1

Unmatched level of molecular convergence among deeply divergent complex multicellular fungi

Zsolt Merényi[1], Arun N. Prasanna[1&], Wang Zheng[2], Károly Kovács[1], Botond Hegedüs[1], Balázs Bálint[1], Balázs Papp[1], Jeffrey P. Townsend[2,3,4], László G. Nagy[1]*

[1]*Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Center, HAS, Szeged, 6726, Hungary*

[2]*Department of Biostatistics, Yale University, New Haven, CT, United States of America,*

[3]*Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, United States of America,*

[4]*Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, United States of America*

*Author for correspondence: lnagy@fungenomelab.com

[&]Current address: Red Sea Science and Engineering Research Center, 4700 King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia

Genetic mechanisms of the convergent evolution of multicellularity suggest that predictability of evolution can be driven by genome composition.

Abstract

Convergent evolution is pervasive in nature, but it is poorly understood how various constraints and natural selection limit the diversity of evolvable phenotypes. Here, we report that, despite >650 million years of divergence, the same genes have repeatedly been co-opted for the development of complex multicellularity in the two largest clades of fungi—the Ascomycota and Basidiomycota. Co-opted genes have undergone duplications in both clades, resulting in >81% convergence across shared multicellularity-related families. This convergence is coupled with a rich repertoire of multicellularity-related genes in ancestors that predate complex multicellular fungi, suggesting that the coding capacity of early fungal genomes was well suited for the repeated evolution of complex multicellularity. Our work suggests that evolution may be predictable not only when organisms are closely related or are under similar selection pressures, but also if the genome biases the potential evolutionary trajectories organisms can take, even across large phylogenetic distances.

Introduction

Darwin suggested that organisms can evolve an unlimited variety of forms (*1*). Contrary to his concept of 'unlimited forms', it is now clear that evolution follows similar paths more often than classic models of genetic change would predict (*2–5*). The independent emergence of similar phenotypes is called convergent evolution, which happens in response to similar selection pressures, bias in the emergence of phenotypic variation (*6, 7*), or both (*3, 5, 8*). Convergence is widespread in nature (e.g. (*9–12*)) suggesting that evolution may be predictable (*13*) and deterministic (*5, 11, 14*) under some circumstances, although what drives divergent lineages to evolve similar phenotypes is poorly understood.

A fascinating example of convergent phenotypes is multicellularity: it has evolved at least 25–30 times across the pro- and eukaryotes (*15–23*), reaching a diversity of complexities that range from simple cell aggregates to the most complex macroscopic organisms (*24*). Instances of the evolution of multicellularity are considered major transitions in evolution—a conceptual label that is difficult to reconcile with repeated origins (*15, 17*). This difficulty follows from the assumption that major transitions are limited by big genetic hurdles and thus should occur rarely during evolution (*17, 25*).

The highest level of multicellular organization is referred to as complex multicellularity (CM), which, unlike unicells and simple multi-celled aggregates (e.g. filaments, colonies, biofilms, etc.), is characterized by a three-dimensional organization, sophisticated mechanisms for cell-cell adhesion and communication, and extensive cellular differentiation (*17, 23, 24, 26*). CM occurs in metazoans, embryophyte plants, and fungi as well as red and brown algae. In fungi, CM refers mostly to sexual fruiting bodies which are found in 8-11 disparate fungal clades and show clear signs of convergent origins (*22*). Although they originated independently, CM fungal clades are phylogenetically close, providing a tractable system for studying the genetics of major evolutionary transitions in complexity. Fruiting bodies in fungal lineages can be developmentally and morphologically highly distinct—yet they evolved for the same general purpose: to enclose sexual reproductive structures in a protective environment and facilitate spore dispersal (*23, 27–30*). Here we seek to explain the convergent evolution of fungal fruiting bodies by analyzing the fate of multicellularity-related gene families across the two largest clades of CM fungi: the Agaricomycotina (mushroom-forming fungi, Basidiomycota) and the Pezizomycotina (Ascomycota).

Results

To study the evolution of complex multicellularity in fungi we first reconstructed ancestral cellularity levels across a phylogenetic tree of 19 representative species (table S1). This analysis strongly suggests that the two most diverse CM clades, Agarico- and Pezizomycotina, acquired fruiting body formation independently (Fig. 1/a). Likelihood proportions imply that the most recent common ancestor (MRCA) of the Dikarya did not form fruiting bodies (marginal probability of non-CM: 0.87, CM: 0.13), followed by two independent acquisitions of CM in the MRCA of the Agaricomycetes and MRCA of Pezizomycotina. Our data suggest that the MRCA of the Ascomycota did not produce fruiting

80 bodies (state non-CM: 0.616, state CM: 0.384), which implies a third independent origin of
81 fruiting body production in *Neolecta* (Taphrinomycotina) (*22*, *31*, *32*) (Fig. 1/a).
82   To identify developmentally relevant genes, we used publicly available fruiting body
83 transcriptomes of 5 Pezizomycotina (*33–36*) and 4 Agaricomycotina (*37–39*) species, with
84 which it was possible to quantify gene expression across 2–13 developmental stages. Based
85 on expression dynamics, we detected 2645–9444 developmentally regulated genes in the nine
86 species (Fig. S1), corresponding to 19.8–66.3% of the proteome. The identified
87 developmentally regulated genes contained 26.9-97.6%, 4.6-69.2% and 92.7% of known
88 developmental genes of *Neurospora, Aspergillus* and *Coprinopsis,* respectively (Note S1,
89 table S2), consistent with previous studies of fruiting body development (*23*, *28–30*, *40*). In a
90 broader dataset of 19 species (see Methods) developmentally regulated genes fell into 21,267
91 families, of which we focused on that ones that showed conserved developmental regulation
92 in the majority of species (Fig. 1/b). We identified 1,026 gene families that were
93 developmentally regulated in ≥75% of the species in either or both clades, resulting in 314,
94 273, and 439 families that have a conserved developmental expression in ≥7 of 9 Dikarya, ≥4
95 of 5 Pezizomycotina and ≥3 of 4 Agaricomycotina species, respectively (table S3). We
96 hereafter focus on these families because these are most likely to have been developmentally
97 regulated also in the most recent common ancestor of Agaricomycotina and that of the
98 Pezizomycotina.
99
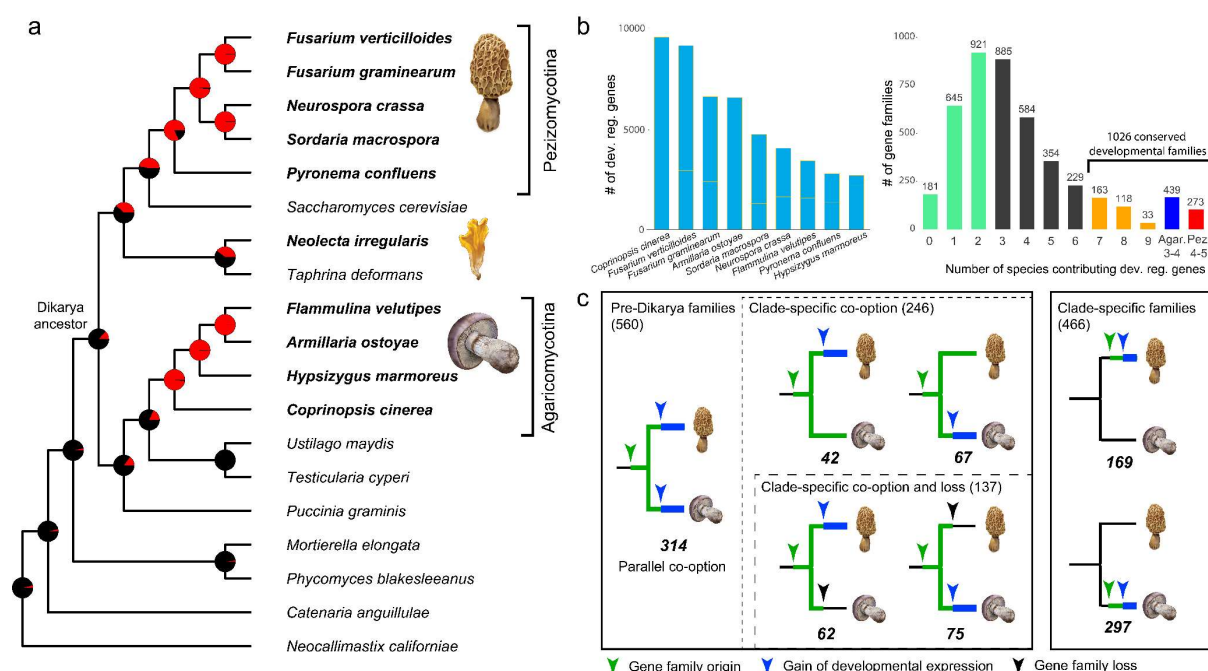
100
101 **Figure 1**. The evolution of complex multicellularity in fungi and conserved developmentally
102 regulated gene families. **(a)** phylogenetic relationships among 19 species analyzed in this study
103 inferred from 86 conserved, single-copy orthologs. Two independent clades of complex multicellular
104 species are marked, and typical fruiting body morphologies are shown. Pie charts at nodes indicate the
105 proportional likelihoods of CM (red) and non-CM (black) ancestral states reconstructed using
106 Maximum Likelihood. Character state coding of extant species are shown as bold (CM) or regular
107 (non-CM) font. **(b)** the number of developmentally regulated genes in each of the nine species (left)
108 and the number of gene families in which these genes grouped (shared by ≥7 species, right). Groups
109 of gene families that are developmentally regulated in ≥3 Agaricomycotina or ≥4 Pezizomycotina are

3

110 also shown. **(c)** developmentally regulated gene families grouped by evolutionary conservation and
111 history.
112

