Natural Product Reports

Biosynthetic gene clusters and the evolution of fungal chemodiversity

Journal:	*Natural Product Reports*
Manuscript ID	NP-HIG-07-2019-000045.R2
Article Type:	Highlight
Date Submitted by the Author:	09-Dec-2019
Complete List of Authors:	Rokas, Antonis; Vanderbilt University Mead, Matthew; Vanderbilt University Steenwyk, Jacob; Vanderbilt University Raja, Huzefa; University of North Carolina at Greensboro, Chemistry and Biochemistry Oberlies, Nicholas; University of North Carolina at Greensboro, Chemistry & Biochemistry

SCHOLARONE™
Manuscripts

1

Biosynthetic gene clusters and the evolution of fungal chemodiversity

3

4

5 Antonis Rokas[1,*], Matthew E. Mead[1], Jacob L. Steenwyk[1], Huzefa A. Raja[2], and Nicholas H.

6 Oberlies[2]

7

8 [1]Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA

9 [2]Department of Chemistry and Biochemistry, University of North Carolina at Greensboro,

10 Greensboro, NC, USA

11

12

13 * To whom correspondence should be addressed: antonis.rokas@vanderbilt.edu

14

15

16 **Keywords:** natural product, mycotoxin, antibiotic, secondary metabolic gene cluster, lateral

17 transfer, fungal genomics, fungal chemical diversity, ecological interactions, fungal

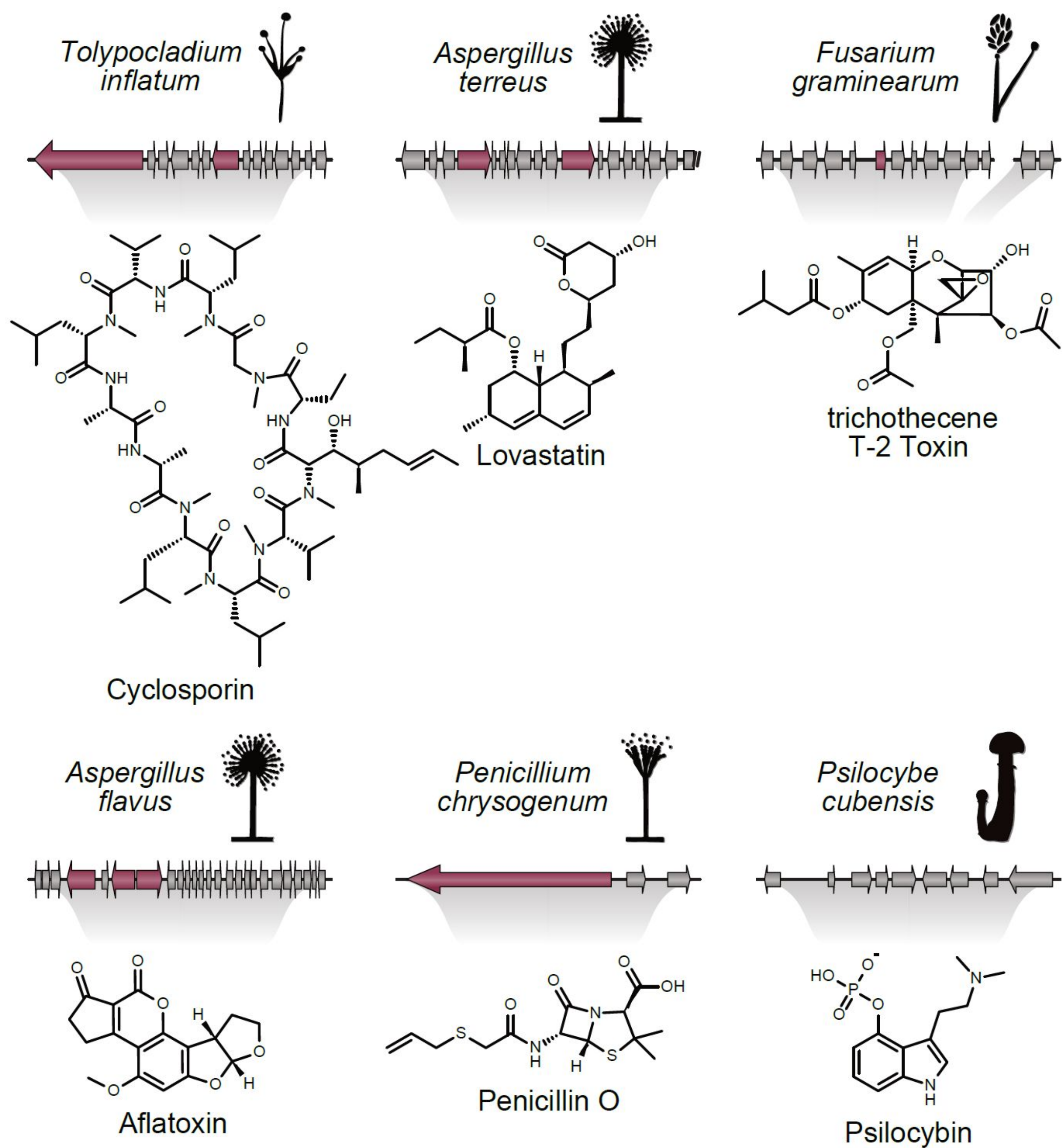
18 communication

19 **Abstract**

20 Fungi produce a remarkable diversity of secondary metabolites: small, bioactive molecules not
21 required for growth but which are essential to their ecological interactions with other organisms.
22 Genes that participate in the same secondary metabolic pathway typically reside next to each
23 other in fungal genomes and form biosynthetic gene clusters (BGCs). By synthesizing state-of-
24 the-art knowledge on the evolution of BGCs in fungi, we propose that fungal chemodiversity
25 stems from three molecular evolutionary processes involving BGCs: functional divergence,
26 horizontal transfer, and *de novo* assembly. We provide examples of how these processes have
27 contributed to the generation of fungal chemodiversity, discuss their relative importance, and
28 outline major, outstanding questions in the field.

29 **1. Fungal biosynthetic gene clusters produce diverse secondary metabolites of broad**
30 **ecological importance and human relevance**

31 Fungi produce a remarkable diversity of secondary metabolites,[1] also known as natural products,
32 such as the immunosuppressant cyclosporin,[2] the cholesterol reducing lovastatin,[3] the antibiotic
33 penicillin,[4] the hallucinogenic prodrug psilocybin,[5] and the mycotoxins trichothecene[6] and
34 aflatoxin[7] (Fig. 1). Although these small molecules are not required for fungal survival and
35 growth, their bioactive properties render them highly relevant to human affairs as drugs, toxins,
36 and pigments. But arguably their raison d'être is to act as crucial intermediaries at the front line
37 of fungal ecology. Numerous secondary metabolites are thought to play key roles in shaping the
38 interactions that fungi have with other organisms across the tree of life, including with other
39 fungi,[8] bacteria,[9, 10] plants,[11, 12] or animals.[13-15] These interactions are varied, and include
40 virulence, defense, quorum sensing, protection, nutrient acquisition and the promotion of growth
41 (Fig. 2).
42
43 --

