$ were estimated using SFS counts based on formulae in Campos et al. (2014). To measure the skewness of SFS we calculated Tajima's *D* (Tajima 1989) and Fay and Wu's *H*, which measures skewness in the derived frequency spectra following Zeng et al. (2006). We measure codon usage bias by calculating the codon adaptation index (CAI, Sharp and Li 1987). Divergence *K* was measured as the Kimura-2-parameter (Kimura 1980) proportion of sites fixed in the ingroup compared to the outgroup (*A. basii*), considering transitions and transversions separately. In order to

505 control for low-frequency ascertainment biases, we excluded segregating sites with a frequency of
506 1/N within species, where N is the total number of sequences per species. All calculations were
507 performed using the script *poly+div_sfs.pl*, available at
508 https://sites.google.com/site/santiagosnchezrmirez/home/software/perl. For comparison, estimations
509 were also computed in DnaSP v5 (Librado and Rozas 2009) and Polymorphorama (Bachtrog and
510 Andolfatto 2006,
511 http://ib.berkeley.edu/labs/bachtrog/data/polyMORPHOrama/polyMORPHOrama.html).
512
513 *McDonald-Kreitman tests*. We estimated the degree of adaptive evolution in CDS for each species by
514 measuring the amount of polymorphism (within species DNA changes) and divergence (fixed DNA
515 changes between species) leading to non-synonymous and synonymous substitutions. We used the
516 McDonald-Kreitman test (MKT) (McDonald and Kreitman 1991) to estimate the Neutrality Index
517 (NI),

$$\frac{p_n / p_s}{d_n / d_s}$$

518 Where p_n and p_s are, respectively, the within-species proportion of non-synonymous and
519 synonymous segregating sites per non-synonymous and synonymous site, and d_n and d_s, the
520 proportion of fixed non-synonymous and synonymous substitutions per non-synonymous and
521 synonymous site. Values closer to 1 indicate neutrality, those close to 0 indicate positive selection,
522 and those > 1 suggest either negative purifying selection or balancing selection. We used the formula
523 derived by (Smith and Eyre-Walker 2002) to estimate the proportion of amino acid substitutions
524 under positive selection:

$$\alpha = 1 - \frac{D_s P_n}{D_n P_s}$$

This model assumes that mutations are either strongly deleterious or neutral, thus model violations can arise in the presence of slightly deleterious mutations. Likewise, demographic changes can also cause α to be under or overestimated (Eyre-Walker 2006). In order, to account for demography and slightly deleterious mutations we used the approach by Eyre-Walker and Keightley (2007, 2009), which uses a maximum likelihood approximation based on the distribution of fitness effects (Keightley and Eyre-Walker 2010; http://www.homepages.ed.ac.uk/pkeightl//dfe_alpha/dfe-alpha-download.html). Here, estimated the derived SFS for each gene and for each species by polarizing alleles using the outgroup (*A. basii*). We first used *est_dfe* on the folded "neutral" and "selected" (e.g. synonymous and non-synonymous sites) SFS to estimate the DFE based on a two-stage demographic model fixing the N_1 to 100 and optimizing N_2. The mean selection coefficient *s* and the β parameter were also optimized using as starting values -0.1 and 0.5 respectively. We then ran *est_alpha_omega* specifying "neutral" and "selected" counts of diverged (fixed) sites. We applied Jukes-Cantor correction for multiple hits when estimating divergence and removed polymorphism contributing to divergence. In order to generate confidence intervals, we re-estimated parameters by bootstrapping 100 times 100 randomly sampled SFS. Every time the 100 SFS were summarized for each allelic bin.

Codon-based model selection. Models that solely rely on K_n/K_s (ω) ratios are also robust and popular methods for detecting natural selection at the molecular level (Nielsen 2001; Yang and Nielsen 2002). Essentially, CDS regions that have an equal number non-synonymous and synonymous mutations are thought to evolve under effective neutrality, having a $\omega = 1$. Deviations from this ratio indicate purifying selection if $\omega < 1$ and diversifying (positive) selection if $\omega > 1$ (Yang 1998; Anisimova et al. 2001). We used likelihood ratio tests (LRT) to discriminate between three codon-based models: (1) a site-wise model with a single ω ratio; (2) a site-wise “nearly neutral” model where there is a proportion of sites p_0 with $\omega_0 < 1$, and a proportion of sites p_1 with $\omega_1 = 1$.

549 This serves as a null model for a third that considers an additional category of sites with p_2 with
550 $\omega_2 > 1$. In the last two models, p_1 and p_2 are fixed as $p_1 = 1 - p_0$ and $p_2 = 1 - p_0 - p_1$,
551 respectively. All models are based on Goldman and Yang's (1994) codon substitution model and
552 were estimated using maximum likelihood in PAML v4.8a (Yang 2007) using *codeml*. Because ω
553 ratios assume that mutations are fixed differences between species and because population
554 polymorphisms may potentially bias ω estimates (Kryazhimskiy and Plotkin 2008), we selected only
555 one sequence per species per gene.
556
557 *Historical demography*. To infer the demographic history of each species, we applied the multi-locus
558 eBSP model (Heled and Drummond 2008) using the Kimura-2-parameter (K80) model with gamma
559 distribution as the substitution model, implemented in BEAST v1.8.2 (Drummond et al. 2012). As
560 data, we selected 29 intron regions that complied to a series of filters, namely a length size of at least
561 400 bp and locations in different linkage groups. This last filter was done to ensure that loci are not in
562 linkage disequilibrium, as the eBSP model assumes free between loci recombination. In each species,
563 we included an alignment the translation elongation factor 1 alpha (*tef1*), which has been used as
564 molecular clock proxy with a substitution rate of 0.00194 substitution/site/Myr (Sánchez-Ramírez et
565 al. 2015b). The substitution rate in the other genes is scaled based on the fixed rate. Each of the eBSP
566 analyses ran for 100 million generations, with a sampling rate of 10%, and a burnin of 10%. We
567 assessed convergence and mixing by looking at likelihood per generation trace plots and effective
568 sample size values in Tracer 1.6 (Rambaut et al. 2013). Population size values in eBSP were
569 produced directly from the analyses in CSV format.
570
571 *Structural and functional annotations*. Whole-gene sequences from *A. jacksonii* were translated into
572 amino acids and then imported into Blast2GO (Conesa et al. 2005), which was used as an annotation

tool. The main amino acid sequence annotation was taken from the non-redundant (nr) protein database in NCBI using BLASTp, only keeping the best 10 matches. In addition, we ran InterProScan v5 (Zdobnov and Apweiler 2001) to identify functional domains, from which we extracted GO terms. For visualization purposes, we used advanced "word clouds" in http://www.wordle.net/advanced, where scaled GO terms were laid out. Molecular function, biological processes, and cellular component, were marked by different colors. For the scaling proportion, we used:

$$\sum_{x=1}^{x}\sum_{k=1}^{k}\alpha \Big/ x$$

where α is the value of the amount of adaptive evolution (α or P_ω) summed over k species which have found $\alpha > 0$ and summed over x times a given GO term was found. In the case of α k and x can be minimaly 1, however for P_ω, which is the proportion of sites with $\omega > 1$, there is only a single value per gene (i.e. k always equals 1), but x can be > 1.