Widespread parallel co-option of developmental families

114 We analyzed the origin of the genetic bases of CM by reconstructing the evolution of
115 developmental gene families along the phylogeny. Of the 1,026 conserved developmental
116 families, 560 predate the origin of CM, consistent with their co-option for multicellularity-
117 related functions, while 297 and 169 families are taxonomically restricted to the Agarico- and
118 Pezizomycotina, respectively (Fig. 1/c). Of the 560 ancient families, 314 (56.1%) are
119 developmentally regulated in both the Agarico- and Pezizomycotina, indicating parallel co-
120 option for fruiting body development. The remaining 246 families can be divided into those
121 that have homologs in only one CM clade and were lost in the other (24.5%, 137 families)
122 and those that have homologs in both CM clades but are developmentally regulated only in
123 one (19.5%, 109 families), consistent with clade-specific co-option. The frequency of clade-
124 specific co-option is low, with 42 and 67 families in the Agarico- and Pezizomycotina (7.5%
125 and 12.0%), respectively. The observation of limited clade-specific, but widespread parallel
126 co-option suggests that gene families with suitable properties for CM rarely escaped
127 integration into the genetic toolkit of CM. It also agrees with genes suitable for a given
128 phenotype being rare and thus mostly being recruited under similar selection regimes (*41*).
129   The observed distribution of developmentally regulated gene families is consistent
130 with two hypotheses. Families with clade-specific phylogenetic distribution or clade-specific
131 developmental expression conform to expectations under the simplest model of convergent
132 evolution at the phenotypic level: two independent gains of CM in the Agarico- and
133 Pezizomycotina. Similarly, shared developmentally expressed families could have been
134 independently co-opted in the two CM clades. However, this set of families could also
135 encode plesiomorphic functions that were present in the Dikarya ancestor and were
136 independently integrated into CM in the Agarico- and Pezizomycotina. Some of those might
137 have served as precursors to CM (e.g. as traits linked with asexual development (*42*)), which
138 could have predisposed lineages for evolving CM independently, leading to a higher
139 likelihood for phenotypic convergence (*23, 43, 44*).
140   Although the Dikarya ancestor most likely did not have fruiting bodies (see Fig. 1/a),
141 we reasoned that its ancestral gene complement could reveal whether evolutionary
142 predisposition is a reasonable hypothesis to explain CM in fungi. The Dikarya had 989 genes
143 in the 314 shared developmental families (Fig. 2), which we functionally characterized by
144 examining *Saccharomyces cerevisiae* orthologs. Analyses of Gene Ontology terms revealed
145 an enrichment of genes for the regulation of growth, filamentous growth in response to
146 starvation, transmembrane transport, cell communication, gene expression regulation and
147 carbohydrate metabolism, reminiscent of general functions required for fungal development
148 (Fig. S2). We find that several gene regulatory circuits, including ones involved in sexual
149 reproduction, mating partner recognition, light, nutrient and starvation sensing, fungal cell-
150 wall synthesis and modification, cell-to-cell signaling and morphogenesis have been present

151   in the Dikarya (and even earlier), suggesting that these might have provided a foundation for
152   the evolution of fruiting bodies.
153
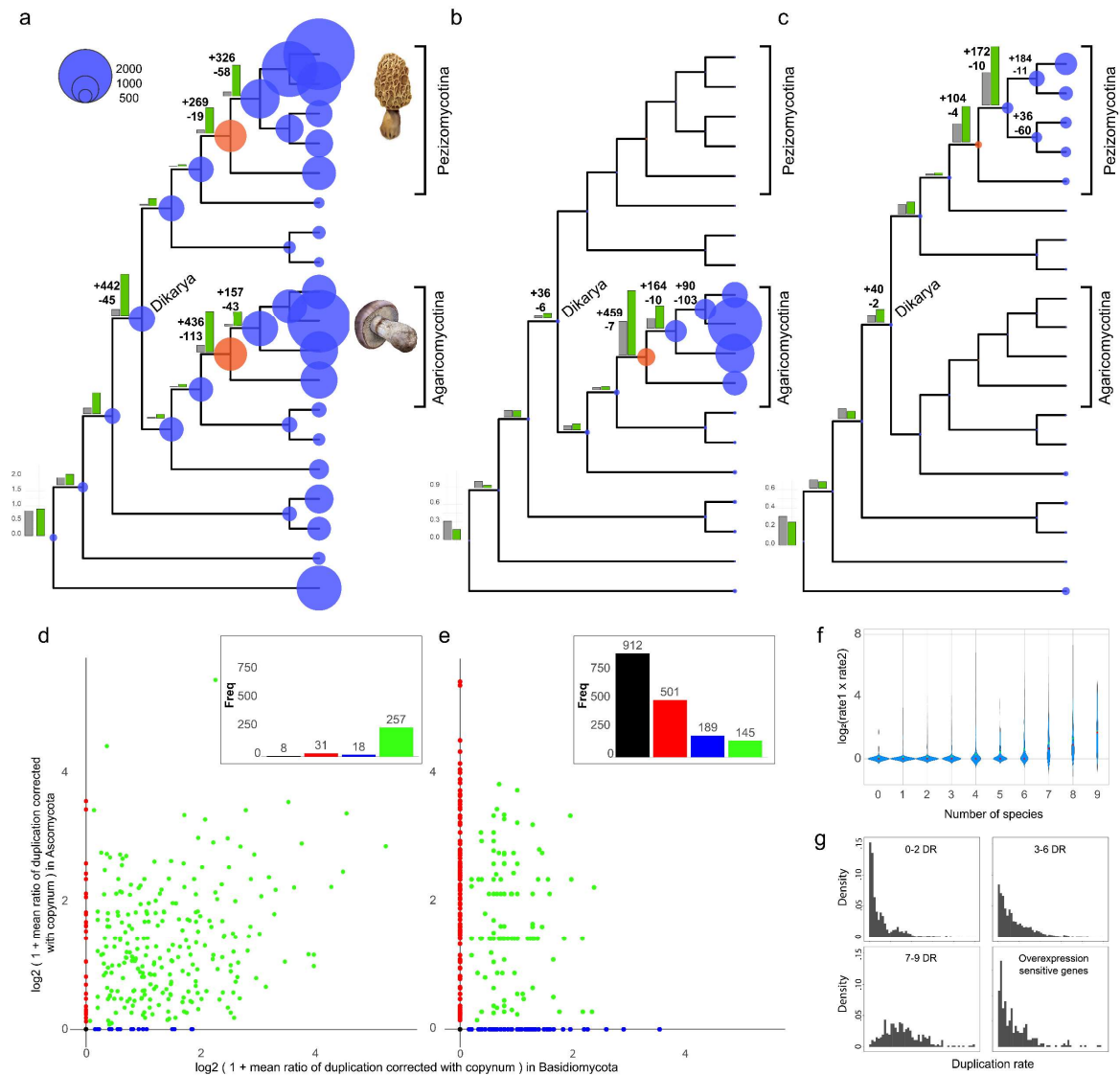
Figure 2. Convergent expansion of developmentally regulated gene families in independent complex
multicellular fungi. **(a-c)** Reconstructed copy number evolution of 314 shared developmentally
regulated gene families (a), 439 families with Agaricomycotina-specific developmental expression
and (c) 273 families with Pezizomycotina-specific developmental expression. Bubble size
proportional to the number of reconstructed ancestral gene copies across the analyzed families.
Numbers next to internal nodes denote the number of inferred duplications and losses. Bar graphs
show genome-wide duplication rates (grey) versus duplication rates of the depicted developmental
families (green). Inferred gains of CM are indicated by red bubbles. **(d-e)** scatterplot of Agarico- and
Pezizomycotina duplication rates across 314 shared developmentally regulated gene families (d) and
1747 families containing ≤2 developmentally regulated species (e). Black, red, blue and green denote
families with no duplications, Ascomycota specific-, Basidiomycota specific- and parallel
duplications, respectively. Bar diagrams show the number of gene families in each category. **(f)**
correlation between the extent of convergence and the number of species contributing
developmentally regulated genes to a family. **(g)** the distribution of gene duplication rates across gene

5

169 families containing developmentally regulated genes from ≤2, 3-6 and ≥7 species and families in
170 which dosage effects constrain duplications rates (*45*).

Gene family expansions correlate with the origins of CM

172 To obtain a higher-resolution picture of the evolution of developmental gene families, we
173 reconstructed gene duplications and losses in the Agarico- and Pezizomycotina and in other
174 parts of the fungal tree. We found characteristic expansions of developmentally regulated
175 gene families in CM Agarico- and Pezizomycotina, but no or significantly smaller expansions
176 in other families (Fig. 2/a, fig. S3). Across the 314 shared families, we inferred a net
177 expansion (duplication minus loss) of 323 and 250 genes in the MRCA of the CM Agarico-
178 and that of CM Pezizomycotina, respectively, indicating that the origin of these groups
179 coincided with significant expansions in developmentally regulated gene families. In the
180 MRCA of Dikarya we found 442 duplications and 45 losses. The observed gene family
181 expansions were driven by increased gene duplication rates, with loss rates remaining
182 approximately constant (Fig. S4). A 6.3 to 8.1-fold higher rate of expansion was found in the
183 314 shared developmental families compared to other families shared by ≥7 of the nine
184 species, indicating that CM-related gene families are one of the most expanding group in the
185 fungal genomes examined here.
186   Gene families with a developmental expression specific to the Agarico- (439 families)
187 or Pezizomycotina (273 families) show higher duplication rates (1.93–1.95-fold) and
188 substantially expanded in their respective clades, but not in the other CM subphylum or in
189 non-CM species (Fig. 2/bc). Of the Agarico- and Pezizomycotina the latter showed a more
190 gradual expansion of gene families: we reconstructed 104 and 162 net gains in the MRCA of
191 Pezizomycotina and that of Sordariomycetes, respectively. Interestingly, we inferred
192 relatively few (73) duplications along the branch leading to *Pyronema*, a representative of
193 apothecium-forming Pezizomycotina. This species has 287 genes in the 273 Pezizomycotina-
194 specific families, whereas other species have 401–849 genes. Given that *Pyronema's* fruiting
195 bodies probably reflect the ancestral morphology (apothecium) in the Pezizomycotina (*46–*
196 *49*), these figures could indicate that the developmental gene repertoire of *Pyronema*
197 resembles the ancestral condition in the Pezizomycotina.

Convergent expansions in shared developmental families

199 We next asked whether the expansions observed in two CM clades were composed of
200 expansions of the same gene families (i.e. convergent) or composed of expansions of
201 differing gene families. We calculated subphylum-specific gene duplication rates in the
202 Agarico- and Pezizomycotina, which we plotted against each other as shown on Fig. 2/d. Of
203 the 314 shared developmental families, 257 (81.8%) showed parallel expansions in the
204 Agarico- and Pezizomycotina. In contrast, only eight (2.5%) showed no duplications in either
205 class and 49 (15.6%) showed duplications in only one. If families that likewise contained at
206 least seven species but developmentally regulated genes from up to 2 species (1747 families)
207 were considered, the pattern was the opposite: only 145 (8.3%) showed parallel duplications
208 in the Agarico- and Pezizomycotina, 1602 (91.7%) showed no duplications or in only one of
209 the subphyla (Fig. 2/e). Convergent expansion in the 314 shared developmental families