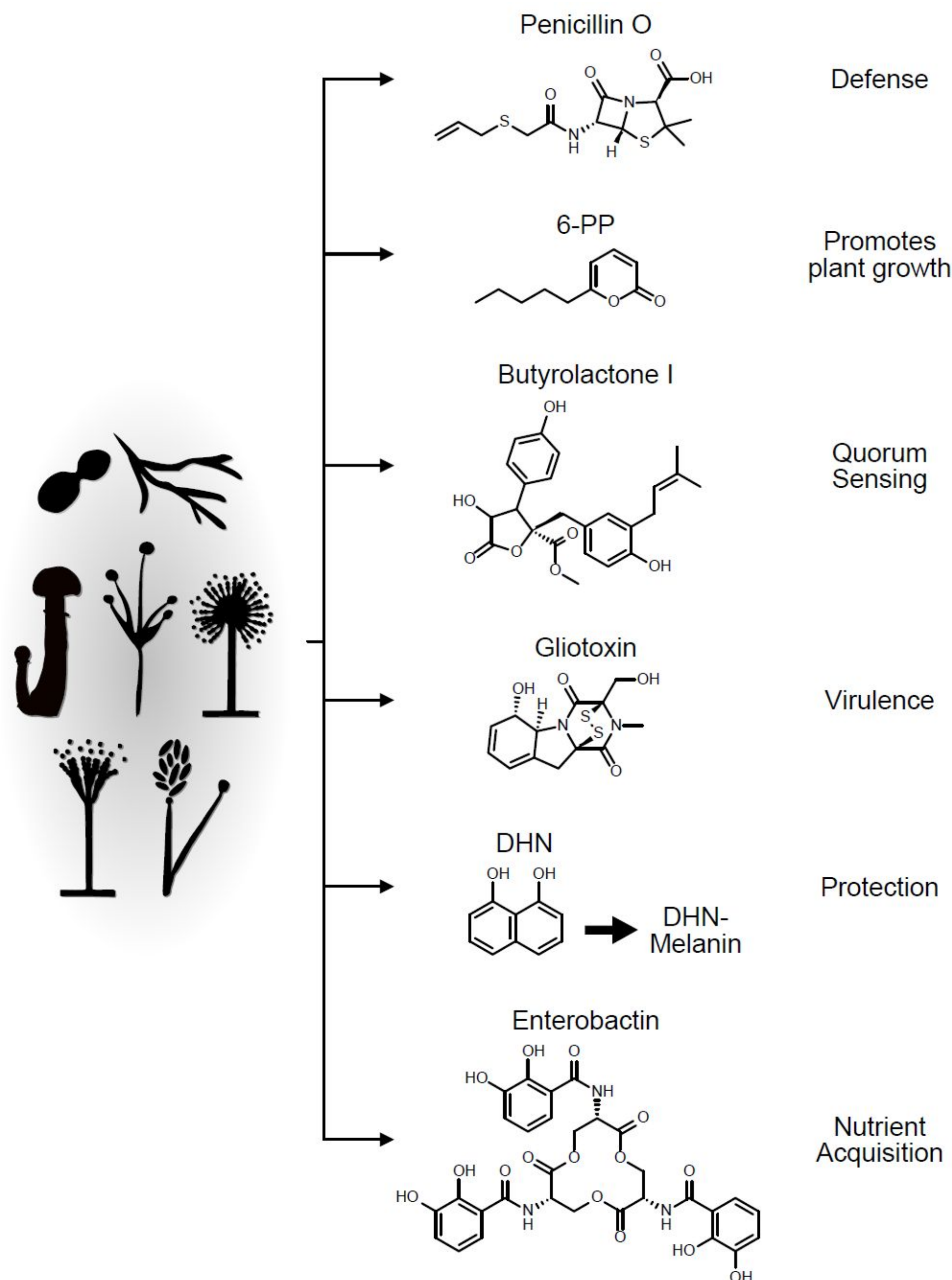


44

45 **Figure 1. Select examples of fungal BGCs, their secondary metabolites, and the organisms**
46 **that produce them.** Genes are represented by arrows; Genes colored maroon denote secondary
47 metabolite backbone biosynthesis genes (such as polyketide synthases, terpene synthases, and
48 non-ribosomal peptide synthases), whereas genes colored grey denote BGC genes with diverse
49 functions, such as metabolite modification, metabolite transport, regulation of BGC expression,

50 and resistance to secondary metabolite activity. Note that psilocybin biosynthesis does not
51 require any of the canonical backbone biosynthesis genes. Data from: Cyclosporin BGC[2],
52 lovastatin BGC[3], trichothecene T-2 toxin BGC[6], aflatoxin BGC[7], penicillin BGC[4], and
53 psilocybin BGC[5, 16].
54 --
55
56 Most fungal secondary metabolites are encoded by biosynthetic gene clusters (BGCs; Fig. 1);
57 each cluster typically contains the majority, if not all, of the genes participating in the production
58 of a given secondary metabolite, with these genes located adjacent to each other (i.e.,
59 "clustered") in the genome.[1, 17-19] A typical fungal BGC contains one or more genes whose
60 protein products catalyze the synthesis of the backbone of the metabolite (such as polyketide
61 synthases, non-ribosomal peptide synthases, and terpene synthases), and one or more genes
62 encoding for: i) enzymes (such as epimerases, methyltransferases, and hydroxylases) that modify
63 this backbone, ii) proteins involved in metabolite transport, iii) transcription factors involved in
64 regulation of BGC expression, and iv) proteins that confer resistance to the activity of the
65 secondary metabolite.[1, 17] Fungal BGCs are generally similar in their genomic organization to
66 bacterial BGCs; the key difference is that bacterial BGCs are typically organized into operons
67 (where multiple genes are transcribed into a single messenger RNA), whereas fungal BGCs are
68 typically transcribed individually.[20, 21]
69
70 Notable secondary metabolites produced by diverse backbone biosynthesis genes and BGCs
71 include: cyclosporin, a non-ribosomal peptide biosynthesized by a 14-gene BGC in the
72 ascomycete fungus *Tolypocladium inflatum*;[2] lovastatin, a polyketide biosynthesized by an 18-

73 gene BGC in the mold *Aspergillus terreus*;[3] the trichothecene T-2 toxin, a terpene biosynthesized
74 by a 12-gene BGC and a 2-gene BGC found in several *Fusarium* species;[6] aflatoxin, a polyketide
75 biosynthesized by a 25-gene BGC in the mold *Aspergillus flavus* and its close relatives;[7]
76 penicillin, a non-ribosomal peptide biosynthesized by a 3-gene BGC in molds in the genera
77 *Penicillium* and *Aspergillus*;[4] and psilocybin, a tryptamine-derived secondary metabolite
78 biosynthesized by a 9-gene BGC in several different basidiomycete genera whose biosynthesis
79 does not require any of the canonical backbone biosynthesis genes (Fig. 1).[5, 16] A comprehensive
80 and up to date compilation of fungal BGCs whose secondary metabolite products have been
81 functionally validated can be found at the MIBiG (Minimum Information about Biosynthetic
82 Gene cluster) repository.[22, 23]
83
84 BGCs vary widely in their numbers across fungal genomes; whereas ascomycete filamentous
85 fungi and basidiomycete fungi typically contain dozens (if not scores) of BGCs, unicellular
86 yeasts in both lineages either lack BGCs altogether or contain very few.[17, 24, 25] A given BGC is
87 often known from only a single species or a few closely related ones, but broadly and
88 discontinuously distributed BGCs, such as sterigmatocystin,[26] also exist. Additionally, BGCs and
89 their secondary metabolites also show extensive variation in their presence / absence patterns
90 within fungal species.[17, 27-29]
91
92 --



93

94 **Figure 2. Secondary metabolites are central to the ecology of many fungi and shape their**

95 **diverse interactions with other organisms.** Penicillin is an antibiotic whose ecological role lies

96 in fungal defense against bacteria,[10] 6-*n*-pentyl-6*H*-pyran-2-one (6-PP) promotes plant growth,[12]

97 butyrolactone I is a quorum sensing molecule,[30] gliotoxin is a virulence factor,[31] DHN-melanin

98 protects again UV light damage,[32] and enterobactin is an iron uptake molecule that contributes to
99 the acquisition of nutrients.[33]
100 --
101
102 A notable feature of BGCs, hinted at by their high variability and narrow taxonomic range, is
103 that they are rapidly evolving.[17, 34] Why is that so? From a molecular perspective, it has been
104 argued that the lower specificity of secondary metabolic enzymes means that new gene
105 duplicates are more likely to catalyze novel substrates and produce novel products that may be
106 favored by natural selection, accelerating their evolution.[35] Additionally, BGCs often reside in
107 fast-evolving genomic regions, such as near the ends of chromosomes[36] or in accessory
108 chromosomes.[37] From an ecological perspective, the involvement of secondary metabolites in
109 mediating interspecific interactions suggests that they are key in "arms races" between fungi and
110 their competitors, which are thought to accelerate evolutionary rates of the genes involved.[38] But
111 secondary metabolite biosynthesis is also energetically costly. Thus, loss of the ability to produce
112 a secondary metabolite and reliance on other fungal relatives in the community for its production
113 may be, at least up to a point, advantageous to individual organisms[39, 40] and further increase the
114 rate of BGC evolution.[17]
115
116 One important question raised by considering the ecological relevance of fungal secondary
117 metabolites, the narrowness of their taxonomic distribution, and the fast pace of BGC evolution,
118 concerns the molecular evolutionary processes that give rise to fungal chemodiversity. In this
119 highlight, we suggest that there are three major molecular evolutionary processes that occur at
120 the level of BGCs and which give rise to fungal chemodiversity: functional divergence,

121 horizontal or lateral transfer, and *de novo* assembly (see glossary in Table 1 for definitions of
122 these terms). While the focus of our highlight is on discussing how variation at the level of BGCs
123 gives rise to variation in secondary metabolism or chemodiversity, we note that all genetic
124 variation at the level of BGCs occurs via the standard battery of mutational types, such as point
125 mutations, insertions, deletions, rearrangements, duplications, and horizontal gene transfer (see
126 glossary in Table 1 for definitions). All of these types of mutations are well established and
127 known to influence fungal genes, genomes, and BGCs.[20, 27, 29, 34, 40]
128