Acknowledgements

We would like to thank Hernán López-Fernández and Katriina Ilves for sharing their experiences with exon-targeting sequencing, and Dax Torti for providing advice on best library preparation and sequencing practices. We also acknowledge SciNet and CAGEF for providing access to high-performance computing infrastructure. Funding was provided through grants to J.M.M. by the Royal Ontario Museum (ROM) Governors and Natural Science and Engineering Research Council (NSERC) of Canada, and a scholarship by the National Science and Technology Council of Mexico (CONACYT) for doctoral studies to S.S.R.

References

595 Aguiar D, Istrail S. 2012. HapCompass: a fast cycle basis algorithm for accurate haplotype assembly
596 of sequence data. J Comp Biol. 19:577–590.
597 Aguilar A, Garza JC. 2007. Patterns of historical balancing selection on the salmonid major
598 histocompatibility complex class II β gene. J Mol Evol. 65:34–43.
599 Aguileta G, Marthey S, Chiapello H, Lebrun MH, Rodolphe F, Fournier E, et al. 2008. Assessing the
600 performance of single-copy genes for recovering robust phylogenies. Syst Biol. 57:613–627.
601 Aguileta G, Hood M, Refrégier Ga, Giraud T. 2009. Genome evolution in plant pathogenic and
602 symbiotic fungi. Adv Bot Res. 49:151–193.
603 Aguileta G, Lengelle J, Chiapello H, Giraud T, Viaud M, Fournier E, Rodolphe F, Marthey S,
604 Ducasse A, Gendrault A, Poulain J, Wincker P, Gout L. 2012. Genes under positive selection
605 in a model plant pathogenic fungus, *Botrytis*. Infect Genet Evol. 12:987–996.
606 Aguileta G, Lengelle J, Marthey S, Chiapello H, Rodolphe F, Gendrault A, Yockteng R, Vercken E,
607 Devier B, Fontaine MC, Wincker P, Dossat C, Cruaud C, Couloux A, Giraud T. 2010.
608 Finding candidate genes under positive selection in non-model species: examples of genes
609 involved in host specialization in pathogens. Mol Ecol. 19:292–306.
610 Anderson JB. 2005. Evolution of antifungal-drug resistance: mechanisms and pathogen fitness. Nat
611 Rev Microbiol. 3:547–556.
612 Anderson JB, Kohn LM, Leslie JF. 1992. Genetic mechanisms in fungal adaptation. In: The fungal
613 community: its organization and role in the ecosystem. New York: Marcel Dekker. p, 73–98.
614 Anderson JB, Funt J, Thompson DA, Prabhu S, Socha A, Sirjusingh C, Dettman JR, Parreiras L,
615 Guttman DS, Regev A, Kohn LM. 2010. Determinants of divergent adaptation and
616 Dobzhansky-Muller interaction in experimental yeast populations. Cur Biol. 20:1383–1388.
617 Andolfatto P. 2001. Adaptive hitchhiking effects on genome variability. Curr Opin Genet Dev.
618 11:635–641.

619 Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. Nature 437:1149–1152.

620 Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in
621 detecting adaptive molecular evolution. Mol Biol Evol. 18:1585–1592.

622 Bachtrog D, Andolfatto P. 2006. Selection, recombination and demographic history in *Drosophila*
623 *miranda*. Genetics 174:2045–2059.

624 Becher R, Hettwer U, Karlovsky P, Deising HB, Wirsel SGR. 2010. Adaptation of *Fusarium*
625 *graminearum* to tebuconazole yielded descendants diverging for levels of fitness, fungicide
626 resistance, virulence, and mycotoxin production. Phytopathology. 100:444–453.

627 Bell G. 2009. Fluctuating selection: the perpetual renewal of adaptation in variable environments.
628 Phil Trans Royal Soc London (B). 365:87–97.

629 Bentrup KZ, Russell DG. 2001. Mycobacterial persistence: adaptation to a changing environment.
630 Trends Microbiol. 9:597–605.

631 Bierne N, Eyre-Walker A. 2004. The genomic rate of adaptive amino acid substitution in *Drosophila*.
632 Mol Biol Evol. 21:1350–1360.

633 Branco S, Gladieux P, Ellison CE, Kuo A, LaButti K, Lipzen A, Grigoriev IV, Liao HL, Vilgalys R,
634 Peay KG, Taylor JW, Bruns TD. 2015. Genetic isolation between two recently diverged
635 populations of a symbiotic fungus. Mol Ecol. 24:2747–2758.

636 Campos JL, Halligan DL, Haddrill PR, Charlesworth B. 2014. The Relation between recombination
637 rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. Mol Biol
638 Evol. 31:1010–1028.

639 Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and patterns of
640 molecular evolution and variation. Nat Rev Genet. 10:195–205.

641 Charlesworth B. 2010. Molecular population genomics: a short history. Genet Res. 92:397–411.

642 Charlesworth J, Eyre-Walker A. 2006. The rate of adaptive evolution in enteric bacteria. Mol Biol
643 Evol. 23:1348–1356.
644 Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool
645 for annotation, visualization and analysis in functional genomics research. Bioinformatics.
646 21:3674–3676.
647 Cummings MP, Neel MC, Shaw KL. 2008. A genealogical approach to quantifying lineage
648 divergence. Evolution. 62:2411–2422.
649 Cutter AD. 2013. Integrating phylogenetics, phylogeography and population genetics through
650 genomes and evolutionary theory. Mol Phylogenet Evol. 69:1172–1185.
651 Cutter AD, Choi JY. 2010. Natural selection shapes nucleotide polymorphism across the genome of
652 the nematode *Caenorhabditis briggsae*. Genome Res. 20:1103–1111.
653 Cutter AD, Payseur BA. 2013. Genomic signatures of selection at linked sites: unifying the disparity
654 among species. Nat Rev Genet. 14:262–274
655 Cutter AD, Wang G-X, Ai H, Peng Y. 2012. Influence of finite-sites mutation, population
656 subdivision and sampling schemes on patterns of nucleotide polymorphism for species with
657 molecular hyperdiversity. Mol Ecol. 21:1345–1359.
658 Davies JM, Lowry CV, Davies KJ. 1995. Transient adaptation to oxidative stress in yeast. Arch
659 Biochem Biophys. 317:1–6.
660 De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, Van de Peer Y. 2013q.
661 Convergent gene loss following gene and genome duplications creates single-copy families in
662 flowering plants. PNAS 110(8):2898–2903. http://doi.org/10.1073/pnas.1300127110
663 Dentinger BTM, Margaritescu S, Moncalvo J-M. 2010. Rapid and reliable high-throughput methods
664 of DNA extraction for use in barcoding and molecular systematics of mushrooms. Mol Ecol
665 Res. 10:628–633.