210   correlates with the number of species represented by developmentally regulated genes in the
211   family (Fig. 2/f). Convergence in shared developmental families is significantly more
212   abundant than in any other combination of gene families we tested ($P < 1.2 \times 10^{-9}$, Fisher's
213   exact test), including controls for gene family size and the number of developmentally
214   regulated proteins per family, among others (table S4, fig. S5). The accelerated duplication
215   rate in these families differs considerably from that of other families (table S4), or families
216   with constraint on duplication imposed by a fitness cost of increased dosage (45) (Note S3,
217   Fig 2/g), collectively suggesting that the convergent gene family expansions were driven by
218   positive selection. Families with clade-specific developmental expression, on the other hand,
219   did not show signs of convergent expansion (Fig. S6). We note that further convergent
220   expanding can certainly be found in groups with <7 species, which renders our estimate of
221   convergence conservative.
222   We also examined the extent of convergence in amino acid sites among CM Agarico-
223   and Pezizomycotina, using approaches that incorporate null models (50) proposed in response
224   to previous criticisms of published cases (51–53). We found 129 families in which
225   convergent shifts in amino acid preference are significantly enriched relative to control
226   analyses (Fig. S7, note S2). Developmentally regulated genes are enriched in 28 of these
227   (Note S2), including genes related to cell division and DNA repair, splicing, ergosterol
228   biosynthesis, among others (see table S5). Nevertheless, the extent of convergence in CM
229   clades was overall similar to that observed in other combinations of clades (note S2), which
230   could indicate that the extent of amino acid convergence in CM is either not outstanding in
231   CM fungi or that other, unknown traits drove convergence also in non-CM clades.
232   The observed extent of convergence in gene family co-option and expansion exceeds
233   expectations based on previous predictions (54) or examples (55–60) at this phylogenetic
234   scale. In closely related populations of the same species or sister species, evolution works
235   with the same standing genetic variation, providing for a higher incidence of (potentially non-
236   adaptive) genetic parallelism (3, 61). Because the probability of repeated recruitment of genes
237   declines with phylogenetic distance, much less convergence is expected among distantly
238   related clades that diverged in the architecture of gene regulatory networks, even if the genes
239   themselves are conserved. Molecular clock estimates suggest that the Agarico- and
240   Pezizomycotina diverged >650 million years ago and their ancestors existed >270 myr after
241   the Dikarya ancestor (62, 63). Because of this deep divergence, the extent of parallel co-
242   option and convergent diversification of developmental families in CM Agarico- and
243   Pezizomycotina is not explainable by phylogenetic proximity or neutral processes alone.

Discussion

244   **Discussion**

245   Our genome-wide analyses revealed extensive parallel co-option of ancient genes and
246   convergent gene family expansions in two complex multicellular clades of fungi. We
247   observed molecular convergence in hundreds of gene families, with ~82% of shared
248   developmentally regulated families showing convergent expansions. Several recent studies
249   suggested that molecular convergence may be widespread in nature (51, 58, 60). However,
250   while most previous examples were restricted to a few genes (55, 60, 64, 65) or to closely

251 related species (*61, 66*), our results suggest that molecular convergence can be pervasive in
252 clades separated by >650 million years of evolution and can affect hundreds of gene families.
253 The repeated emergence of CM in fungi suggests that evolution can be deterministic,
254 which, in the context of Gould's famous thought experiment (*67*), means that if we replayed
255 life's tape, CM would again evolve in fungal clades. This predictability has been attributed to
256 shared genetic variation (*10*), similar selective regimes (*5, 8, 11*) or constraints on the array of
257 acceptable changes (*8*) and on how novelty arises (*4, 8*). A special case of bias in the
258 emergence of novelty is when genes with suitable biochemical properties are available in
259 ancestral species and can easily be co-opted for the same functionalities. Our results are
260 compatible with the scenario that ancient fungi have been predisposed for evolving CM by a
261 rich repertoire of genes in the Dikarya ancestor that are used by extant species for CM-related
262 functions. Predisposed lineages are more likely to show phenotypic convergence (*8, 23, 43,*
263 *68*), purely because of the availability of genetic tools that can be recruited for the same
264 functions. It follows that if predisposition indeed happened, then the repeated evolution of
265 CM is not as surprising as it may seem, given the availability of genetic mechanisms that are
266 crucial for the evolution of such multigenic phenotypes.
267 Already Haldane speculated that similar phenotypes emerge not only as a result of
268 similar selection pressures but also as a result of shared genetic biases (*69*). There is probably
269 a finite number of ways by which CM-associated functions, such as cell adhesion,
270 communication or differentiation can evolve, which explains why the same gene families
271 were co-opted and started diversifying in complex multicellular Agarico- and
272 Pezizomycotina. Our study provides an example on how the genomic repertoire may channel
273 phenotypic evolution towards similar solutions and how this can lead to extensive genetic
274 convergence even at large phylogenetic scales. Such genetic biases on phenotypic evolution
275 suggest that the tireless tinkering of evolution is not only limited by the environment, but also
276 by the genetic ingredients at hand.
277

References

279 1. C. Darwin, E. Mayr, *On the origin of species* (Harvard University Press, 1995;
280   http://www.hup.harvard.edu/catalog.php?isbn=9780674637528).
281 2. N. Shubin, C. Tabin, S. Carroll, Deep homology and the origins of evolutionary
282   novelty. *Nature*. **457**, 818–823 (2009).
283 3. D. L. Stern, V. Orgogozo, Is Genetic Evolution Predictable? *Science*. **323**, 746–751
284   (2009).
285 4. G. R. McGhee, *Convergent evolution : limited forms most beautiful* (MIT Press, 2011).
286 5. Z. D. Blount, R. E. Lenski, J. B. Losos, Contingency and determinism in evolution:
287   Replaying life's tape. *Science*. **362**, eaam5979 (2018).
288 6. D. P. Rice, J. P. Townsend, A Test for Selection Employing Quantitative Trait Locus
289   and Mutation Accumulation Data (2012), doi:10.1534/genetics.111.137075.
290 7. B. Park *et al.*, Distributions of Mutational Effects and the Estimation of Directional
291   Selection in Divergent Lineages of Arabidopsis thaliana. *Genetics*. **206**, 2105–2117
292   (2017).
293 8. J. B. Losos, Convergence, adaptation, and constraint. *Evolution (N. Y)*. **65**, 1827–1840
294   (2011).

295 9. H. D. Rundle, L. Nagel, J. W. Boughman, D. Schluter, Natural Selection and Parallel
296   Speciation in Sympatric Sticklebacks. *Science*. **287**, 306–308 (2000).
297 10. M. Muschick, A. Indermaur, W. Salzburger, Convergent Evolution within an Adaptive
298   Radiation of Cichlid Fishes. *Curr. Biol.* **22**, 2362–2368 (2012).
299 11. D. Luke Mahler, J. B. L. Travis Ingram, Liam J. Revell, Exceptional Convergence on
300   the Macroevolutionary Landscape in Island Lizard Radiations. *Science* (2013),
301   doi:10.1126/science.1239431.
302 12. R. van Velzen *et al.*, Comparative genomics of the nonlegume Parasponia reveals
303   insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci.*
304   *U. S. A.* **115**, E4700–E4709 (2018).
305 13. M. S. Pankey, V. N. Minin, G. C. Imholte, M. A. Suchard, T. H. Oakley, Predictable
306   transcriptome evolution in the convergent and complex bioluminescent organs of
307   squid. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4736-42 (2014).
308 14. J. B. Losos, Contingency and Determinism in Replicated Adaptive Radiations of
309   Island Lizards. *Science*. **279**, 2115–2118 (1998).
310 15. R. K. Grosberg, R. R. Strathmann, The Evolution of Multicellularity: A Minor Major
311   Transition? *Annu. Rev. Ecol. Evol. Syst.* **38**, 621–654 (2007).
312 16. A. Rokas, The Origins of Multicellularity and the Early History of the Genetic Toolkit
313   For Animal Development. *Annu. Rev. Genet.* **42**, 235–251 (2008).
314 17. A. H. Knoll, The Multiple Origins of Complex Multicellularity. *Annu. Rev. Earth*
315   *Planet. Sci.* **39**, 217–239 (2011).
316 18. D. J. Dickinson, W. J. Nelson, W. I. Weis, An epithelial tissue in Dictyostelium
317   challenges the traditional origin of metazoan multicellularity. *BioEssays*. **34**, 833–840
318   (2012).
319 19. I. Ruiz-Trillo *et al.*, The origins of multicellularity: a multi-taxon genome initiative.
320   *Trends Genet.* **23**, 113–118 (2007).
321 20. D. Claessen, D. E. Rozen, O. P. Kuipers, L. Søgaard-Andersen, G. P. van Wezel,
322   Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies.
323   *Nat. Rev. Microbiol.* **12**, 115–124 (2014).
324 21. J. C. Coates, U.-E. Aiman, B. Charrier, Understanding green multicellularity: do
325   seaweeds hold the key? *Front. Plant Sci.* **5**, 737 (2015).
326 22. L. G. Nagy, G. M. Kovács, K. Krizsán, Complex multicellularity in fungi:
327   evolutionary convergence, single origin, or both? *Biol. Rev.* **93**, 1778–1794 (2018).
328 23. L. G. Nagy, Many roads to convergence. *Science*. **361**, 125–126 (2018).
329 24. A. Sebé-Pedrós, B. M. Degnan, I. Ruiz-Trillo, The origin of Metazoa: a unicellular
330   perspective. *Nat. Rev. Genet.* **18**, 498–512 (2017).
331 25. J. M. Smith, S. Eörs, *The Major Transitions in Evolution* (New York:W. H. Freeman
332   and Company, 1995).
333 26. K. M. Lord, N. D. Read, Perithecium morphogenesis in Sordaria macrospora. *Fungal*
334   *Genet. Biol.* **48**, 388–399 (2011).
335 27. D. S. Hibbett, After the gold rush, or before the flood? Evolutionary morphology of
336   mushroom-forming fungi (Agaricomycetes) in the early 21st century. *Mycol. Res.* **111**,
337   1001–1018 (2007).
338 28. F. Trail, Z. Wang, K. Stefanko, C. Cubba, J. P. Townsend, The ancestral levels of
339   transcription and the evolution of sexual phenotypes in filamentous fungi. *PLOS*
340   *Genet.* **13**, e1006867 (2017).
341 29. S. Pöggeler, M. Nowrousian, I. Teichert, A. Beier, U. Kück, in *Physiology and*
342   *Genetics* (Springer International Publishing, Cham, 2018;
343   http://link.springer.com/10.1007/978-3-319-71740-1_1), pp. 1–56.
344 30. K. Krizsan *et al.*, Transcriptomic atlas of mushroom development highlights an

345   independent origin of complex multicellularity. *bioRxiv*, 349894 (2018).

346   31.   T. A. Nguyen *et al.*, Innovation and constraint leading to complex multicellularity in
347   the Ascomycota. *Nat. Commun.* **8**, 14444 (2017).

348   32.   L. G. Nagy, Evolution: Complex Multicellular Life with 5,500 Genes. *Curr. Biol.* **27**,
349   R609–R612 (2017).

350   33.   U. R. Sikhakolli *et al.*, Transcriptome analyses during fruiting body formation in
351   Fusarium graminearum and Fusarium verticillioides reflect species life history and
352   ecology. *Fungal Genet. Biol.* **49**, 663–673 (2012).

353   34.   I. Teichert, G. Wolff, U. Kück, M. Nowrousian, "Combining laser microdissection and
354   RNA-seq to chart the transcriptional landscape of fungal development" (2012), ,
355   doi:10.1186/1471-2164-13-511.

356   35.   S. Traeger *et al.*, The Genome and Development-Dependent Transcriptomes of
357   Pyronema confluens: A Window into Fungal Evolution. *PLoS Genet.* **9**, e1003820
358   (2013).