129 **2. The evolutionary processes underlying fungal chemodiversity**

130 **2.1 BGC functional divergence**

131 Functional divergence is the process by which the accumulation of molecular differences
132 between evolutionarily related or homologous (see glossary in Table 1) genes and pathways
133 leads to a change in their function or phenotype. In the context of BGCs, functional divergence
134 refers to the accumulation of molecular differences between the gene sequences of homologous
135 BGCs that then give rise to chemical differences in their secondary metabolite products and
136 generate secondary metabolite structural diversity. Functional divergence has influenced both the
137 evolution of orthologous (see glossary in Table 1) BGCs that have originated through speciation
138 events as well as paralogous (see glossary in Table 1) BGCs that have originated through
139 duplication events.
140

141 **2.1.1 Functional divergence of orthologous BGCs**

142 Orthologous BGCs can functionally diverge via the accumulation of amino acid differences in
143 the enzymes encoded by BGCs. For example, the chemodiversity of fumonisin mycotoxins

144 among *Fusarium* fungi stems from amino acid sequence variation in a protein encoded by a
145 single gene from the fumonisin BGC (Fig. 3A).[41] Some *Fusarium* species, such as *Fusarium*
146 *verticillioides*, are known to produce primarily fumonisin B, whereas other species, such as
147 *Fusarium oxysporum*, produce primarily fumonisin C. The only difference in the structures of
148 fumonisin B and C is in the length of their backbones; the fumonisin B backbone is 20 carbon
149 atoms long, whereas the backbone of fumonisin C is 19 carbon atoms long. Comparison of the
150 fumonisin BGCs in *F. verticillioides* and *F. oxysporum* showed that the two species contain
151 orthologous BGCs with the same 19 (orthologous) genes; gene swapping experiments further
152 showed that sequence variation within the *fum8* gene, which encodes for an α-oxoamine
153 synthase, is responsible for the observed difference in the type of fumonisin (B or C) produced
154 by the two species.[41] The precise amino acid difference(s) between the *F. verticillioides* and *F.*
155 *oxysporum* Fum8 protein orthologs responsible for observed divergence in fumonisin structure
156 are not known and the two orthologs exhibit 91% similarity in their amino acid sequences.[41]
157 However, it appears that the *F. verticillioides* Fum8 enzyme preferentially binds the amino acid
158 alanine (and catalyzes its condensation to an 18-carbon linear polyketide to produce the 20-
159 carbon-long fumonisin B), whereas the *F. oxysporum* Fum8 preferentially binds glycine,
160 resulting in the production of the 19-carbon-long fumonisin C.[41] Sequence comparisons of
161 *Fusarium* α-oxoamine synthase sequences show that the amino acid residue at position 580 of
162 the protein is strongly associated with the type of fumonisin produced; presence of alanine at
163 position 580 is associated with fumonisin B production, whereas presence of valine at the same
164 position is associated with fumonisin C production.[42] Consistent with this association, mutations
165 of this residue in human α-oxoamine synthase have been shown to alter the enzyme's binding
166 affinity to its amino acid substrate.[43]

167

168 Alternatively, orthologous BGCs can functionally diverge through gains and losses of genes
169 (Fig. 3B). For example, some *Aspergillus* species, such as *Aspergillus flavus*, produce the
170 mycotoxin aflatoxin, whereas other species, including *Aspergillus nidulans*, produce the
171 mycotoxin sterigmatocystin. The two mycotoxins, as well as their BGCs, are similar to each
172 other. The difference in the mycotoxin produced is due to at least three genes (*aflP*, *aflU*, and
173 *aflQ*; shown in bold in Fig. 3B) present in the aflatoxin BGC that are not found in the
174 sterigmatocystin BGC. The AflP protein is an *O*-methyltransferase that converts sterigmatocystin
175 to *O*-methylsterigmatocystin, whereas the cytochrome P450 monooxygenase AflU and the P-450
176 monooxygenase AflQ catalyze the conversion of *O*-methylsterigmatocystin to aflatoxin G and
177 aflatoxin B, respectively.[7, 44] Note that the differences in gene content between the aflatoxin and
178 sterigmatocystin BGCs include additional genes (Fig. 3B); however, only *aflP*, *aflU*, and *aflQ*
179 have been shown to be involved in the conversion of sterigmatocystin to the aflatoxins.

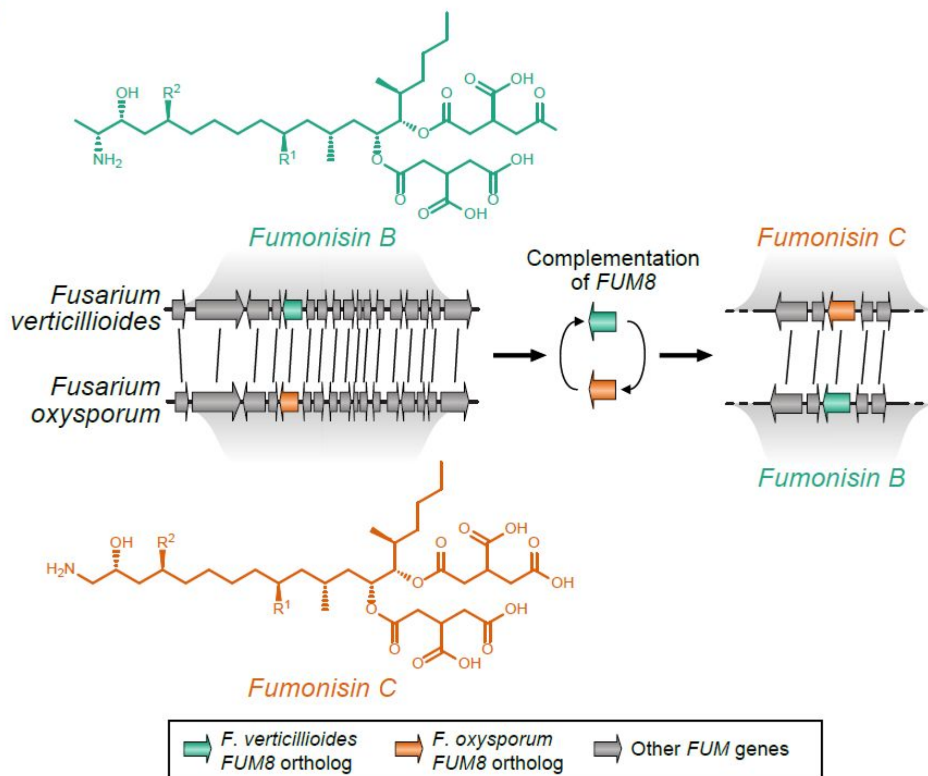
180

181 Finally, some orthologous BGCs have functionally diverged through both the accumulation of
182 amino acid differences in the protein products of their genes as well as through gains and losses
183 of genes. The combined effect of these two processes is thought to account for the observed
184 structural diversity of yanuthone antimicrobial compounds in *Penicillium* molds[45], as well as for
185 the diversity of the echinocandin class of antifungal drugs,[46] trichothecene mycotoxins,[47] and
186 ergot alkaloids[48] produced by diverse fungi.