666 Dettman JR, Anderson JB, Kohn LM. 2010. Genome□wide investigation of reproductive isolation in
667 experimental lineages and natural species of *Neurospora*: identifying candidate regions by
668 microarray□based genotyping and mapping. Evolution 64:694-709.
669 Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the
670 BEAST 1.7. Mol Biol Evol. 29:1969–1973.
671 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
672 Nucleic Acids Res. 32:1792–1797.
673 Egea R, Casillas S, Barbadilla A. 2008. Standard and generalized McDonald-Kreitman test: a website
674 to detect selection by comparing different classes of DNA sites. Nucleic Acids Res.
675 36:W157–W162.
676 Ekblom R, Galindo J. 2011. Application of next generation sequencing in molecular ecology of non-
677 model organisms. Heredity. 107:1–15.
678 Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. Trends
679 Ecol Evol. 29:51–63.
680 Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, Taylor JW. 2011. Population genomics
681 and local adaptation in wild isolates of a model microbial eukaryote. Proc Natl Acad Sci
682 USA. 108:2831–2836.
683 Eyre-Walker A. 2006. The genomic rate of adaptive evolution. Trends Ecol Evol 21:569–575.
684 Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat Rev
685 Genet. 8:610–618.
686 Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the
687 presence of slightly deleterious mutations and population size change. Mol Biol Evol.
688 26:2097–2108.
689 Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413.

690 Fay JC, Wu C-I. 2003. Sequence divergence, functional constraint, and selection in protein evolution.
691 Annu Rev Genom Human Genet. 4:213–235.
692 Fay JC, Wyckoff GJ, Wu CI. 2001. Positive and negative selection on the human genome. Genetics
693 158:1227–1234.
694 Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012. Emerging
695 fungal threats to animal, plant and ecosystem health. Nature 484:186–194.
696 Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. Genetics 133:693–709.
697 Fu YX, Li WH. 1999. Coalescing into the 21st century: An overview and prospects of coalescent
698 theory. Theor Popul Biol. 56:1–10.
699 Gao X, Gulari E, Zhou X. 2004. In situ synthesis of oligonucleotide microarrays. Biopolymers.
700 73:579–596.
701 Gerstein AC, Kuzmin A, Otto SP. 2014. Loss-of-heterozygosity facilitates passage through Haldane's
702 sieve for *Saccharomyces cerevisiae* undergoing adaptation. Nat Comm. 5:3819.
703 Gladieux P, Devier B, Aguileta G, Cruaud C, Giraud T. 2013. Purifying selection after episodes of
704 recurrent adaptive diversification in fungal pathogens. Infect Genet Evol. 17:123–131.
705 Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, de Vienne DM, Rodríguez de la Vega RC,
706 Branco S, Giraud T. 2014. Fungal evolutionary genomics provides insight into the
707 mechanisms of adaptive divergence in eukaryotes. Mol Ecol. 23:753–773.
708 Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA
709 sequences. Mol Biol Evol. 11:725–736.
710 Gossmann TI, Waxman D, Eyre-Walker A. 2014. Fluctuating selection models and McDonald-
711 Kreitman type analyses. PLoS ONE 9:e84540–5.
712 Han F, Peng Y, Xu L, Xiao P. 2014. Identification, characterization, and utilization of single copy
713 genes in 29 angiosperm genomes. BMC Genomics 15:504–9.

714 Hancock-Hanser BL, Frey A, Leslie MS, Dutton PH, Archer FI, Morin PA. 2013. Targeted multiplex
715 next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA
716 sequencing for population genomics. Mol Ecol Res. 13:254–268.
717 Heled J, Drummond AJ. 2008. Bayesian inference of population size history from multiple loci.
718 BMC Evol Biol. 8:289.
719 Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. Mol Biol
720 Evol. 27:570–580.
721 Hewitt GM. 1996. Some genetic consequences of ice ages, and their role in divergence and
722 speciation. Biol J Linn Soc. 58:247–276.
723 Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907–913.
724 Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. Phil Tran R Soc
725 B. 359:183–195.
726 Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. Nature 470:479–485.
727 Hough J, Williamson RJ, Wright SI. 2013. Patterns of selection in plant genomes. Ann Rev Ecol
728 Evol Syst. 44:31–49.
729 Huerta-Sanchez E, Durrett R, Bustamante CD. 2008. Population genetics of polymorphism and
730 divergence under fluctuating selection. Genetics 178:325–337.
731 Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex
732 class I loci reveals overdominant selection. Nature 335:167–170.
733 Kimura M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of
734 molecular evolution. Nature 267:275–276.
735 Kimura M. 1986. DNA and the neutral theory. Philos. Trans. R. Soc. Lond. B. 312:343–354.
736 Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum
737 A, Colpaert J, Copeland A, Costa MD, Doré J, Floudas D, Gay G, Girlanda M, Henrissat B,

738 Herrmann S, Hess J, Högberg N, Johansson T, Khouja HR, LaButti K, Lahrmann U,
739 Levasseur A, Lindquist EA, Lipzen A, Marmeisse R, Martino E, Murat C, Ngan CY, Nehls U,
740 Plett JM, Pringle A, Ohm RA, Perotto S, Peter M, Riley R, Rineau F, Ruytinx J, Salamov A,
741 Shah F, Sun H, Tarkka M, Tritt A, Veneault-Fourrey C, Zuccaro A, Tunlid A, Grigoriev IV,
742 Hibbett DS, Martin F. 2015. Convergent losses of decay mechanisms and rapid turnover of
743 symbiosis genes in mycorrhizal mutualists. Nat Genet 47:410–415.

744 Kontoyiannis DP, Lewis RE. 2002. Antifungal drug resistance of pathogenic fungi. Lancet
745 359:1135–1144.

746 Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. PLoS Genet 4:e1000304.

747 Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform.
748 Bioinformatics 26:589–595.

749 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000
750 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and
751 SAMtools. Bioinformatics 25:2078–2079.

752 Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism
753 data. Bioinformatics 25:1451–1452.

754 Liu ZL. 2006. Genomic adaptation of ethanologenic yeast to biomass conversion inhibitors. Appl
755 Microbiol Biotechnol 73:27–36.

756 McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature
757 351:652–654.

758 Messer PW, Petrov DA. 2013a. Frequent adaptation and the McDonald-Kreitman test. Proc Natl
759 Acad Sci USA. 110:8615–8620.