359   36.   Z. Wang *et al.*, Global Gene Expression and Focused Knockout Analysis Reveals
360   Genes Associated with Fungal Fruiting Body Development in Neurospora crassa.
361   *Eukaryot. Cell*. **13**, 154–169 (2014).

362   37.   Y.-J. Park *et al.*, Whole Genome and Global Gene Expression Analyses of the Model
363   Mushroom Flammulina velutipes Reveal a High Capacity for Lignocellulose
364   Degradation. *PLoS One*. **9**, e93560 (2014).

365   38.   H. Muraguchi *et al.*, Strand-Specific RNA-Seq Analyses of Fruiting Body
366   Development in Coprinopsis cinerea. *PLoS One*. **10**, e0141586 (2015).

367   39.   J. Zhang *et al.*, Transcriptome Analysis and Its Application in Identifying Genes
368   Associated with Fruiting Body Development in Basidiomycete Hypsizygus
369   marmoreus. *PLoS One*. **10**, e0123025 (2015).

370   40.   Y. Sakamoto, K. Nakade, N. Konno, Endo-β-1,3-Glucanase GLU1, from the Fruiting
371   Body of Lentinula edodes, Belongs to a New Glycoside Hydrolase Family. *Appl.*
372   *Environ. Microbiol.* **77**, 8350–8354 (2011).

373   41.   P.-A. Christin, D. M. Weinreich, G. Besnard, Causes and evolutionary significance of
374   genetic convergence. *Trends Genet.* **26**, 400–405 (2010).

375   42.   Z. Wang, P. R. Johnston, Z. L. Yang, J. P. Townsend, Evolution of Reproductive
376   Morphology in Leaf Endophytes. *PLoS One*. **4**, 4246 (2009).

377   43.   L. G. Nagy *et al.*, Latent homology and convergent regulatory evolution underlies the
378   repeated emergence of yeasts. *Nat. Commun.* **5**, 4471 (2014).

379   44.   M. Griesmann *et al.*, Phylogenomics reveals multiple losses of nitrogen-fixing root
380   nodule symbiosis. *Science*. **361**, eaat1743 (2018).

381   45.   R. Sopko *et al.*, Mapping Pathways and Phenotypes by Systematic Gene
382   Overexpression. *Mol. Cell*. **21**, 319–330 (2006).

383   46.   Y. J. Liu, B. D. Hall, "Body plan evolution of ascomycetes, as inferred from an RNA
384   polymerase II phylogeny" (2004), (available at
385   www.pnas.orgcgidoi10.1073pnas.0400938101).

386   47.   J. W. Spatafora *et al.*, "A five-gene phylogeny of Pezizomycotina" (# 2006 by The
387   Mycological Society of America, 2006), (available at
388   http://www.fieldmuseum.org/myconet).

389   48.   C. L. Schoch *et al.*, The Ascomycota Tree of Life: A Phylum-wide Phylogeny
390   Clarifies the Origin and Evolution of Fundamental Reproductive and Ecological Traits.
391   *Syst. Biol.* **58**, 224–239 (2009).

392   49.   N. Zhang, Z. Wang, J. W. McLaughlin, David, Spatafora, Ed. (Springer, Second.,
393   2015), pp. 57–88.

394   50.   C. Rey, L. Guéguen, M. Sémon, B. Boussau, Accurate Detection of Convergent

395   Amino-Acid Evolution with PCOC. *Mol. Biol. Evol.* **35**, 2296–2306 (2018).

396 51. Z. Zou, J. Zhang, No Genome-Wide Protein Sequence Convergence for Echolocation.
397   *Mol. Biol. Evol.* **32**, 1237–1241 (2015).

398 52. Z. Zou, J. Zhang, Are Convergent and Parallel Amino Acid Substitutions in Protein
399   Evolution More Prevalent Than Neutral Expectations? *Mol. Biol. Evol.* **32**, 2085–2096
400   (2015).

401 53. J. F. Storz, Causes of molecular convergence and parallelism in protein evolution. *Nat.*
402   *Rev. Genet.* **17**, 239–250 (2016).

403 54. G. L. Conte, M. E. Arnegard, C. L. Peichel, D. Schluter, The probability of genetic
404   parallelism and convergence in natural populations. *Proc. R. Soc. B Biol. Sci.* **279**,
405   5039–5047 (2012).

406 55. A. L. Hughes, R. Friedman, Parallel evolution by gene duplication in the genomes of
407   two unicellular fungi. *Genome Res.* **13**, 794–9 (2003).

408 56. Y. Zhen, M. L. Aardema, E. M. Medina, M. Schumer, P. Andolfatto, Parallel
409   Molecular Evolution in an Herbivore Community. *Science.* **337**, 1634–1637 (2012).

410 57. T. A. Castoe *et al.*, Evidence for an ancient adaptive episode of convergent molecular
411   evolution. *PNAS.* **106**, 8986–8991 (2009).

412 58. A. Rokas, S. B. Carroll, Frequent and Widespread Parallel Evolution of Protein
413   Sequences. *Mol. Biol. Evol.* **25**, 1943–1953 (2008).

414 59. J. Parker *et al.*, Genome-wide signatures of convergent evolution in echolocating
415   mammals. *Nature.* **502**, 228–231 (2013).

416 60. Y.-Y. Shen, L. Liang, G.-S. Li, R. W. Murphy, Y.-P. Zhang, Parallel Evolution of
417   Auditory Genes for Echolocation in Bats and Toothed Whales. *PLoS Genet.* **8**,
418   e1002788 (2012).

419 61. P. F. Colosimo *et al.*, Widespread Parallel Evolution in Sticklebacks by Repeated
420   Fixation of Ectodysplasin Alleles. *Science.* **307**, 1928–1933 (2005).

421 62. D. Floudas *et al.*, The Paleozoic Origin of Enzymatic Lignin Decomposition
422   Reconstructed from 31 Fungal Genomes. *Science.* **336**, 1715–1719 (2012).

423 63. A. Kohler *et al.*, Convergent losses of decay mechanisms and rapid turnover of
424   symbiosis genes in mycorrhizal mutualists. *Nat. Genet.* **47**, 410–415 (2015).

425 64. D. M. Emms, S. Covshoff, J. M. Hibberd, S. Kelly, Independent and Parallel Evolution
426   of New Genes by Gene Duplication in Two Origins of C4 Photosynthesis Provides
427   New Insight into the Mechanism of Phloem Loading in C4 Species. *Mol. Biol. Evol.*
428   **33**, 1796–1806 (2016).

429 65. M. Wirthlin *et al.*, Parrot Genomes and the Evolution of Heightened Longevity and
430   Cognition. *Curr. Biol.* **28**, 4001–4008.e7 (2018).

431 66. K. R. Elmer *et al.*, Parallel evolution of Nicaraguan crater lake cichlid fishes via non-
432   parallel routes. *Nat. Commun.* **5**, 5168 (2014).

433 67. S. J. Gould, *Wonderful life : the Burgess Shale and the nature of history* (WW Norton
434   & Company, New York, 1990; https://www.amazon.com/Wonderful-Life-Burgess-
435   Nature-History/dp/039330700X).

436 68. A. A. Agrawal, Toward a Predictive Framework for Convergent Evolution: Integrating
437   Natural History, Genetic Mechanisms, and Consequences for the Diversity of Life.
438   *Am. Nat.* **190**, S1–S12 (2017).

439 69. Haldane J. B. S, *The causes of evolution* (Harper and Brothers, London, 1932).

440 70. S. Andrews, FastQC: a quality control tool for high throughput sequence data.
441   *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/* (2010), (available at
442   http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

443 71. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina
444   sequence data. *Bioinformatics.* **30**, 2114–2120 (2014).

11

445 72. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq
446   quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
447 73. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome
448   comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**,
449   157 (2015).
450 74. P. Jones *et al.*, InterProScan 5: genome-scale protein function classification.
451   *Bioinformatics.* **30**, 1236–40 (2014).
452 75. E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: a tool for discovery
453   and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics.* **10**,
454   48 (2009).
455 76. K. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version 7:
456   Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
457 77. S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated
458   alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* **25**, 1972–
459   1973 (2009).
460 78. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of
461   large phylogenies. *Bioinformatics.* **30**, 1312–3 (2014).
462 79. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in
463   R language. *Bioinformatics.* **20**, 289–290 (2004).
464 80. C. A. Darby, M. Stolzer, P. J. Ropp, D. Barker, D. Durand, Xenolog classification.
465   *Bioinformatics.* **33**, btw686 (2016).
466 81. H. V Colot *et al.*, A high-throughput gene knockout procedure for Neurospora reveals
467   functions for multiple transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10352–
468   10357 (2006).
469

482 **Supplementary Materials:**

483 Materials and Methods

484 Figures S1-S9

485 Tables S1 and S3 to S9

486 References (*70-81*)

487

Supplementary Materials for

Unmatched level of molecular convergence among deeply divergent complex multicellular fungi

Zsolt Merényi, Arun N. Prasanna, Wang Zheng, Károly Kovács, Botond Hegedüs, Balázs Bálint, Balázs Papp, Jeffrey P. Townsend, László G. Nagy

Correspondence to: lnagy@fungenomelab.com

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

13

Materials and Methods

Bioinformatic analysis of RNA-Seq data

We downloaded publicly available transcriptome data related to different developmental stages of nine fruiting body (FB) forming species from the Pezizomycotina (*Fusarium graminearum*, *F. verticillioides*, *Sordaria macrospora*, *Neurospora crassa*, *Pyronema confluens*) and the Agaricomycotina (*Armillaria ostoyae*, *Coprinopsis cinerea*, *Hypsizygus marmoreus*, *Flammulina velutipes*, table S1). We ran a quality check on raw fastq files using fastqc v0.11.5 (*70*) and trimmed the adapter sequences and low-quality bases with trimmomatic v0.36 (*71*). Next, we used kallisto v0.43.1 (*72*) to quantify the abundance of transcripts for each stage. Specifically, we utilized the estimated counts from abundance data to calculate Fragments Per Kilobase Million (FPKM) and used it as the quantification metric. As a pre-filter, an FPKM value less than two was considered insignificant. The homogeneity of biological replicates was checked by constructing MDS plots based on overall expression levels.

Identification of Developmentally Regulated Genes

The RNA-Seq data comprised 2-13 stages, which was used to identify developmentally regulated genes: those that show at least four-fold change in expression between any two fruiting body stages or tissue types and that show an expression level FPKM > 4. Fold change values were calculated for all biologically relevant pairwise comparisons (for see details fig. S1).

Comparative genomic approaches

In addition to the nine above mentioned fruiting body forming species, 10 additional species were included in the analysis for comparative purposes (table S1). This 19 genome dataset was clustered into gene families using OrthoFinder v1.1.8 (*73*) with the default inflation parameter of 1.5 to facilitate interspecies comparison.