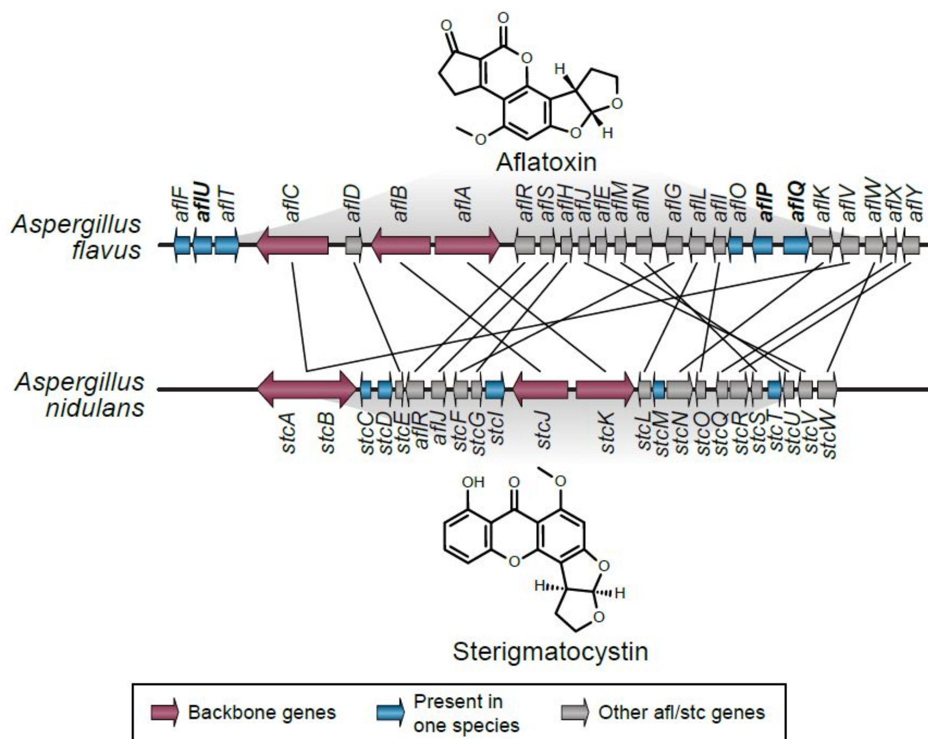
187

188 --

A



B



189

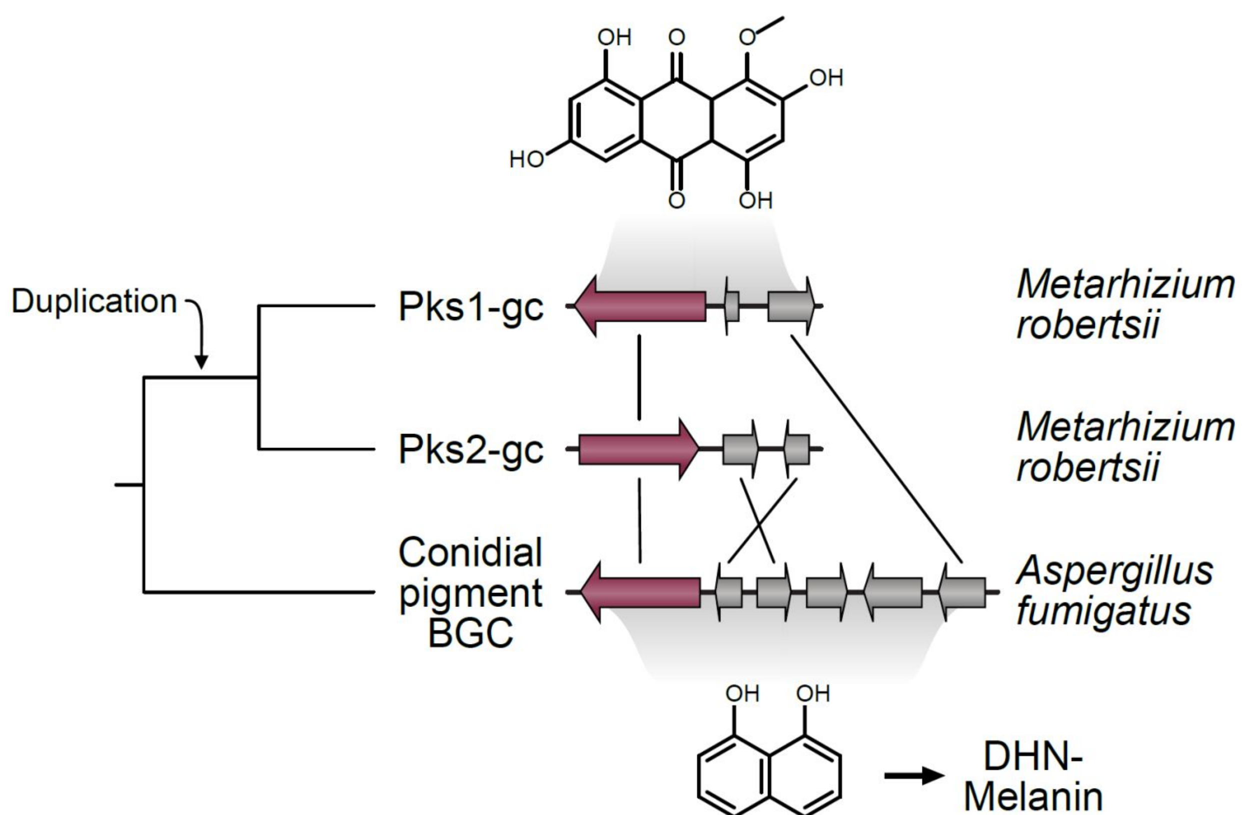
190 **Figure 3. Fumonisins and sterigmatocystin / aflatoxins; two notable examples of fungal**
191 **chemodiversity that stems from the functional divergence of orthologous BGCs.** Genes are
192 represented by arrows. Lines between genes from different species refer to orthologous genes.
193 --
194

195 ### 2.1.2 Functional divergence of paralogous BGCs

196 BGC functional divergence that gives rise to the evolution of new secondary metabolites can also
197 occur via the duplication of genomic regions containing entire BGCs. Even though duplication of
198 genes in BGCs has been widely documented and it is now well established that gene duplication
199 is a major driver of both the diversity of individual backbone genes present in BGCs[49, 50] as well
200 as of genes in BGCs in general,[34] much less is known about the duplication of entire BGCs.
201

202 An example of BGC duplication concerns the duplication of two polyketide-producing BGCs,
203 *Pks1-gc* and *Pks2-gc*, in *Metarhizium* entomopathogenic fungi, one of which is known to
204 produce an anthraquinone derivative.[51] Genomic and functional analyses of the two paralogous
205 BGCs show that they have functionally diverged through the reciprocal loss of genes in each
206 BGC as well as through the accumulation of substitutions in both the promoter and protein-
207 coding regions of their polyketide synthase genes (Fig. 4).[51] Interestingly, the only shared
208 paralogous gene pair between the *Pks1-gc* and the *Pks2-gc* is the *Pks1 – Pks2* pair. In contrast,
209 the *Pks1-gc* and the *Pks2-gc* BGCs share two and three homologous genes, respectively, with the
210 *A. fumigatus* conidial pigment BGC (Fig. 4). Consistent with these differences in gene sequence
211 and content, the two BGCs show distinct expression patterns (the genes of *Pks1-gc* are expressed
212 during asexual spore formation, whereas the genes of *Pks2-gc* are expressed during the

213 establishment of infection in insects) and produce distinct secondary metabolites. The
214 anthraquinone derivative product of *Pks1-gc* is involved in the pigmentation of asexual spores
215 and in abiotic stress tolerance, such as tolerance to UV light, whereas the uncharacterized
216 product of *Pks2-gc* appears to contribute to pathogenicity and not pigmentation or abiotic
217 stress.[51]
218
219 -- **Figure 4** ---