760 Messer PW, Petrov DA. 2013b. Population genomics of rapid adaptation by soft selective sweeps.
761 Trends Ecol Evol. 28:659–669.

762 Nehls U, Grunze N, Willmann M, Reich M, Küster H. 2007. Sugar for my honey: carbohydrate
763 partitioning in ectomycorrhizal symbiosis. Phytochemistry 68:82–91.
764 Nei M. 1987. Molecular Evolutionary Genetics. Columbia University Press.
765 Nei M., Li W-H. 1979. Mathematical model for studying genetic variation in terms of restriction
766 endonuclease. Proc Natl Acad Sci USA. 89:1477–1481.
767 van der Nest MA, Beirn LA, Crouch JA, Demers JE, de Beer ZW, De Vos L, Gordon TR, Moncalvo
768 J-M, Naidoo K, Sánchez-Ramírez S, Roodt D, Santana QC, Slinski SL, Stata M, Taerum SJ,
769 Wilken PM, Wilson AM, Wingfield MJ, Wingfield BD. 2014. IMA Genome-F 3: Draft
770 genomes of *Amanita jacksonii, Ceratocystis albifundus, Fusarium circinatum, Huntiella*
771 *omanensis, Leptographium procerum, Rutstroemia sydowiana, and Sclerotinia echinophila*.
772 IMA Fungus. 5:472–485.
773 Nielsen R. 1997. The ratio of replacement to silent divergence and tests of neutrality. J. Evol. Biol.
774 10:217–231.
775 Nielsen R. 2001. Statistical tests of selective neutrality in the age of genomics. Heredity.
776 Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for
777 selective sweeps using SNP data. Genome Res. 15:1566–1575.
778 Ohta T. 1992. The nearly neutral theory of molecular evolution. Annu Rev Ecol Syst. 23:263–286.
779 Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language.
780 Bioinformatics. 20:289–290.
781 Piper P, Calderon CO, Hatzixanthis K, Mollapour M. 2001. Weak acid adaptation: the stress response
782 that confers yeasts with resistance to organic acid food preservatives. Microbiology.
783 147:2635–2642.
784 Rambaut A, Suchard MA, Xie W, Drummond AJ. 2013. Tracer v16.
785 http://tree.bio.ed.ac.uk/software/tracer/

786 Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A. 2008. Statistical power
787 analysis of neutrality tests under demographic expansions, contractions and bottlenecks with
788 recombination. Genetics. 179:555–567.
789 Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M.
790 A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic
791 inference and model choice across a large model space. Syst Biol. 61:539–542.
792 Sánchez-Ramírez S, Tulloss RE, Amalfi M, Moncalvo JM. 2015a. Palaeotropical origins,
793 boreotropical distribution and increased rates of diversification in a clade of edible
794 ectomycorrhizal mushrooms (*Amanita* section *Caesareae*). J Biogeogr. 42:351–363.
795 Sánchez-Ramírez S, Tulloss RE, Guzmán-Dávalos L, Cifuentes-Blanco J, Valenzuela R, Estrada-
796 Torres A. et al. 2015b. In and out of refugia: historical patterns of diversity and demography
797 in the North American Caesar's mushroom species complex. Mol Ecol. 24:5938–5956.
798 Schoustra SE, Debets AJM, Slakhorst M, Hoekstra RF. 2007. Mitotic recombination accelerates
799 adaptation in the fungus *Aspergillus nidulans*. PLoS Genet. 3:e68.
800 Selbmann L, Egidi E, Isola D, Onofri S, Zucconi L, de Hoog GS, Chinaglia S, Testa L, Tosi S,
801 Balestrazzi A, Lantieri A, Compagno R, Tigini V, Varese GC. 2013. Biodiversity, evolution
802 and adaptation of fungi in extreme environments. Plant Biosyst. 147:237–246.
803 Sella G, Petrov DA, Przeworski M, Andolfatto P. 2009. Pervasive natural selection in the *Drosophila*
804 genome? PLoS Genet 5:e1000495.
805 Slotte T, Foxe JP, Hazzouri KM, Wright SI. 2010. Genome-wide evidence for efficient positive and
806 purifying selection in *Capsella grandiflora*, a plant species with a large effective population
807 size. Mol Biol Evol. 27:1813–1821.
808 Smith, N. G. C., and A. Eyre-Walker. 2002. Adaptive protein evolution in Drosophila. Nature
809 415:1022–1024.

Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. 2009. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. Genetics. 182:205–216.

Stajich JE, Hahn MW. 2005. Disentangling the effects of demography and selection in human history. Mol Biol Evol. 22:63–73.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34:W435–439.

Suchard MA, Rambaut A. 2009. Many-core algorithms for statistical phylogenetics. Bioinformatics. 25:1370–1376.

Suutari M, Liukkonen K, Laakso S. 1990. Temperature adaptation in yeasts: the role of fatty acids. J Gen Microbiol. 136:1469–1474.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.

Templeton JEL, Brotherton PM, Llamas B, Soubrier J, Haak W, Cooper A, Austin JJ. 2013. DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. Investig Genet. 4:26.

Tulloss RE. 2005. *Amanita*-distribution in the Americas, with comparison to eastern and southern Asia and notes on spore character variation with latitude and ecology. Mycotaxon. 93:189–231.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 7:256–276.

Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, Wright SI. 2014. Evidence for widespread positive and negative selection in coding and conserved noncoding regions of *Capsella grandiflora*. PLoS Genet. 10:e1004622.

834 Wolfe BE, Tulloss RE, Pringle A. 2012. The irreversible loss of a decomposition pathway marks the
835 single origin of an ectomycorrhizal symbiosis. PLoS One. 7:e39597.
836 Wright SI, Gaut BS. 2005. Molecular population genetics and the search for adaptive evolution in
837 plants. Mol Biol Evol. 22:506–519.
838 Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate
839 lysozyme evolution. Mol Biol Evol. 15:568–573.
840 Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 24:1586–
841 1591.
842 Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under
843 realistic evolutionary models. Mol Biol Evol. 17:32–43.
844 Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous
845 selection pressure at amino acid sites. Genetics 155:431–449.
846 Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual
847 sites along specific lineages. Mol Biol Evol. 19:908–917.
848 Yang Z, Wong WSW, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under
849 positive selection. Mol Biol Evol. 22:1107–1118.
850 Zdobnov EM, Apweiler R. 2001. InterProScan –an integration platform for the signature-recognition
851 methods in InterPro. Bioinformatics. 17:847–848.
852 Zeng K, Fu Y-X, Shi S, Wu CI. 2006. Statistical tests for detecting positive selection by utilizing
853 high-frequency variants. Genetics 174:1431–1439.
854

855 **Figure captions**

856 Figure 1. A, Species distribution maps based on geographic data from Sánchez-Ramírez et al.
857 (2015b) and MaxEnt (Phillips et al. 2006). B, Species tree showing phylogenetic relationships and