For functional annotation of genes and gene families InterProscan search was performed with InterProscan version 5.28-67.0 (*74*) across the 19 fungal proteomes (table S3).

Gene Ontology (GO) enrichment analysis for yeast orthologs was performed using GOrilla ((*75*) http://cbl-gorilla.cs.technion.ac.il/) with *Saccharomyces cerevisiae* as the reference organism, a 10^{-3} P-value threshold and false discovery rate correction for multiple testing. Terms in all three ontologies (Biological process, Cellular component, Molecular function) were considered. Experimentally verified gene function from *Aspergillus nidulans* and *Neurospora crassa* were also considered during the functional annotation of developmentally regulated gene families. We also used known developmentally regulated gene set of *Coprinopsis* from literature to verify the efficiency of our designation.

Phylogenetic analyses and ancestral state reconstructions

Altogether 86 clusters were single copy and shared by all 19 species; these clusters were used to reconstruct a species tree. After multiple sequence alignment using the L-INS-I algorithm of MAFFT (*76*) and trimming with trimAL (−gt 0.6) (*77*) sequences were concatenated into a supermatrix and used for phylogenetic reconstruction in raxmlHPC-PTHREADS-SSE3 (*78*). The supermatrix was partitioned by gene and the PROTGAMMAWAG model was used with 100 rapid bootstrap replicates, to estimate branch support.

To reconstruct the evolution of fruiting body formation, maximum likelihood ancestral state reconstructions were performed with the ace (ancestral character estimations) function of the ape R package ((*79*); R Development Core Team, 2018). As the more

14

561 parametrized ARD (all-rates-different) model didn't yield significantly greater likelihoods,
562 we used the ER (equal rates) model.
563
564 <u>Evolutionary</u> <u>history</u> <u>of</u> <u>gene</u> <u>families</u>
565 To reconstruct the duplication and loss events of gene families across the species tree, the
566 COMPARE pipeline (*43*) was used. For this analysis gene trees were reconstructed for
567 clusters containing at least four proteins (9686 clusters). Sequences in each cluster were
568 aligned using the L-INS-I method of MAFFT and trimmed with trim-AL (-gt 0.2). Gene tree
569 reconstructions were performed in RAxML under the PROTGAMMAWAG model with 100
570 rapid bootstrap replicates. Gene trees were rerooted and reconciled with the species tree using
571 Notung 2.9 (*80*) with 80% bootstrap support as the edge-weight threshold for topological
572 rearrangements. After ortholog coding, duplications and losses for each orthogroup were
573 mapped onto the species tree using Dollo parsimony (*43*). The visualization of reconstructed
574 duplication/loss histories and further statistical analyses (Fisher Exact test) were performed
575 with custom R scripts (available from the authors upon request).
576   To quantify convergent gene family expansions, we filtered families with the
577 following criteria: a) a gene family has genes conserved in ≥7 of the 9 Dikarya species, ≥4 of
578 the 5 Pezizomycotina species or ≥3 of the 4 Agaricomycotina species or b) a cluster has
579 developmental expression conserved in ≥7 of the 9 Dikarya species, ≥4 of the 5
580 Pezizomycotina species or ≥3 of the 4 Agaricomycotina species.
581
582 <u>Convergence</u> <u>in</u> <u>gene</u> <u>family</u> <u>expansions</u>
583 To quantify the level of convergent gene family expansions, subphylum-specific gene
584 duplication rates were compared between the Agarico- and Pezizomycotina. Rates were
585 calculated by normalizing the raw number of inferred duplications for a given node by both
586 the length of the preceding branch and by gene family size, because along longer branches
587 and in larger gene families the probability of duplications is naturally higher. After this
588 correction step, duplication rates were averaged across nodes (gene duplication rate for
589 Agarico- and Pezizomycotina) and plotted using custom R scripts. The numbers of four
590 possible events were recorded: duplications in only one (Agarico- or Pezizomycotina), both
591 or none of the CM clades.
592 To assess if developmental gene families show more or less convergence than expected by
593 chance, different control groups of gene families were generated and compared using Fisher's
594 exact test. Control families were always compared to the conserved developmentally
595 regulated cluster set (developmentally expression conserved in ≥7 of the 9 Dikarya species).
596 The first control groups comprised families that similarly contained ≥7 species but only 0-2
597 (1747 cluster) or 3-6 species (2052 cluster) with developmental expression. Next, to test if
598 gene family size (i.e. number of proteins) impacts convergence, we also generated control
599 groups with similar gene family size distribution but containing less developmentally
600 regulated genes than the 314 shared developmental gene families. A custom R script was
601 used to find a non-developmentally regulated gene family for each of the developmentally
602 regulated gene family with a matching size one by one. If it was not possible to find a gene
603 family with similar size (permitted maximum difference of 10%), the gene family was
604 excluded from the comparison. If there were more than one gene families with the same size,
605 the one with most similar species composition and least developmentally regulated genes
606 (according to number of species represent developmentally regulated genes) was chosen.
607
608 <u>Detecting</u> <u>convergent</u> <u>amino</u> <u>acid</u> <u>changes</u>
609 In order to gain insights into amino acid convergence between the Agaricomycotina and
610 Pezizomycotina, we followed Rey et al.'s approach (*50*) to identify convergent shifts in amino

15

611 acid preference at a given site. Convergence is defined not only as changes to identical amino
612 acids from different ancestral states, but also as changes to amino acids with similar
613 biochemical properties (referred to as convergent shifts in amino acid composition). We
614 identified such shifts across all gene families in the 19 species' genomes using the model
615 "Profile Change with One Change" (PCOC) (50) (downloaded from
616 https://github.com/CarineRey/pcoc on 2018.11.05). We used reconciled gene trees with
617 branch lengths re-estimated with RAxML (raxmlHPC-PTHREADS-SSE3) as input. Each of
618 the most inclusive clades that contained only CM species were designated as phenotypically
619 convergent clade. Automated designation of convergent clades was done using a custom R
620 script, followed by execution of PCOC with default settings. For considering a site as
621 convergent we chose the PCOC model with a posterior probability threshold of 0.8. We
622 performed the analysis on 3799 gene families that contained at least 10 proteins and at least
623 one protein from both the Agaricomycotina and the Pezizomycotina. Three sets of control
624 analyses were run assess the amount of convergence caused by chance events. In control 1,
625 the basal lineages of Ascomycota and Basidiomycota were designated as convergent clades
626 (Ustilagomycotina, Pucciniomycotina, Saccharomycotina, Taphrinomycotina). In control 2,
627 CM Agaricomycotina species were paired with the basal clades of Ascomycota
628 (Saccharomycotina, Taphrinomycotina) while in control 3 CM Pezizomycotina were paired
629 with the basal clades of Basidiomycota (Ustilagomycotina, Pucciniomycotina). This resulted
630 in three control analyses in which CM is not shared by clades designated as convergent (note
631 however, that other traits might be). The numbers of detected amino acid sites showing
632 convergent shifts in each gene family were recorded and correlations between CM and
633 control groups were evaluated with a Pearson correlation test. We also compared these values
634 after correction by branch lengths between or in the designated clades to avoid the effect of
635 divergence (i.e. branch length) on the amount of amino acid changes.
636   We assumed that gene families which contain more convergent amino acid sites in
637 CM lineages than in non-CM clades might be involved in the shaping of convergent
638 phenotypes. For identification of these gene families, a linear model was fit to predict the
639 number of convergent sites between CM clades from the corresponding values of non-CM
640 clades (control 1). Gene families with more convergent sites than the upper limit of 95%
641 prediction interval of the linear model were considered as displaying significant number of
642 convergent sites in CM clades.

643 **Supplementary Text**
644 <u>Comparison of developmentally regulated genes with known developmental genes</u>
645   To validate our approach to identifying developmental genes, we compared our sets of
646 developmentally regulated genes to suites of genes known to be involved in fruiting body
647 development in *Aspergillus nidulans*, *Neurospora crassa* and *Coprinopsis cinerea*. First we
648 compared our dataset to that of (28). They ranked genes with the largest evolved difference in
649 gene expression change across perithecium development of 5 Ascomycota species (from
650 *Fusarium* and *Neurospora*). Of the 41 genes they found to have an aberrant phenotype in
651 sexual development of either *Neurospora crassa* or *Fusarium graminearum*, our analysis we
652 identified 40 (97.6%, table S2). Another study (81) identified 234 genes causing phenotypic
653 changes during fruiting body formation when knocked out in Neurospora, of which 63 were
654 detected as developmentally regulated (26.9%) in our study (table S2). Secondly, we
655 examined homologs of genes involved in the initiation and coordination of sexual
656 reproduction of *Aspergillus nidulans* (84). Homologs of these genes were identified in our
657 nine species based on best BLAST hits (e-value $< 10^{-6}$). Depending on the species, 4.6-69.2%
658 (table S2) of homolog genes were developmentally regulated in our dataset. The lowest ratio
659 is found in *Pyronema confluens* where only two mixed stage (both samples contains