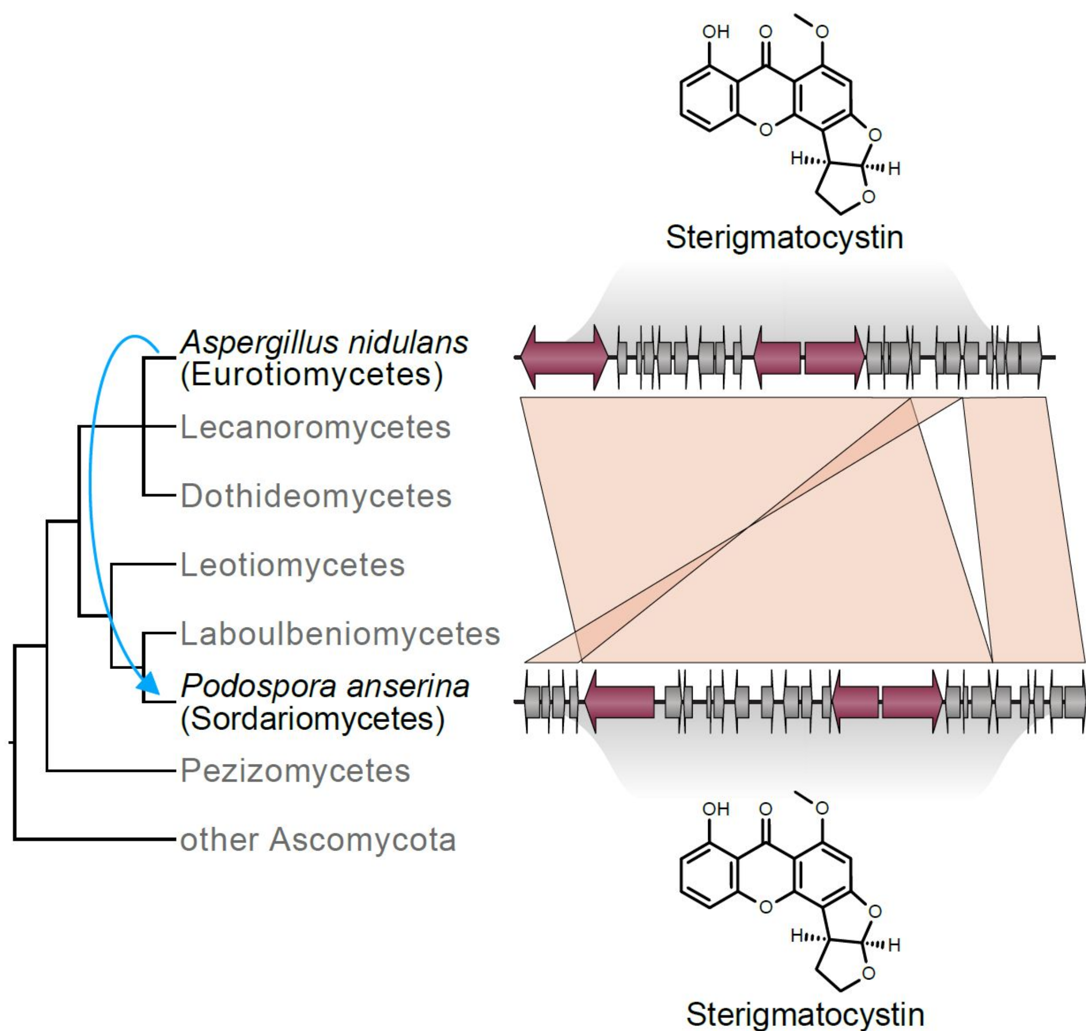


220
221 **Figure 4. Duplication and subsequent functional divergence of *Pks1-gc* and *Pks2*-gc, two**
222 **paralogous polyketide BGCs present in *Metarhizium* insect pathogens.** Two polyketide
223 synthase-containing BGCs, *Pks1-gc* and *Pks2-gc*, in *Metarhizium robertsii* appear to be the result
224 of an ancient duplication of an entire BGC that likely resembled the conidial pigment BGC
225 found in the mold *Aspergillus fumigatus*. The *Pks1-gc* BGC produces an anthraquinone

226 derivative, whereas the product of the *Pks2-gc* has yet to be characterized.[51] Data from: *M.*
227 *robertsii Pks1-gc* BGC[51] and chemical structure; *M. robertsii Pks2-gc* BGC;[51] and *A. fumigatus*
228 conidial pigment BGC.[52] Lines between genes from different species refer to orthologous genes.
229 --
230
231 Another example of functional divergence of paralogous BGCs are the patulin and yanuthone
232 BGCs in *Penicillium* molds, whose secondary metabolite products share a 6-methylsalicylic acid
233 (6-MSA) core.[45] The 15-gene patulin BGC and the 10-gene yanuthone BGCs contain several
234 pairs of paralogous genes thought to catalyze the same reactions leading to the formation of the
235 6-MSA core structure as well as several additional genes that lack sequence similarity to genes in
236 the other BGC and are presumably responsible for the structural differences between patulin and
237 yanuthones.[45] Thus, a proto-BGC responsible for the production of 6-MSA likely originated and
238 duplicated prior to the origin of *Penicillium*, followed by additional recruitment of non-
239 homologous genes in both BGCs. Interestingly, phylogenetic analysis of the 6-MSA synthase
240 protein suggests that the patulin 6-MSA synthase is more closely related to the 6-MSA synthases
241 found in the aculinic acid BGC from *Aspergillus aculeatus*[53] and in the terreic acid BGC in
242 *Aspergillus terreus*,[54] both of which produce 6-MSA-based secondary metabolites.[45] Thus, the
243 duplication and subsequent functional divergence of the patulin and yanuthone BGCs is part of a
244 broader series of duplication and functional divergence events of 6-MSA-based BGCs.
245
246 **2.3 BGC horizontal transfer**
247 Fungal chemodiversity can also originate via the horizontal transfer of entire BGCs from other
248 organisms.[55] For example, horizontal transfer of the sterigmatocystin BGC from *Aspergillus* to

249 *Podospora* resulted in the ability of the latter to produce sterigmatocystin (Fig. 5).[56, 57] In the
250 aftermath of horizontal transfer, the acquired BGCs can accumulate changes in their sequence
251 and genomic organization without altering the structure of the metabolic product. For example,
252 the average amino acid sequence similarity between the proteins encoded by the *Aspergillus*
253 *nidulans* and *Podospora anserina* sterigmatocystin BGCs is 63% and the two BGCs also differ
254 somewhat in their genomic organization, yet both produce the same metabolite. Thus, in contrast
255 to BGC functional divergence (section 2.2) and BGC *de novo* assembly (section 2.4 below), both
256 of which result in BGCs that produce new compounds, BGC horizontal transfer typically results
257 in the production of an existing compound in a new, typically distantly related, organism.
258
259 In the last decade, several examples of BGC horizontal transfer have been reported; most
260 transfers of entire BGCs are between fungi, such as the transfers of the BGC for the pigment
261 bikaverin from the ascomycete genus *Fusarium* to that of *Botrytis*,[58-60] of the BGC for the
262 hallucinogen psilocybin among basidiomycete fungi,[16] of the fumonisin BGC across *Fusarium*
263 species,[42] of the chaetoglobosin-like BGC from *Penicillium* to *Mycosphaerella populorum*,[61] or
264 the multiple transfers of the BGC for the histone deacetylase inhibitor depudecin among
265 ascomycete fungi.[62] In contrast, horizontal transfer of entire BGCs from bacteria, the lineage in
266 which secondary metabolism first originated,[63] appears to be less common and only one clear-cut
267 example of transfer of the siderophore enterobactin from enterobacteria to budding yeasts is
268 known to date.[33]
269
270 The examples discussed above all concern transfers of BGCs in the absence of functional
271 divergence (i.e., the same secondary metabolite is produced in both the donor and the recipient

272 organisms). The identification of examples of BGCs that functionally diverged after HGT is
273 more challenging because, following functional divergence, donor and recipient BGCs can
274 exhibit substantial divergence in gene content and arrangement.[55] BGC horizontal transfer
275 followed by functional divergence is thought to account for the diversification of
276 epipolythiodioxopiperazine (ETP) mycotoxins, such as gliotoxin, sirodesmin and their
277 relatives.[64]
278
279 -- **Figure 5** --



280

281 **Figure 5. Horizontal transfer of the sterigmatocystin BGC from *Aspergillus* to *Podospora***
282 **resulted in the presence of the sterigmatocystin BGC in the *Podospora* genome and its**
283 **ability to produce sterigmatocystin.** Evolutionary analyses of the history of the genes in the
284 sterigmatocystin BGC suggest that the *Podospora* BGC was horizontally acquired from an
285 *Aspergillus* ancestor.[56] Subsequent functional and chemical studies have validated that
286 *Podospora* fungi produce the sterigmatocystin mycotoxin.[10, 57, 65] Large orthologous blocks of
287 genetic sequence are depicted using orange trapezoids.
288 --
289
290 **2.4 *De novo* BGC assembly**
291 The final, and least well-documented, evolutionary process involved in the generation of fungal
292 chemodiversity is *de novo* BGC assembly, under which new secondary metabolites originate
293 from scratch in fungal genomes. The genes that become part of the newly formed secondary
294 metabolic pathway originate either through duplication and relocation of native genes or through
295 horizontal acquisition. One important source of genes for BGCs are duplicates of genes encoding
296 for enzymes already involved in primary and secondary metabolism, such as the isopropyl-
297 malate synthase gene in the echinocandin BGC of *Aspergillus rugulosus*,[66] and the citrate
298 synthase gene in the zaragozic acid BGC of *Curvularia lunata*.[67]
299
300 *De novo* assembled BGCs are unlikely to be highly similar in their gene or sequence content to
301 already existing BGCs, making their identification through comparisons of genome sequences
302 (the major way all cases of BGC functional divergence and BGC horizontal transfer have been
303 identified) much more challenging. Several lines of evidence support that this mechanism also