858 divergence times with relation to the outgroup (modified from Sánchez-Ramírez et al. 2015b).
859 Branches show posterior probability support.
860
861 Figure 2. Polymorphism and divergence data for 502 genes. A, box-plots of Tajima's *D* separated by
862 site-classes (non-synonymous, white; synonymous, light grey; introns, dark grey) for the *A. jacksonii*
863 complex. B, box-plots of the ratios of non-synonymous to synonymous polymorphism and
864 divergence. C, α values from the standard MKT test.
865
866 Figure 3. Estimates of (A) the proportion of adaptive substitutions (α_{dfe}) and (B) the rate of adaptive
867 evolution (ω_{α}) based on the distribution of fitness effects.
868
869 Figure 4. The proportion of deleterious mutations under four different $N_e s$ categories based on the
870 distribution of fitness effects model.
871
872 Figure 5. Demographic trends through time for each species based on the eBSP model. Grey shading
873 represents 95% posterior confidence intervals (CI). Solid lines represent mean population size values
874 and dotted lines represent median values.
875
876 Figure 6. Gene Ontology terms *Wordles* for (A) all genes in the data set, (B) genes with $\alpha > 1$, and
877 genes with a significant proportion of sites under diversifying selection. The size of the term is
878 equivalent to the extent of selection and how often the term was found across genes and species.
879
880
881

882

883

884 **Supplementary data**

885 Table S1. GO terms of singleton genes with $\alpha > 1$.

886 Table S2. Species, specimen voucher name, and geographic location data for all samples.

887 Figure S1. Variable sites and percentage of missing data plotted against gene lenght.

888 Figure S2. Sample of 100 genealogies from 502 estimated trees. Terminal colors indicate different

889 species.

890 Figure S3. Distribution of *gsi* values for all genes and species. Species with higher amounts of

891 conflict are highlight.

892 Figure S4. Scatterplot of Tajima's *D* in synonymous and non-synonymous sites against the *gsi*. Grey

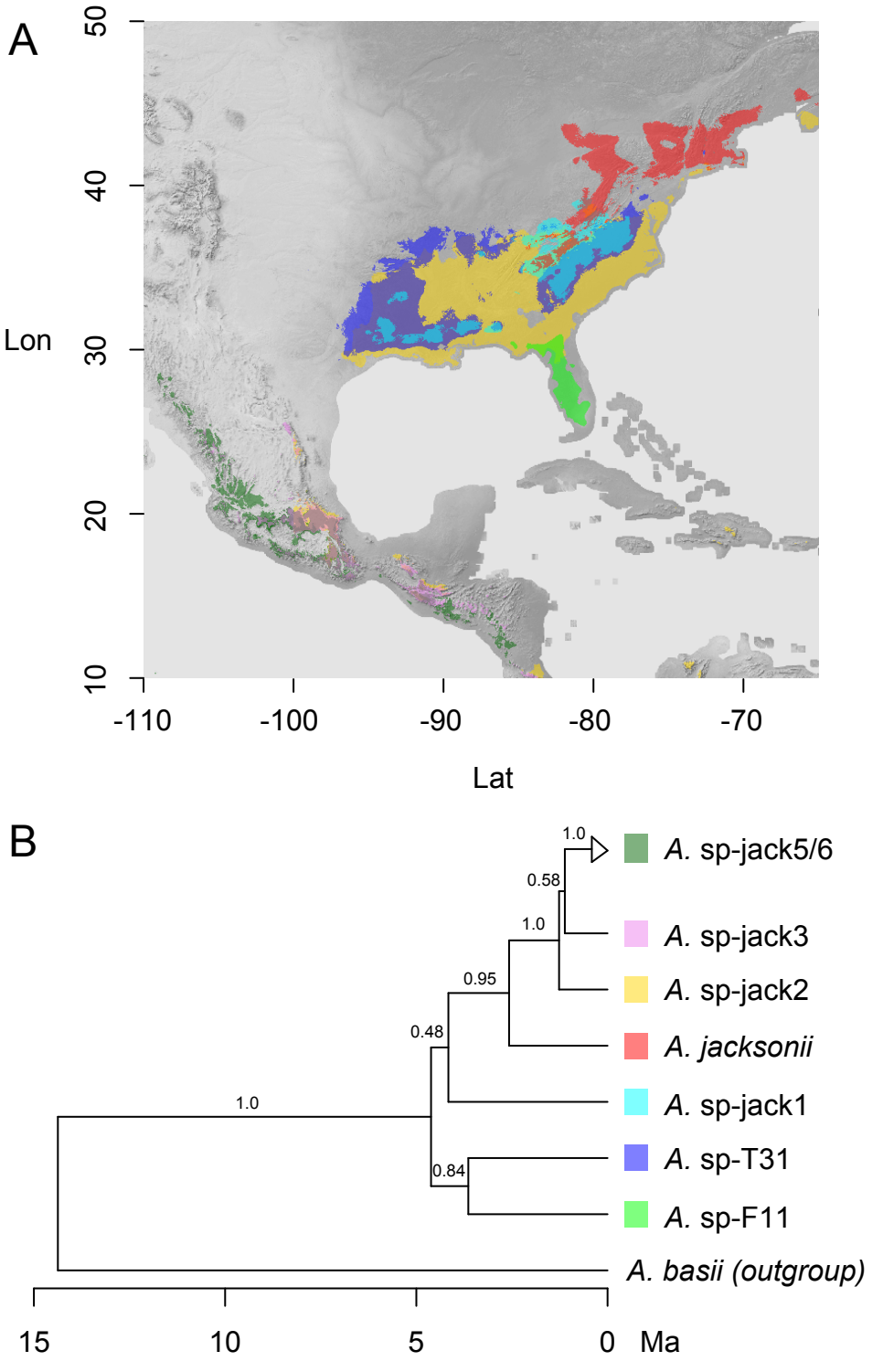
893 points highlight reciprocally monophyletic genes.