660 vegetative mycelium) were sampled for transcriptomic analysis (*35*). We also collected genes
661 from the literature with a proven role in fruiting body development of *Coprinopsis cinerea*.
662 Our identified developmentally regulated genes contained 92.7% of known developmental
663 genes of *Coprinopsis* (table S2).
664
665 <u>Convergent amino acid shifts</u>
666   Altogether 3541 gene families were analyzed with PCOC (*50*) to identify convergent
667 amino acid shifts related to the convergent phenotype of complex multicellularity (CM).
668 When CM species were designated as convergent clades the gene trees of families had on
669 average 2.52±1.65 (max: 41) convergent clades (median: 2) and PCOC identified on average
670 10.23±10.6 sites as convergent (median: 7, max: 117). In the case of the control comparison
671 consisting of basal, non-CM lineages of the Asco- and Basidiomycota, the gene trees had on
672 average 4.03±2.08 (max: 66) convergent clades (median: 4) and on average 10.66±11.45 sites
673 were detected as convergent (median: 8, max: 123). The numbers of amino acid sites
674 identified by the PCOC model were compared between CM clades and control groups in each
675 of the gene families in order to decide whether the number of sites in CM species is higher
676 than that in the control clades, indicating adaptive amino acid convergence related to CM.
677 The number of detected amino acid sites showed strong correlation (Pearson r > 0.92, p-value
678 $< 2.2 \times 10^{-16}$) between the CM and control comparisons (Fig. S8), indicating similar levels of
679 amino acid convergence in CM and non-CM clades. This could be explained two ways. First,
680 it is possible that there isn't more amino-acid convergence in relation with CM than
681 background levels. Alternatively, unmeasured traits in the control clades could be associated
682 with similar amounts of adaptive amino acid convergence as CM, explaining their similar
683 levels in the two sets of clades. Despite different normalizations (with branch length between
684 or in the designated clades) (Fig. S8), the results did not generally support a higher-than-
685 expected contribution of amino acid convergence to the independent emergence of CM.
686 However, 129 gene families (3.64% of all examined) had more convergent sites than the
687 upper border of prediction intervals of a linear model, in comparison with the control analysis
688 ($R^2 = 0.85$, p-value $< 2.2 \times 10^{-16}$). From these 129 gene families only 28 contain more
689 developmentally regulated proteins than expected by chance (table S5). One of them
690 (OG0002402) is homologous with the *Aspergillus nimO/AN1779*, protein kinase, contained 7
691 developmentally regulated genes (out of 10 genes in total), and 25 convergently evolved
692 amino acid site between the two subphyla (in contrast: 13-22 in control groups, fig. S9). This
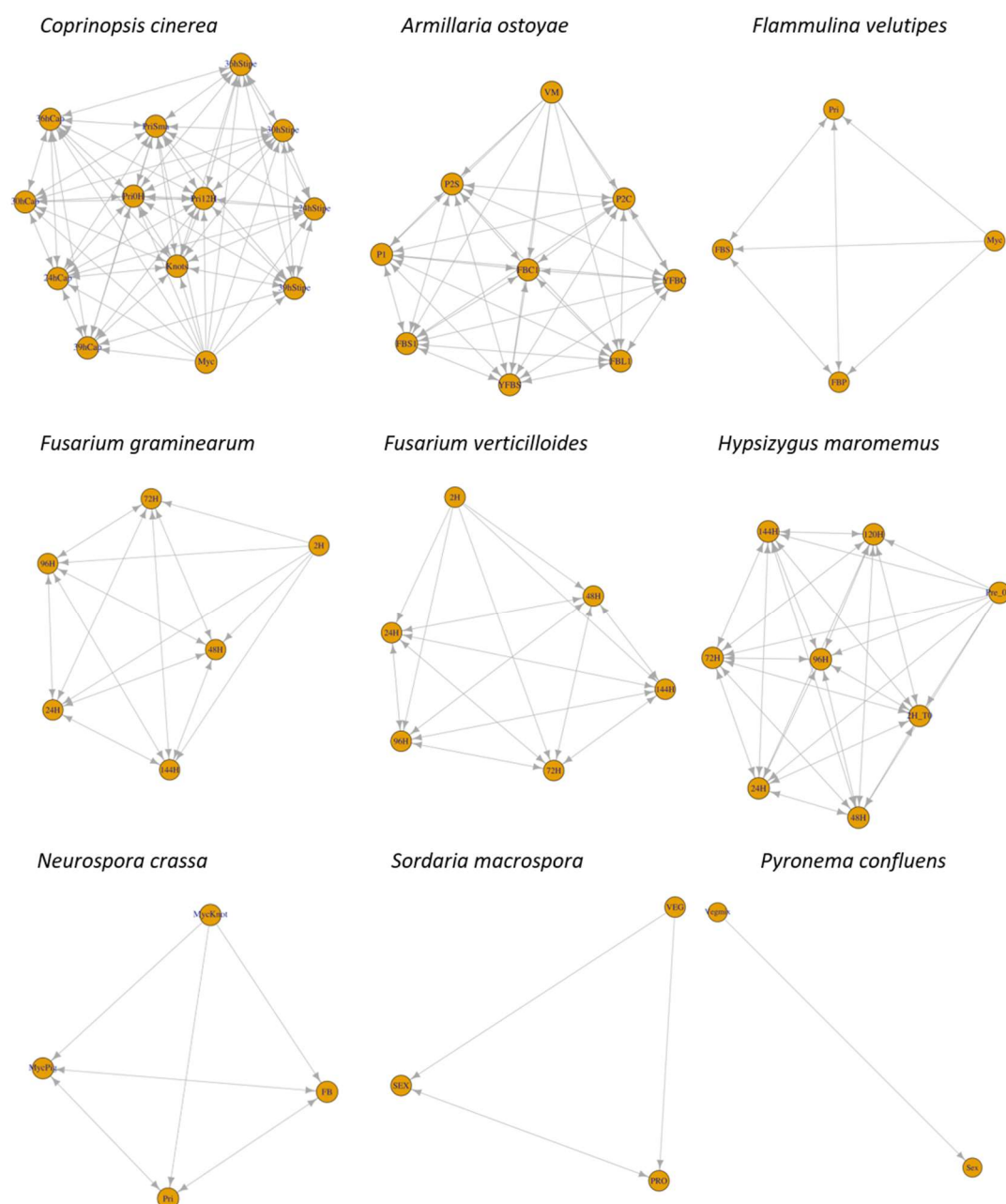693 gene family might be an example of amino acid convergence related to fruiting body
694 formation.
695
696 <u>Gene families with constraint</u>
697   We used ranked gene lists of *S. cerevisiae* from gene overexpression assays of Sopko
698 et al. (*45*), who analyzed and scored (toxicity score from lethal to wild type: 1-5) the growth
699 rate of strongly overexpressing each of 5917 genes in *Saccharomyces*. These overexpression
700 sensitive genes were identified in our dataset using BLAST with the *Saccharomyces* strain
701 S288C as a query (downloaded from
702 https://downloads.yeastgenome.org/sequence/S288C_reference/orf_protein/). After the
703 detection of best hits for each query protein, and an 80% identity filter, we could identify
704 unequivocally 3907 gene families in which all members of the family had the same toxicity
705 score. In 718 cases members of the gene families had different toxicity scores, in such cases
706 we accepted those toxicity scores which were supported by the most BLAST hits, or we took
707 into account the largest score. Finally, the duplication rates of 296 gene families with a
708 toxicity score <3 (most sensitive for gene overexpression in *Saccharomyces*) were compared
709 to the duplication rates of conserved developmental gene families (containing ≥7 of the 9 CM

17

710   species). In this comparison, the duplication rate of a gene family was averaged across all
711   nodes of the tree and calculated by normalizing the raw number of inferred duplications for a
712   given node by both the length of the preceding branch and gene family size.
713

714   **References**
715

716   70.   S. Andrews, FastQC: a quality control tool for high throughput sequence data.
717   *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/* (2010), (available at
718   http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).
719   71.   A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina
720   sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).
721   72.   N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq
722   quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
723   73.   D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome
724   comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**,
725   157 (2015).
726   74.   P. Jones *et al.*, InterProScan 5: genome-scale protein function classification.
727   *Bioinformatics*. **30**, 1236–40 (2014).
728   75.   E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: a tool for discovery
729   and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. **10**,
730   48 (2009).
731   76.   K. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version 7:
732   Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
733   77.   S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated
734   alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. **25**, 1972–
735   1973 (2009).
736   78.   A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of
737   large phylogenies. *Bioinformatics*. **30**, 1312–3 (2014).
738   79.   E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in
739   R language. *Bioinformatics*. **20**, 289–290 (2004).
740   80.   C. A. Darby, M. Stolzer, P. J. Ropp, D. Barker, D. Durand, Xenolog classification.
741   *Bioinformatics*. **33**, btw686 (2016).
742   81.   H. V Colot *et al.*, A high-throughput gene knockout procedure for Neurospora reveals
743   functions for multiple transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10352–
744   10357 (2006).
745

746

Fig. S1. Stage-wise comparisons used for identifying developmentally regulated genes, for each of the nine species. Developmental stages (nodes) and allowed comparisons (arrows) for the identification of developmentally regulated genes are shown. Arrows pointing in only one direction represent unidirectional comparisons, which were done in order to exclude genes showing highest expression in vegetative mycelium and no dynamics later on. Comparisons between tissue types were only allowed within the same developmental stage. H = hour, Pri = P = Primordia, S = Stipe, FB = Fruiting body, Y = Young, VEG= VM = Myc = Mycelium, C = Cap, L = Lamellae.