304 gives rise to fungal BGCs. The same general evolutionary process of *de novo* pathway assembly
305 is thought to be responsible for the origin of novel pathways that break down anthropogenic
306 chemicals[68] as well as of certain catabolic pathways.[11, 69]
307
308 *De novo* secondary metabolic pathways may originate in a similar manner via a two-step
309 process; step one involves the assembly of the secondary metabolic pathway through the
310 recruitment of native genes, duplicates of native genes, and horizontally acquired genes, and step
311 two involves their clustering into a BGC. Consistent with this model, several fungal secondary
312 metabolic pathways are comprised of two or more BGCs,[17] suggesting that the clustering of
313 fungal secondary metabolic pathways is not an absolute requirement for their function. For
314 example, a 12-gene and a 2-gene BGC found in distinct genomic locations are involved in the
315 biosynthesis of the trichothecene T-2 toxin in *F. graminearum* (Fig. 1).[6] Additionally, several
316 BGCs contain distinct smaller clusters of genes (often referred to as modules) responsible for the
317 production of functional intermediates within the pathway, suggesting that the entire BGC
318 evolved via the merging of distinct, pre-existing smaller BGCs. For example, BGCs associated
319 with the production of echinocandins typically contain a 4-gene cluster for the production of L-
320 homotyrosine, one of the intermediates required for echinocandin biosynthesis.[46] Similarly, the
321 genes of BGCs responsible for the production of distinct secondary metabolites can be
322 intertwined in the genome, as in the case of the fumagillin and pseurotin BGCs in *A. fumigatus*,[70]
323 providing empirical evidence of the evolutionary merging of distinct BGCs.
324
325 The second line of evidence is that gene relocation has been implicated in the diversification of
326 BGCs, such as the expansion of a trichothecene BGC in *Fusarium* species.[71] Perhaps the best

327 candidate of *de novo* assembly of a BGC involved in secondary metabolism is the fumonisin
328 BGC found in certain *Fusarium* and *Aspergillus* species.[72] While the presence of the BGC in
329 *Aspergillus* is best explained by horizontal transfer from another fungus, one hypothesis for the
330 origin of the *Fusarium* BGC, based on examination of phylogenies of genes in the BGC, is that it
331 arose through the relocation and clustering of genes that were originally dispersed in the
332 genome.[72]
333

334 **3. Perspective and Major unanswered questions**

335 Even though the remarkable breadth of fungal chemodiversity was well appreciated before the
336 advent of the genomics revolution,[73] the sequencing of diverse fungal genomes from 2003
337 onward quickly began revealing that fungal genomes contained even larger numbers of BGCs
338 responsible for the biosynthesis of yet-unknown secondary metabolite products and provided
339 unprecedented opportunities for studying the origins and evolution of fungal chemodiversity at
340 the DNA sequence level.[17] Currently, the molecular evolutionary processes by which fungal
341 BGCs evolve are becoming established (Fig. 3) and the relationship between chemical diversity
342 and BGC diversity for several secondary metabolites is being increasingly refined.[47]
343

344 Comparison of the genetic and evolutionary mechanisms underpinning the evolution of fungal
345 chemodiversity with those inferred from the study of bacterial chemodiversity[74, 75] suggests that
346 similar mechanisms operate in both lineages. Arguably the biggest difference is the extent of the
347 contribution of BGC horizontal transfer in driving chemodiversity in the two lineages. Although
348 the role of BGC horizontal transfer is increasingly appreciated in fungi (see section 2.3),

349 bacterial BGC horizontal transfer occurs at far higher rates and plays a bigger role in shaping
350 bacterial chemodiversity.[74]
351
352 While the major contours of the molecular evolutionary basis of fungal chemodiversity are
353 increasingly well understood, several major outstanding questions and opportunities remain. For
354 example, we still lack an understanding of why fungal secondary metabolic pathways are
355 typically arranged in the genome as BGCs (three genetic models, namely co-regulation, genetic
356 linkage, and selfishness, and one phenotypic model, namely toxicity avoidance, have been put
357 forward as explanations)[17, 19, 76] and whether this clustering is associated with fungal
358 chemodiversity. We similarly lack a complete knowledge of the distribution and genomic
359 arrangement of secondary metabolic pathways in fungal genomes, especially from less-studied
360 and less-sequenced lineages located outside a few select genera of filamentous fungi (e.g.,
361 *Aspergillus*, *Fusarium*, *Penicillium*) from the phylum Ascomycota.[77]
362
363 In the context of this highlight article, arguably the biggest challenges and opportunities lie in
364 uncovering examples of *de novo* BGC assembly, understanding the relative contribution of the
365 three different processes in sculpting BGC diversity, and elucidating how this diversity translates
366 to chemodiversity. Recently developed computational algorithms now allow the construction of
367 networks of fungal BGCs on the basis of their sequence similarity and gene order, enabling the
368 grouping of BGCs into BGC families, of families into clans, and so on.[78] Reconciling this
369 network view of BGC evolution with the evolutionary processes that we discuss promises to
370 illuminate their relative importance in sculpting BGC diversity and how that translated to
371 chemodiversity. For example, a recent examination of 37 *Aspergillus* and *Penicillium* genomes

372 identified more than 2,700 BGCs that could be grouped into 455 BGC families that presumably
373 produce distinct groups of secondary metabolites; strikingly, nearly half of these families
374 contained only a single BGC.[79] How did these single-BGC families originate and how common
375 are they when the entirety of fungal genomes is examined? And how do these 455 BGC families
376 relate to the ~15,600 described fungal secondary metabolites?[10] These are exciting questions but
377 also non-trivial to address, not least because of the challenges associated with handling and
378 analyzing the ever increasing volume of publicly available fungal genomes (there are 5,064 draft
379 fungal genomes in GenBank as of October 30, 2019).
380
381 But the opportunity does not stop here; by considering the mechanisms that give rise to BGC
382 diversity we begin to set the foundations of an evolutionary framework to bridge genotype
383 (BGCs) with chemotype (their secondary metabolites). Establishing such a framework will not
384 only advance our understanding of how genomic diversity translates to chemodiversity, but will
385 also be useful in genetic engineering- and directed evolution-based efforts to discover and
386 produce new leads in the pharmaceutical and agrochemical research areas.[45] Connecting BGC
387 diversity with chemodiversity, and elucidating the relationship between BGC sequence
388 divergence and chemical structure divergence, is even more daunting due to the current lack of
389 structures for most fungal BGCs[17] and vice versa (i.e., the BGCs responsible for making most
390 fungal secondary metabolites are unknown). With the sequences of tens of thousands of fungal
391 BGCs, thousands of fungal secondary metabolite chemical structures, and a smorgasbord of
392 novel synthetic biology, chemical, and bioinformatic tools that accelerate the discovery of new
393 secondary metabolites[80] at hand, exciting discoveries lay ahead.
394

395 ## 4. Conflicts of interest

396 There are no conflicts of interest to declare.

397

398 ## 5. Acknowledgements

399 We are grateful to Marnix Medema for the invitation to contribute this article, as well as to two
400 anonymous reviewers and an editorial board member for their constructive feedback. We thank
401 Sonja Knowles from UNCG for assistance with the structural drawings. This work was
402 supported by a Vanderbilt University Discovery Grant, by the Howard Hughes Medical Institute
403 through the James H. Gilliam Fellowships for Advanced Study program (J.L.S. and A.R.), and a
404 Regular Faculty Grant (from the University of North Carolina at Greensboro). Research in
405 A.R.'s lab is also supported by the National Science Foundation (DEB1442113), the
406 Guggenheim Foundation, and the Burroughs Wellcome Fund.. Research on bioactive fungal
407 metabolites in N.H.O.'s lab is also supported by the National Cancer Institute (P01 CA125066).

408

409 ## 6. References

410 1. N. P. Keller, *Nat. Rev. Microbiol.*, 2019, **17**, 167-180.
411 2. K. E. Bushley, R. Raja, P. Jaiswal, J. S. Cumbie, M. Nonogaki, A. E. Boyd, C. A.
412 Owensby, B. J. Knaus, J. Elser, D. Miller, Y. Di, K. L. McPhail and J. W. Spatafora,
413 *PLoS Genet.*, 2013, **9**, e1003496.
414 3. K. C. Mulder, F. Mulinari, O. L. Franco, M. S. Soares, B. S. Magalhaes and N. S.
415 Parachin, *Biotechnol Adv*, 2015, **33**, 648-665.
416 4. S. Gutierrez, F. Fierro, J. Casqueiro and J. F. Martin, *Antonie Van Leeuwenhoek*, 1999,
417 **75**, 81-94.
418 5. J. Fricke, F. Blei and D. Hoffmeister, *Angew Chem Int Ed Engl*, 2017, **56**, 12352-12355.
419 6. N. J. Alexander, R. H. Proctor and S. P. McCormick, *Toxin Rev*, 2009, **28**, 198-215.
420 7. J. Yu, P. K. Chang, K. C. Ehrlich, J. W. Cary, D. Bhatnagar, T. E. Cleveland, G. A.
421 Payne, J. E. Linz, C. P. Woloshuk and J. W. Bennett, *Appl. Environ. Microbiol.*, 2004,
422 **70**, 1253-1262.
423 8. S. L. Knowles, H. A. Raja, A. J. Wright, A. M. L. Lee, L. K. Caesar, N. B. Cech, M. E.
424 Mead, J. L. Steenwyk, L. N. A. Ries, G. H. Goldman, A. Rokas and N. H. Oberlies, *Front*
425 *Microbiol*, 2019, **10**, 285.
426 9. E. B. Van Arnam, C. R. Currie and J. Clardy, *Chem Soc Rev*, 2018, **47**, 1638-1651.