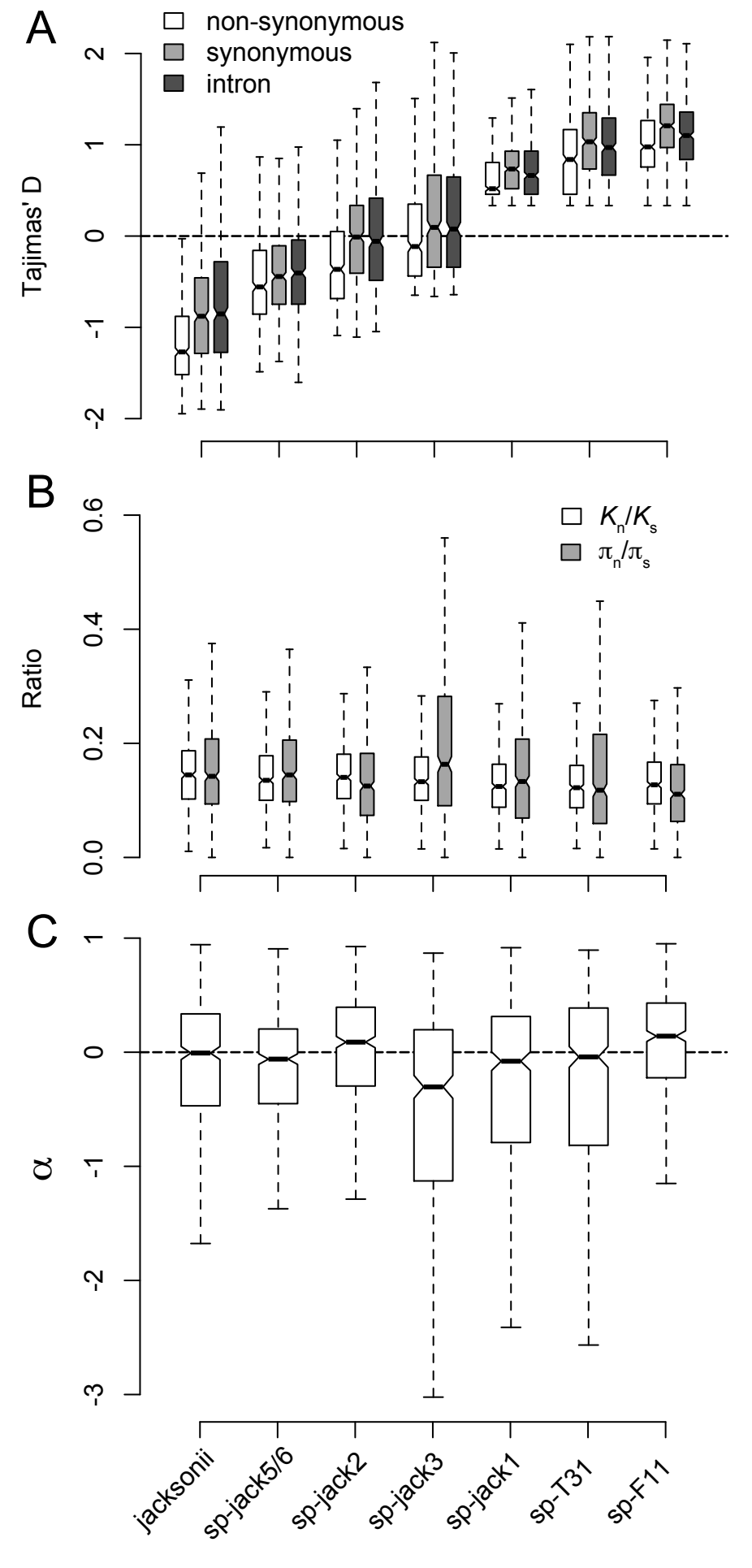
894 Figure S5. Scatterplot of α values against the *gsi*.

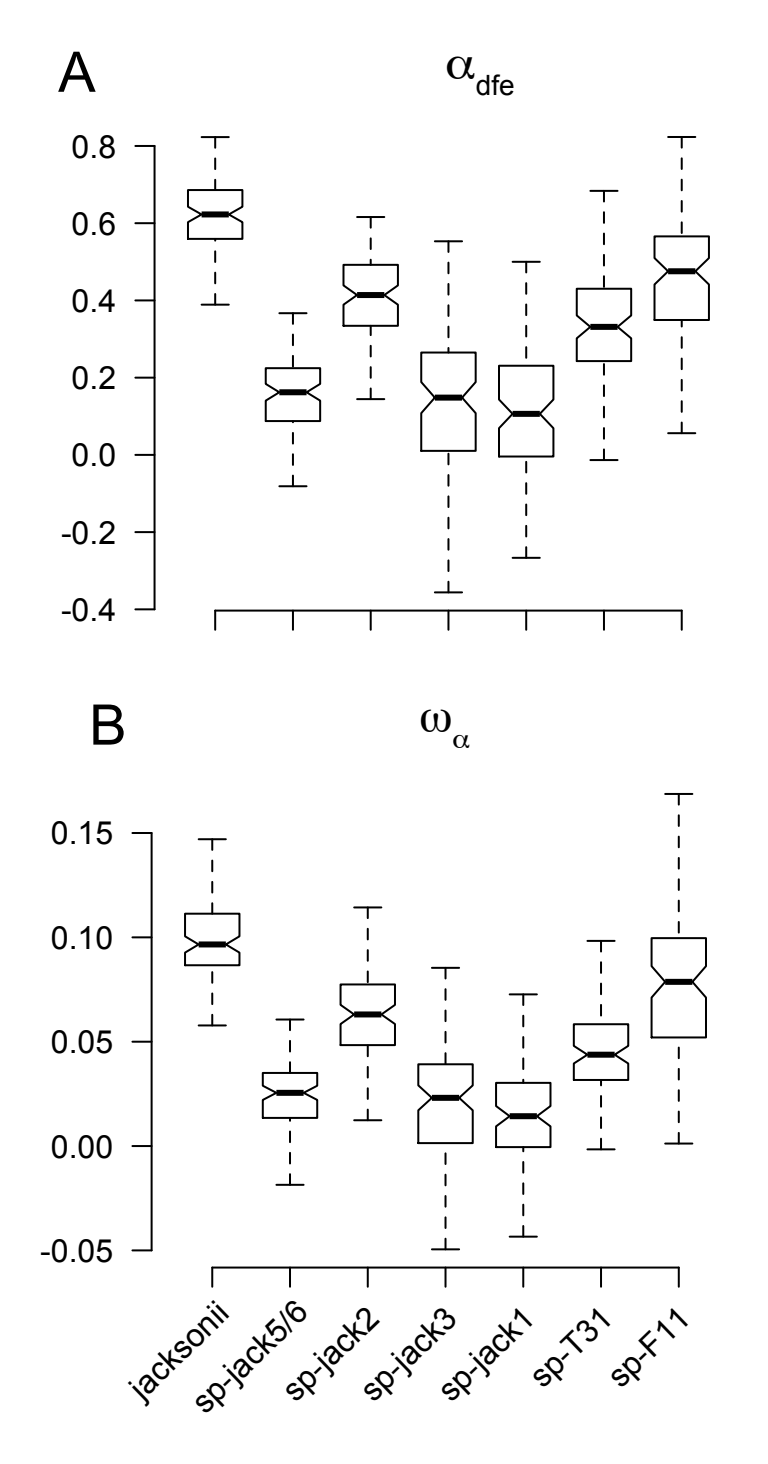
895 Figure S6. Scatterplot of α values and synonymous nucleotide diversity against codon usage bias