19
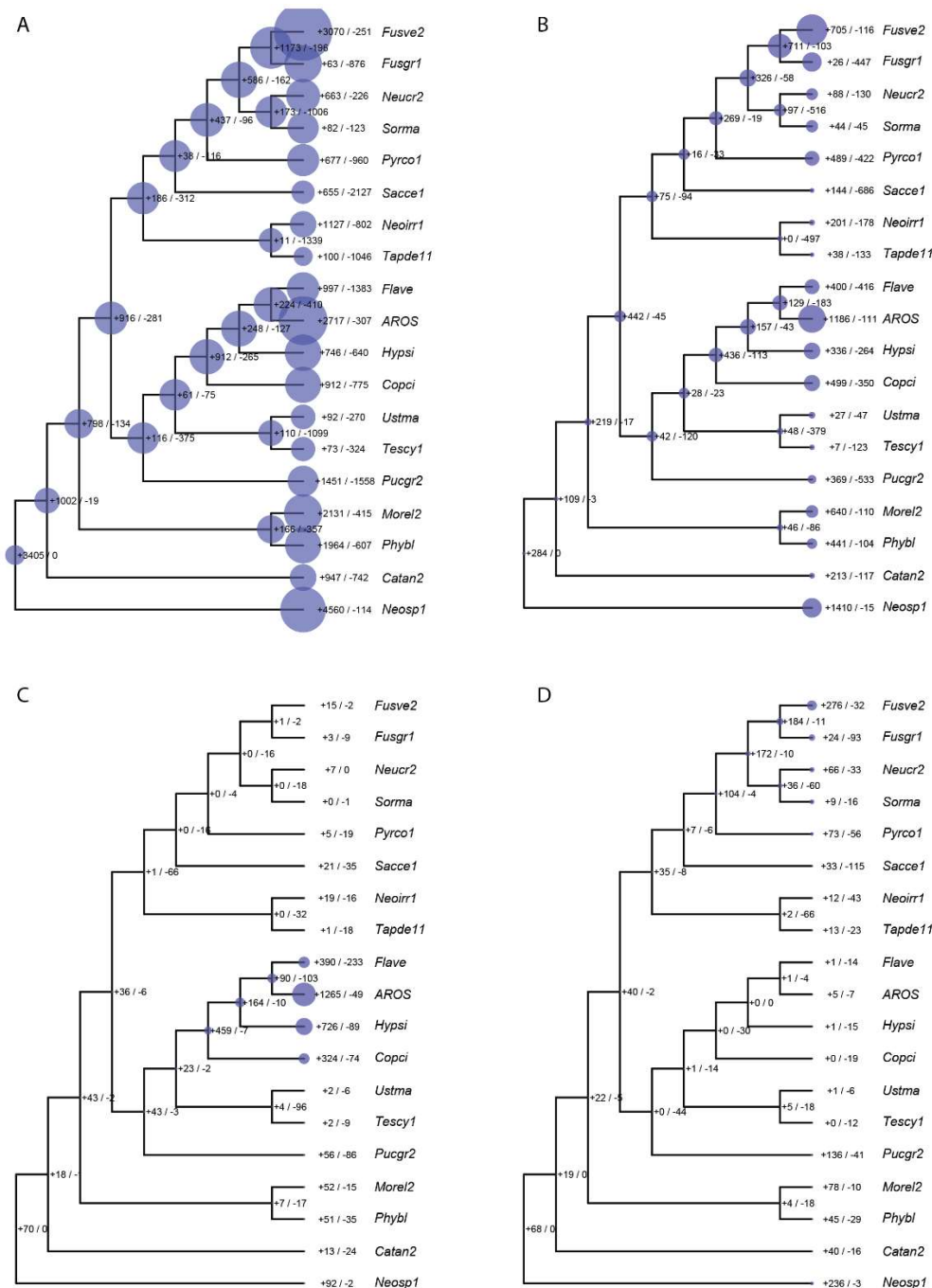
Fig. S3. Convergent expansion of developmentally regulated gene families in independent complex multicellular fungi. (A-C) Reconstructed copy number evolution of (A) 4113 shared gene families, (B) 314 shared developmentally regulated gene families, (C) 439 families with Agaricomycotina-specific developmental expression and (D) 273 families with Pezizomycotina-specific developmental expression. Bubble size proportional to the number of reconstructed ancestral gene copies across the analyzed families. Numbers next to internal nodes denote the number of inferred duplications (+) and losses (-).
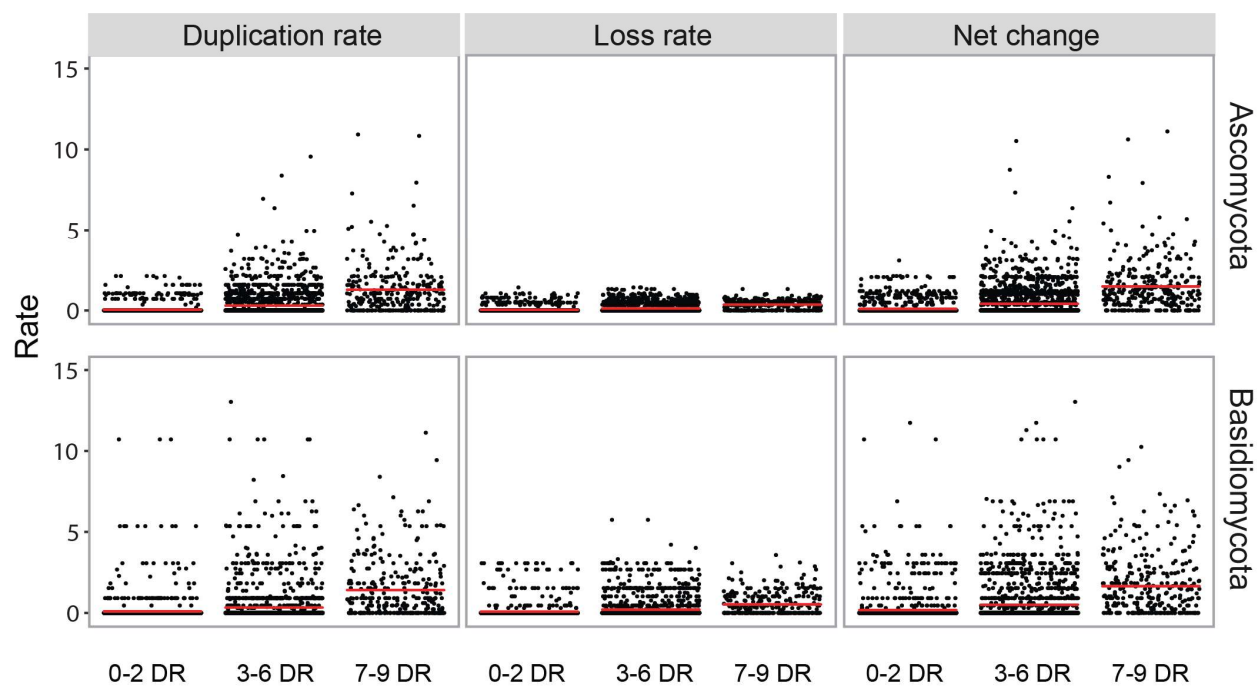
Fig. S4. Gene duplication, gene loss and net gene family expansion rates across gene families containing developmentally regulated genes from ≤2 (0-2DR), 3-6 (3-6DR) and ≥7 (7-9DR) species. Red lines mark mean values.

Fig. S5. Convergent gene family expansions in conserved Dikarya families containing 3-6 species with developmental expression. Each dot represents a family, while the x and y axes represent the duplication rate in the Agaricomycotina and Pezizomycotina, respectively.
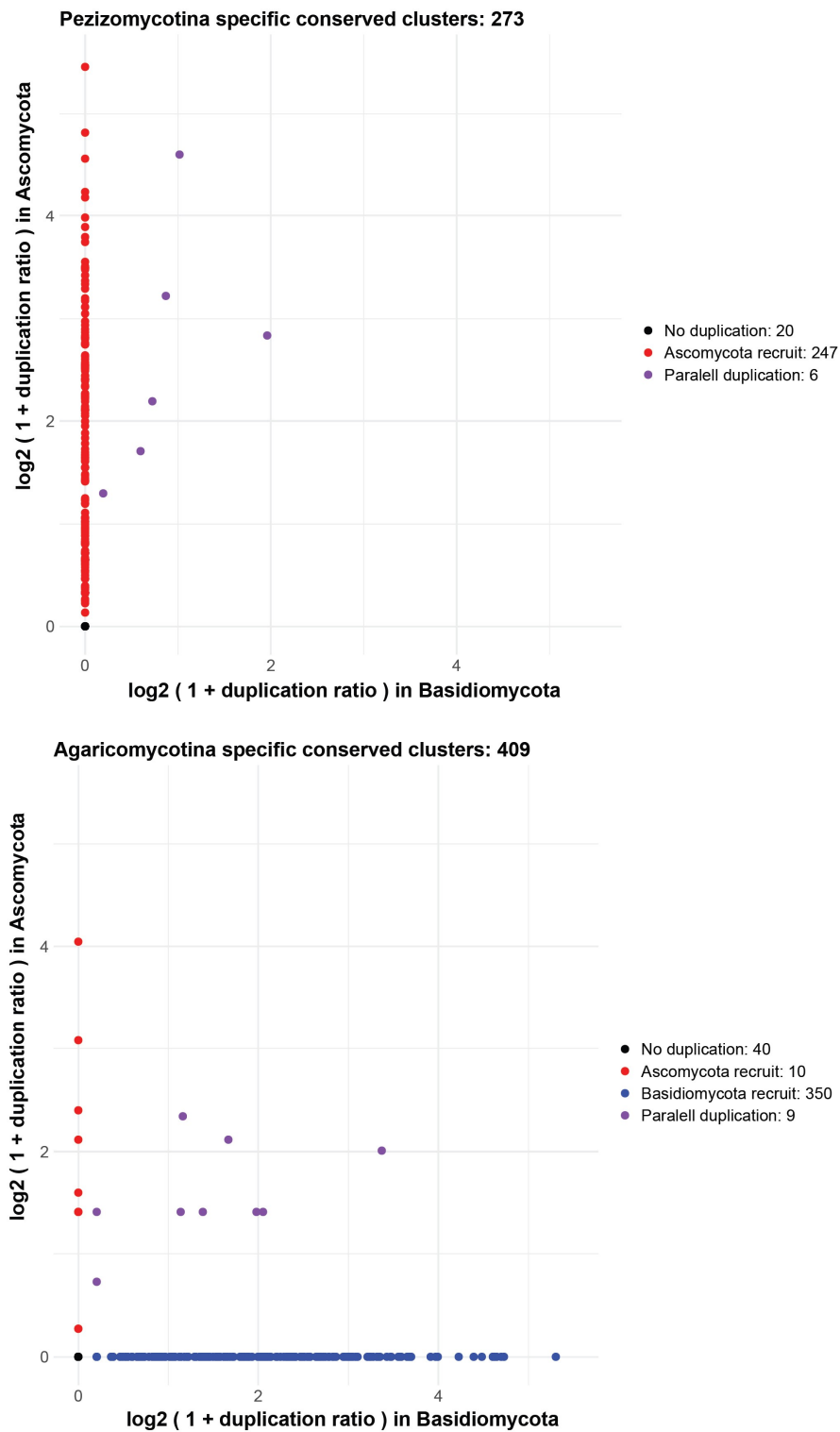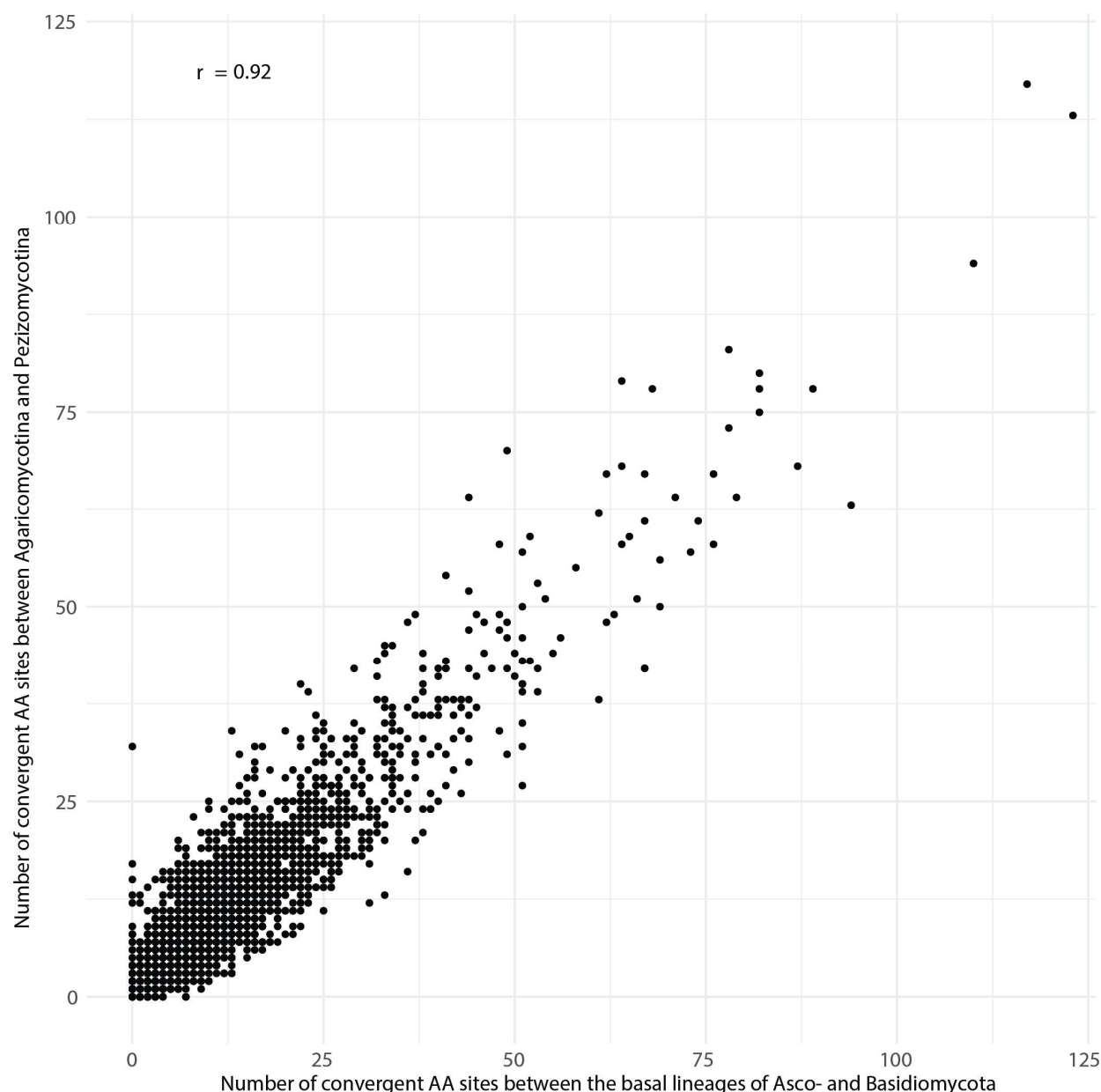
Fig. S6. Convergent gene family expansions in gene families with developmental expression specific to the Pezizomycotina (top) or the Agaricomycotina (bottom). Each dot represents a family, while the x and y axes represent duplication rate in Agaricomycotina and Pezizomycotina, respectively.