427 10. G. F. Bills and J. B. Gloer, *Microbiol Spectr*, 2016, **4**.
428 11. E. Gluck-Thaler, V. Vijayakumar and J. C. Slot, *Mol. Ecol.*, 2018, **27**, 5120-5136.
429 12. F. Vinale, K. Sivasithamparam, E. L. Ghisalberti, R. Marra, M. J. Barbetti, H. Li, S. L.
430 Woo and M. Lorito, *Physiol Mol Plant P*, 2008, **72**, 80-86.
431 13. M. T. Drott, T. Debenport, S. A. Higgins, D. H. Buckley and M. G. Milgroom, *mBio*,
432 2019, **10**.
433 14. S. Caballero Ortiz, M. Trienens, K. Pfohl, P. Karlovsky, G. Holighaus and M. Rohlfs,
434 *Ecol Evol*, 2018, **8**, 4328-4339.
435 15. M. Rohlfs, *Front Microbiol*, 2014, **5**, 788.
436 16. H. T. Reynolds, V. Vijayakumar, E. Gluck-Thaler, H. B. Korotkin, P. B. Matheny and J.
437 C. Slot, *Evolution Letters*, 2018, DOI: 10.1002/evl3.42, in press.
438 17. A. Rokas, J. H. Wisecaver and A. L. Lind, *Nat. Rev. Microbiol.*, 2018, **16**, 731-744.
439 18. J. C. Slot, *Adv Genet*, 2017, **100**, 141-178.
440 19. H. W. Nutzmann, C. Scazzocchio and A. Osbourn, *Ann. Rev. Genet.*, 2018, **52**, 159-183.
441 20. A. Osbourn, *Trends Genet.*, 2010, **26**, 449-457.
442 21. Q. Yue, L. Chen, Y. Li, G. F. Bills, X. Zhang, M. Xiang, S. Li, Y. Che, C. Wang, X. Niu,
443 Z. An and X. Liu, *MBio*, 2015, **6**.
444 22. Y. F. Li, K. J. S. Tsai, C. J. B. Harvey, J. J. Li, B. E. Ary, E. E. Berlew, B. L. Boehman,
445 D. M. Findley, A. G. Friant, C. A. Gardner, M. P. Gould, J. H. Ha, B. K. Lilley, E. L.
446 McKinstry, S. Nawal, R. C. Parry, K. W. Rothchild, S. D. Silbert, M. D. Tentilucci, A.
447 M. Thurston, R. B. Wai, Y. Yoon, R. S. Aiyar, M. H. Medema, M. E. Hillenmeyer and L.
448 K. Charkoudian, *Fungal Genet. Biol.*, 2016, **89**, 18-28.
449 23. M. H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J. B. Biggins, K. Blin, I. de
450 Bruijn, Y. H. Chooi, J. Claesen, R. C. Coates, P. Cruz-Morales, S. Duddela, S. Dusterhus,
451 D. J. Edwards, D. P. Fewer, N. Garg, C. Geiger, J. P. Gomez-Escribano, A. Greule, M.
452 Hadjithomas, A. S. Haines, E. J. N. Helfrich, M. L. Hillwig, K. Ishida, A. C. Jones, C. S.
453 Jones, K. Jungmann, C. Kegler, H. U. Kim, P. Kotter, D. Krug, J. Masschelein, A. V.
454 Melnik, S. M. Mantovani, E. A. Monroe, M. Moore, N. Moss, H. W. Nutzmann, G. H.
455 Pan, A. Pati, D. Petras, F. J. Reen, F. Rosconi, Z. Rui, Z. H. Tian, N. J. Tobias, Y.
456 Tsunematsu, P. Wiemann, E. Wyckoff, X. H. Yan, G. Yim, F. G. Yu, Y. C. Xie, B.
457 Aigle, A. K. Apel, C. J. Balibar, E. P. Balskus, F. Barona-Gomez, A. Bechthold, H. B.
458 Bode, R. Borriss, S. F. Brady, A. A. Brakhage, P. Caffrey, Y. Q. Cheng, J. Clardy, R. J.
459 Cox, R. De Mot, S. Donadio, M. S. Donia, W. A. van der Donk, P. C. Dorrestein, S.
460 Doyle, A. J. M. Driessen, M. Ehling-Schulz, K. D. Entian, M. A. Fischbach, L. Gerwick,
461 W. H. Gerwick, H. Gross, B. Gust, C. Hertweck, M. Hofte, S. E. Jensen, J. H. Ju, L.
462 Katz, L. Kaysser, J. L. Klassen, N. P. Keller, J. Kormanec, O. P. Kuipers, T. Kuzuyama,
463 N. C. Kyrpides, H. J. Kwon, S. Lautru, R. Lavigne, C. Y. Lee, B. Linquan, X. Y. Liu, W.
464 Liu, A. Luzhetskyy, T. Mahmud, Y. Mast, C. Mendez, M. Metsa-Ketela, J. Micklefield,
465 D. A. Mitchell, B. S. Moore, L. M. Moreira, R. Muller, B. A. Neilan, M. Nett, J. Nielsen,
466 F. O'Gara, H. Oikawa, A. Osbourn, M. S. Osburne, B. Ostash, S. M. Payne, J. L.
467 Pernodet, M. Petricek, J. Piel, O. Ploux, J. M. Raaijmakers, J. A. Salas, E. K. Schmitt, B.
468 Scott, R. F. Seipke, B. Shen, D. H. Sherman, K. Sivonen, M. J. Smanski, M. Sosio, E.
469 Stegmann, R. D. Sussmuth, K. Tahlan, C. M. Thomas, Y. Tang, A. W. Truman, M.
470 Viaud, J. D. Walton, C. T. Walsh, T. Weber, G. P. van Wezel, B. Wilkinson, J. M.
471 Willey, W. Wohlleben, G. D. Wright, N. Ziemert, C. S. Zhang, S. B. Zotchev, R.
472 Breitling, E. Takano and F. O. Glockner, *Nature Chemical Biology*, 2015, **11**, 625-631.