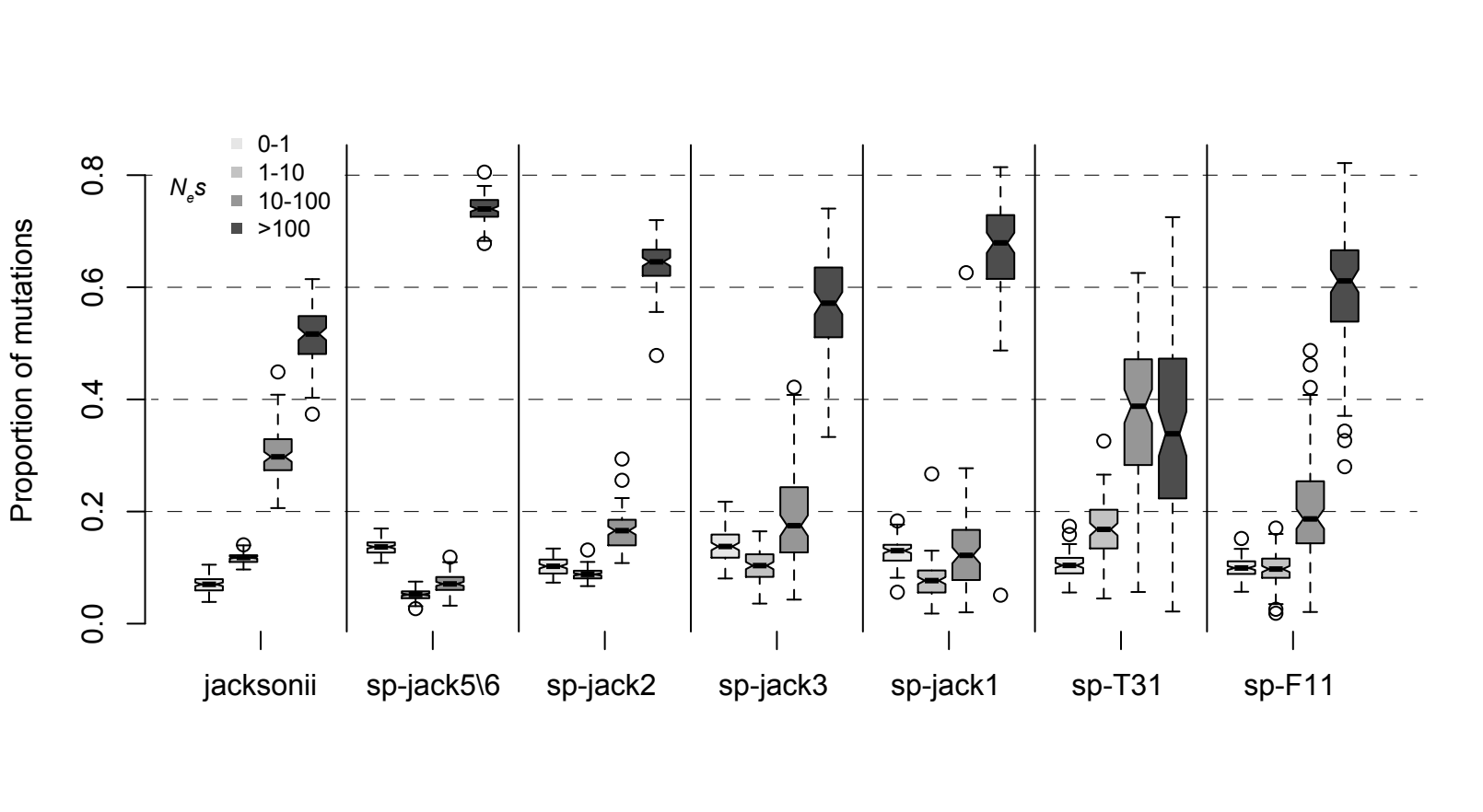
896 (CAI).