23

779

780

Fig. S7. Correlation between the numbers of convergent amino acid (AA) sites identified by the PCOC model in CM clades and in basal Asco- and Basidiomycota clades, as controls groups, across 3541 gene families. Each dot corresponds to a gene family tested for amino acid convergence. The y-axis represents the number of convergent AA sites detected among the Agaricomycotina and Pezizomycotina, while the x-axis represents a control estimate basal lineages of Asco- and Basidiomycota. r denotes Pearson correlation coefficient.

787
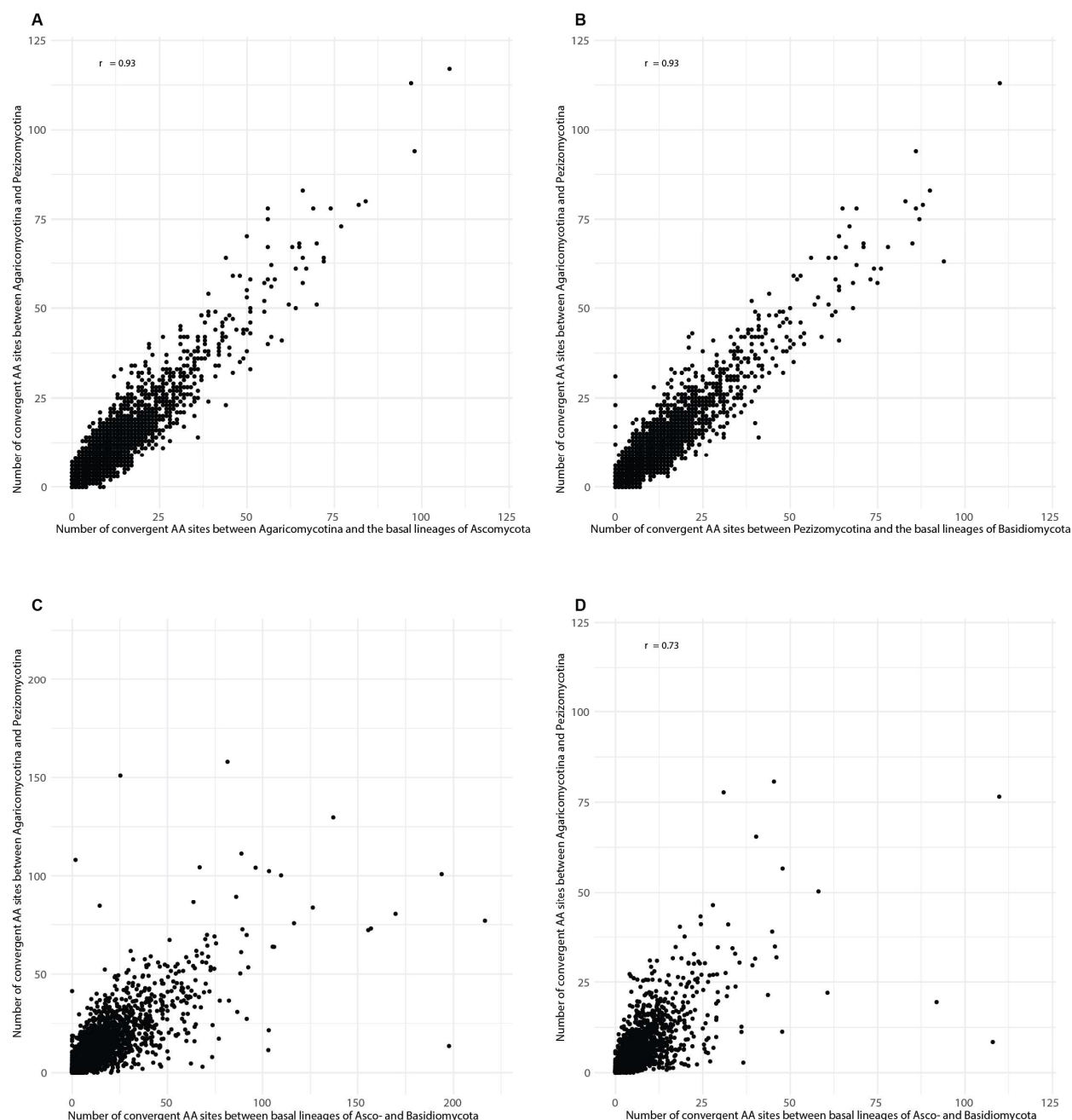
Fig. S8. Numbers of convergent amino acid (AA) sites identified by the PCOC model. In each scatterplot the y-axis represents the number of convergent AA sites detected among the fruiting body forming clades (Agaricomycotina and Pezizomycotina), while the x-axis represents a control estimate: A) Agaricomycotina paired with the basal lineages of Ascomycotina (Saccharo- and Taphrinomycotina) B) Pezizomycotina paired with the basal lineages of Basidiomycota (Ustilago- and Pucciniomycotina). C) basal lineages of Asco- and Basidiomycota normalized with the branch length in convergent lineages; D) basal lineages of Asco- and Basidiomycota normalized by the patristic distance between the crown nodes of convergent lineages. r denotes Pearson correlation coefficient.
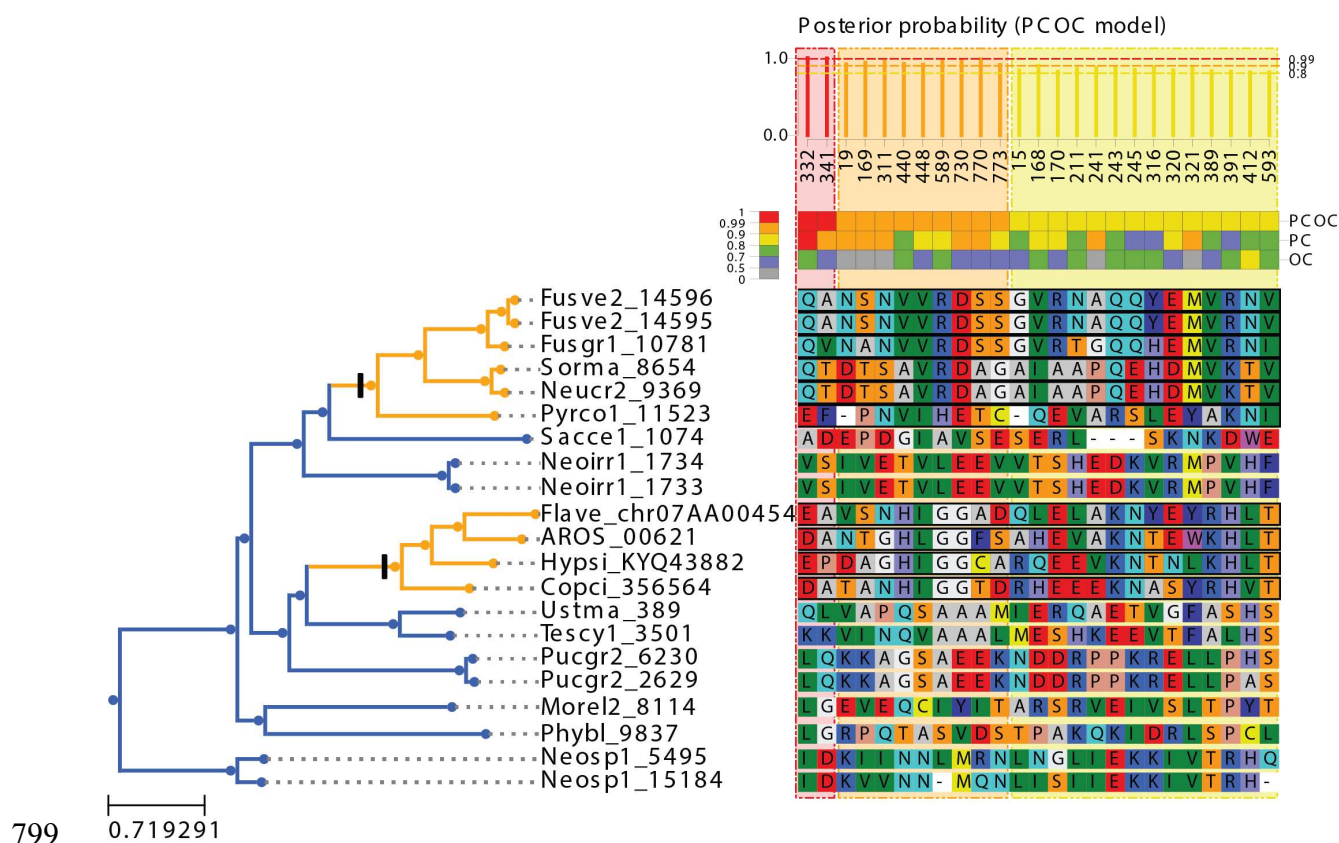
Fig. S9. An example of convergent amino acid shifts in the gene family containing *Aspergillus nimO/AN1779* homologs (OG0002402) identified using PCOC. Only sites where the posterior probability of being convergent was above 0.8 are shown. In the gene tree (left side) clades showing convergent phenotypes (fruiting body formation) are highlighted in orange.

806   **Table S1. (separate file).** List of 19 species used in comparative genomic analyses

807   **Table S2. (separate file).** Comparisons developmentally regulated genes to homologues with
808   known developmental roles in model organisms of fungal CM.

809   **Table S3. (separate file).** List of 1026 gene families with conserved developmental
810   expression

811   **Table S4. (separate file).** Statistical comparisons of the amount of parallel duplication in
812   developmental versus control gene families.

813   **Table S5. (separate file).** Gene families with significantly more amino acid shifts than
814   expected by chance.

815   **Fig S2. (separate file).** Gene Ontology analyses for 314 conserved developmental gene
816   families