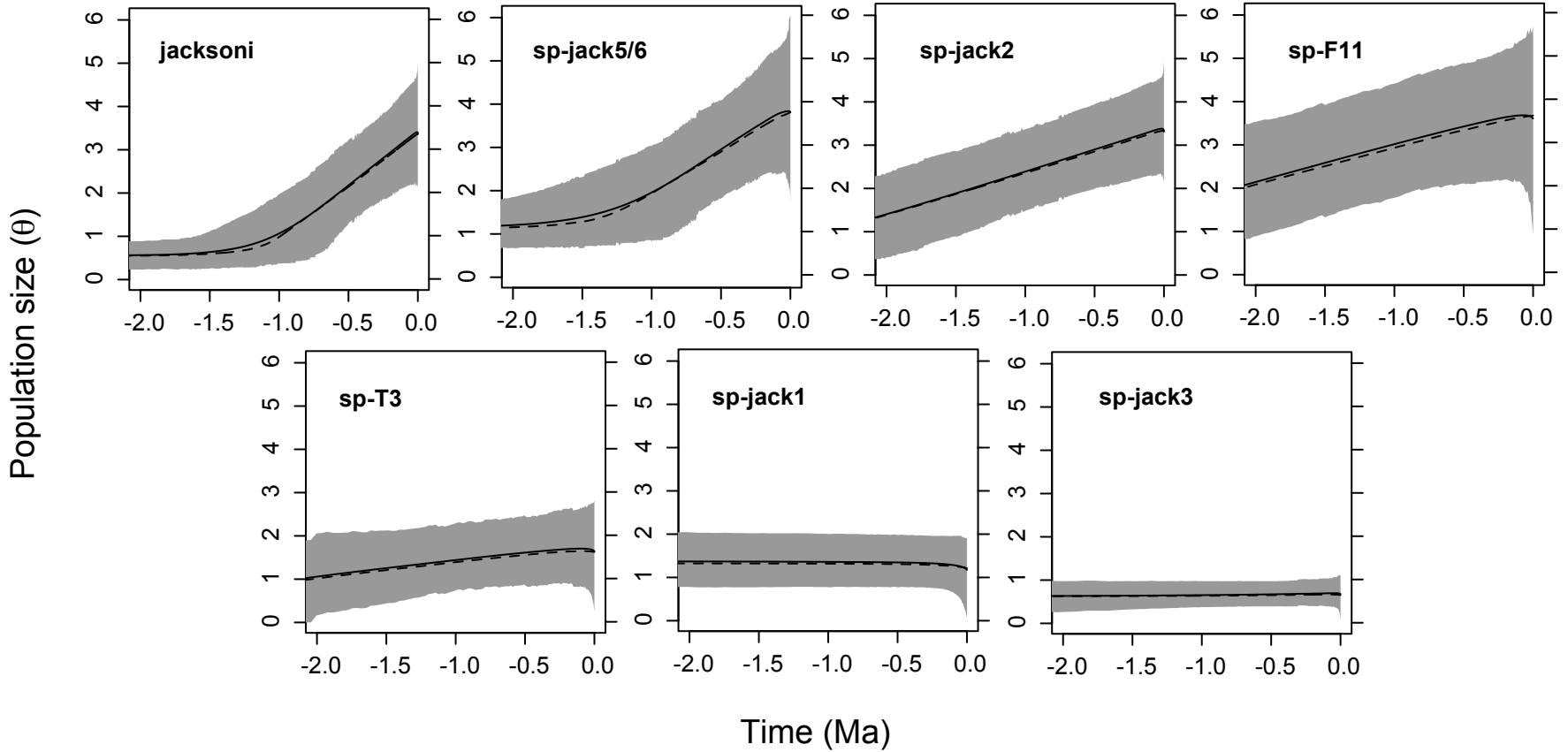
897


A
Lon
50
40
30
20
10
-110
-100
-90
-80
-70
Lat
B
1.0
0.58
1.0
0.95
0.48
1.0
0.84
A. sp-jack5/6
A. sp-jack3
A. sp-jack2
A. jacksonii
A. sp-jack1
A. sp-T31
A. sp-F11
A. basii (outgroup)
15
10
5
0
Ma

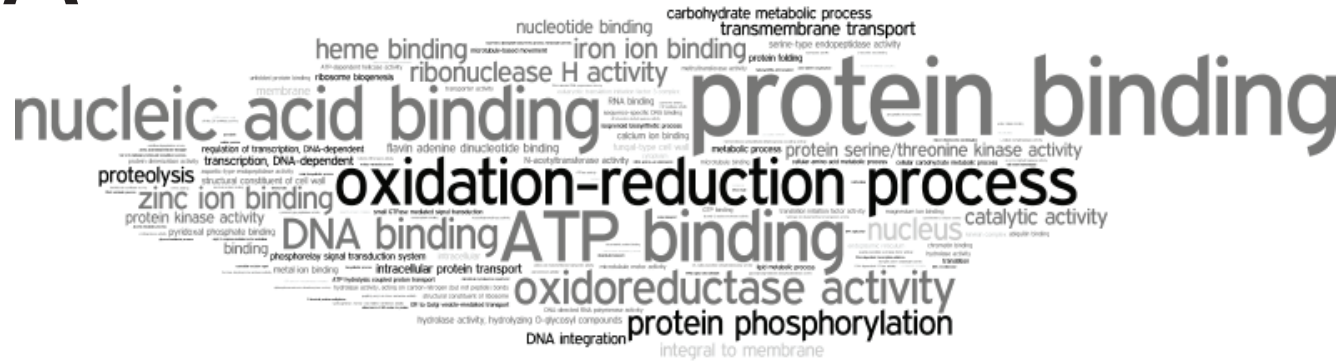



A
α_{dfe}
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
B
ω_{α}
0.15
0.10
0.05
0.00
-0.05
jacksonii
sp-jack5/6
sp-jack2
sp-jack3
sp-jack1
sp-T31
sp-F11


Proportion of mutations
0.0
0.2
0.4
0.6
0.8
$N_e s$
0-1
1-10
10-100
>100
jacksonii
sp-jack5\6
sp-jack2
sp-jack3
sp-jack1
sp-T31
sp-F11


jacksoni
sp-jack5/6
sp-jack2
sp-F11
sp-T3
sp-jack1
sp-jack3
Population size (θ)
Time (Ma)
0
1
2
3
4
5
6
-2.0
-1.5
-1.0
-0.5
0.0

A


carbohydrate metabolic process
nucleotide binding
transmembrane transport
heme binding
iron ion binding
ribonuclease H activity
protein binding
nucleic acid binding
membrane
ribosome biogenesis
regulation of transcription, DNA-dependent
flavin adenine dinucleotide binding
protein serine/threonine kinase activity
transcription, DNA-dependent
proteolysis
oxidation-reduction process
zinc ion binding
protein kinase activity
DNA binding
ATP binding
nucleus
catalytic activity
binding
phosphorelay signal transduction system
intracellular protein transport
oxidoreductase activity
protein phosphorylation
DNA integration
integral to membrane

B


iron ion binding
ribosome biogenesis
catalytic activity
protein phosphorylation
metal ion binding
tricarboxylic acid cycle
intracellular
urease activity
DNA integration
protein serine/threonine kinase activity
cis-Golg network
proteolysis
heme binding
DNA-directed RNA polymerase activity
nucleotide-excision repair
nucleotide binding
chromatin binding
peroxiredoxin activity
tRNA processing
metabolic process
catalytic activity
mitochondrial outer membrane
cysteinyl-tRNA aminoacylation
binding
protein serine/threonine kinase activity
unfolded protein binding
translation initiation factor activity
endonuclease activity
CTP synthase activity
protein peptidyl-prolyl isomerization
ubiquitin binding
metabolic process
tRNA dihydrouridine synthase activity
response to stress
tetrahydrofolylpolyglutamate synthase activity
prefoldin complex
proton-transporting ATPase activity, rotational mechanism
proton-transporting two-sector ATPase complex, catalytic domain
phosphatidylinositol-3-phosphate binding
urease activity
vesicle docking involved in exocytosis
DNA metabolic process
oxidation-reduction process
DNA integration
eukaryotic translation initiation factor 3 complex
protein binding
proteasome regulatory particle
binding
spliceosomal complex
endonuclease activity
cysteine-tRNA ligase activity
ATP binding
anion transport
sulfate transport
urea catabolic process
ribonucleoside binding
protein folding
aspartic-type endopeptidase activity
protein folding
mRNA splicing, via spliceosome
protein heterodimerization activity
sequence-specific DNA binding
regulation of small GTPase mediated signal transduction
pyrimidine nucleotide biosynthetic process
DNA-dependent ATPase activity
N-acetyltransferase activity
folic acid-containing compound biosynthetic process
nickel cation binding
proteolysis
structural constituent of ribosome
N-acetyltransferase activity
guanyl-nucleotide exchange factor activity
metal ion binding
CTP synthase activity
pyridoxal phosphate binding
nucleotide binding
single-stranded DNA binding
intracellular protein transport
citrate (Si)-synthase activity
translation
intracellular
ribosome
oxidoreductase activity
cytochrome-c oxidase activity
peptidyl-prolyl cis-trans isomerase activity
transaminase activity
tRNA processing
translation
ribosome

C


amino acid binding
lysine biosynthetic process via diaminopimelate
telomeric template RNA reverse transcriptase activity
nucleus
small GTPase mediated signal transduction
retromer complex
threonine biosynthetic process
guanyl-nucleotide exchange factor activity
sequence-specific DNA binding transcription factor activity
protein kinase activity
regulation of transcription, DNA-dependent
N-acetyltransferase activity
protein phosphorylation
phosphorylation
DNA binding
aspartate kinase activity
zinc ion binding