473 24. D. J. Krause, J. Kominek, D. A. Opulente, X. X. Shen, X. Zhou, Q. K. Langdon, J.
474 DeVirgilio, A. B. Hulfachor, C. P. Kurtzman, A. Rokas and C. T. Hittinger, *Proc Natl*
475 *Acad Sci U S A*, 2018, **115**, 11030-11035.
476 25. S. E. Helaly, B. Thongbai and M. Stadler, *Nat. Prod. Rep.*, 2018, **35**, 992-1014.
477 26. C. Rank, K. F. Nielsen, T. O. Larsen, J. Varga, R. A. Samson and J. C. Frisvad, *Fungal*
478 *Biol.*, 2011, **115**, 406-420.
479 27. A. L. Lind, J. H. Wisecaver, C. Lameiras, P. Wiemann, J. M. Palmer, N. P. Keller, F.
480 Rodrigues, G. H. Goldman and A. Rokas, *PLoS Biol.*, 2017, **15**, e2003583.
481 28. F. E. Hartmann and D. Croll, *Mol. Biol. Evol.*, 2017, **34**, 2808-2822.
482 29. R. A. Olarte, J. Menke, Y. Zhang, S. Sullivan, J. C. Slot, Y. Huang, J. P. Badalamenti, A.
483 C. Quandt, J. W. Spatafora and K. E. Bushley, *BMC Genomics*, 2019, **20**, 120.
484 30. T. G. Schimmel, A. D. Coffman and S. J. Parsons, *Appl. Environ. Microbiol.*, 1998, **64**,
485 3707-3712.
486 31. J. A. Sugui, J. Pardo, Y. C. Chang, K. A. Zarember, G. Nardone, E. M. Galvez, A.
487 Mullbacher, J. I. Gallin, M. M. Simon and K. J. Kwon-Chung, *Eukaryot. Cell*, 2007, **6**,
488 1562-1569.
489 32. C. Kawamura, T. Tsujimoto and T. Tsuge, *Mol Plant Microbe Interact*, 1999, **12**, 59-63.
490 33. J. Kominek, D. T. Doering, D. A. Opulente, X. X. Shen, X. Zhou, J. DeVirgilio, A. B.
491 Hulfachor, M. Groenewald, M. A. McGee, S. D. Karlen, C. P. Kurtzman, A. Rokas and
492 C. T. Hittinger, *Cell*, 2019, **176**, 1356-1366 e1310.
493 34. J. H. Wisecaver, J. C. Slot and A. Rokas, *PLoS Genet.*, 2014, **10**, e1004816.
494 35. J. K. Weng, R. N. Philippe and J. P. Noel, *Science*, 2012, **336**, 1667-1670.
495 36. J. M. Palmer and N. P. Keller, *Curr. Opin. Microbiol.*, 2010, **13**, 431-436.
496 37. K. Hoogendoorn, L. Barra, C. Waalwijk, J. S. Dickschat, T. A. J. van der Lee and M. H.
497 Medema, *Front Microbiol*, 2018, **9**, 1158.
498 38. S. Paterson, T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M.
499 Quail, F. Smith, D. Walker, B. Libberton, A. Fenton, N. Hall and M. A. Brockhurst,
500 *Nature*, 2010, **464**, 275-278.
501 39. J. J. Morris, R. E. Lenski and E. R. Zinser, *MBio*, 2012, **3**.
502 40. T. A. Richards and N. J. Talbot, *Nat. Rev. Microbiol.*, 2013, **11**, 720-727.
503 41. R. H. Proctor, M. Busman, J. A. Seo, Y. W. Lee and R. D. Plattner, *Fungal Genet. Biol.*,
504 2008, **45**, 1016-1026.
505 42. R. H. Proctor, F. Van Hove, A. Susca, G. Stea, M. Busman, T. van der Lee, C. Waalwijk,
506 A. Moretti and T. J. Ward, *Mol. Microbiol.*, 2013, **90**, 290-306.
507 43. P. M. Shoolingin-Jordan, S. Al-Daihan, D. Alexeev, R. L. Baxter, S. S. Bottomley, I. D.
508 Kahari, I. Roy, M. Sarwar, L. Sawyer and S. F. Wang, *Biochim Biophys Acta*, 2003,
509 **1647**, 361-366.
510 44. J. Yu, *Toxins (Basel)*, 2012, **4**, 1024-1057.
511 45. J. C. Nielsen, S. Grijseels, S. Prigent, B. Ji, J. Dainat, K. F. Nielsen, J. C. Frisvad, M.
512 Workman and J. Nielsen, *Nat Microbiol*, 2017, **2**, 17044.
513 46. Q. Yue, L. Chen, X. Zhang, K. Li, J. Sun, X. Liu, Z. An and G. F. Bills, *Eukaryot. Cell*,
514 2015, **14**, 698-718.
515 47. R. H. Proctor, S. P. McCormick, H. S. Kim, R. E. Cardoza, A. M. Stanley, L. Lindo, A.
516 Kelly, D. W. Brown, T. Lee, M. M. Vaughan, N. J. Alexander, M. Busman and S.
517 Gutierrez, *PLoS Pathog.*, 2018, **14**, e1006946.

518 48. C. A. Young, C. L. Schardl, D. G. Panaccione, S. Florea, J. E. Takach, N. D. Charlton, N.
519 Moore, J. S. Webb and J. Jaromczyk, *Toxins (Basel)*, 2015, **7**, 1273-1302.
520 49. B. J. Condon, Y. Leng, D. Wu, K. E. Bushley, R. A. Ohm, R. Otillar, J. Martin, W.
521 Schackwitz, J. Grimwood, N. MohdZainudin, C. Xue, R. Wang, V. A. Manning, B.
522 Dhillon, Z. J. Tu, B. J. Steffenson, A. Salamov, H. Sun, S. Lowry, K. LaButti, J. Han, A.
523 Copeland, E. Lindquist, K. Barry, J. Schmutz, S. E. Baker, L. M. Ciuffetti, I. V.
524 Grigoriev, S. Zhong and B. G. Turgeon, *PLoS Genet.*, 2013, **9**, e1003233.
525 50. G. Koczyk, A. Dawidziuk and D. Popiel, *Genome Biol Evol*, 2015, **7**, 3132-3154.
526 51. G. Zeng, P. Zhang, Q. Zhang, H. Zhao, Z. Li, X. Zhang, C. Wang, W. B. Yin and W.
527 Fang, *PLoS Genet.*, 2018, **14**, e1007472.
528 52. H. F. Tsai, M. H. Wheeler, Y. C. Chang and K. J. Kwon-Chung, *Journal of Bacteriology*,
529 1999, **181**, 6469-6477.
530 53. L. M. Petersen, D. K. Holm, C. H. Gotfredsen, U. H. Mortensen and T. O. Larsen,
531 *Chembiochem*, 2015, **16**, 2200-2204.
532 54. C. J. Guo, W. W. Sun, K. S. Bruno and C. C. Wang, *Org. Lett.*, 2014, **16**, 5250-5253.
533 55. J. H. Wisecaver and A. Rokas, *Front Microbiol*, 2015, **6**, 161.
534 56. J. C. Slot and A. Rokas, *Curr. Biol.*, 2011, **21**, 134-139.
535 57. L. Shen, F. H. Poree, T. Gaslonde, H. Lalucque, F. Chapeland-Leclerc and G. Ruprich-
536 Robert, *Environmental microbiology*, 2019, DOI: 10.1111/1462-2920.14698.
537 58. M. A. Campbell, M. Staats, J. L. A. van Kan, A. Rokas and J. C. Slot, *Mycologia*, 2013,
538 **105**, 1126-1134.
539 59. M. A. Campbell, A. Rokas and J. C. Slot, *Genome Biol Evol*, 2012, **4**, 289-293.
540 60. J. E. Spraker, P. Wiemann, J. A. Baccile, N. Venkatesh, J. Schumacher, F. C. Schroeder,
541 L. M. Sanchez and N. P. Keller, *MBio*, 2018, **9**.
542 61. B. Dhillon, N. Feau, A. L. Aerts, S. Beauseigle, L. Bernier, A. Copeland, A. Foster, N.
543 Gill, B. Henrissat, P. Herath, K. M. LaButti, A. Levasseur, E. A. Lindquist, E. Majoor, R.
544 A. Ohm, J. L. Pangilinan, A. Pribowo, J. N. Saddler, M. L. Sakalidis, R. P. de Vries, I. V.
545 Grigoriev, S. B. Goodwin, P. Tanguay and R. C. Hamelin, *Proc Natl Acad Sci U S A*,
546 2015, **112**, 3451-3456.
547 62. H. T. Reynolds, J. C. Slot, H. H. Divon, E. Lysoe, R. H. Proctor and D. W. Brown, *Mol.*
548 *Biol. Evol.*, 2017, **34**, 2002-2015.
549 63. T. Cavalier-Smith, *Ciba Found Symp*, 1992, **171**, 64-80; discussion 80-67.
550 64. N. J. Patron, R. F. Waller, A. J. Cozijnsen, D. C. Straney, D. M. Gardiner, W. C. Nierman
551 and B. J. Howlett, *BMC Evol. Biol.*, 2007, **7**, 174.
552 65. J. C. Matasyoh, B. Dittrich, A. Schueffler and H. Laatsch, *Parasitol. Res.*, 2011, **108**,
553 561-566.
554 66. R. A. Cacho, W. Jiang, Y. H. Chooi, C. T. Walsh and Y. Tang, *J Am Chem Soc*, 2012,
555 **134**, 16781-16790.
556 67. N. Liu, Y. S. Hung, S. S. Gao, L. Hang, Y. Zou, Y. H. Chooi and Y. Tang, *Org. Lett.*,
557 2017, **19**, 3560-3563.
558 68. S. D. Copley, *Nat. Chem. Biol.*, 2009, **5**, 559-566.
559 69. S. Wong and K. H. Wolfe, *Nat. Genet.*, 2005, **37**, 777-782.
560 70. P. Wiemann, C. J. Guo, J. M. Palmer, R. Sekonyela, C. C. Wang and N. P. Keller, *Proc*
561 *Natl Acad Sci U S A*, 2013, **110**, 17065-17070.
562 71. R. H. Proctor, S. P. McCormick, N. J. Alexander and A. E. Desjardins, *Mol. Microbiol.*,
563 2009, **74**, 1128-1142.

564 72. N. Khaldi and K. H. Wolfe, *Int. J. Evol. Biol.*, 2011, **2011**, 423821.
565 73. C. Pearce, *Adv. Appl. Microbiol.*, 1997, **44**, 1-80.
566 74. A. C. Ruzzini and J. Clardy, *Curr. Biol.*, 2016, **26**, R859-R864.
567 75. M. A. Fischbach, C. T. Walsh and J. Clardy, *Proc Natl Acad Sci U S A*, 2008, **105**, 4601-
568 4608.
569 76. K. L. McGary, J. C. Slot and A. Rokas, *Proc. Natl. Acad. Sci. USA*, 2013, **110**, 11481-
570 11486.
571 77. M. Stadler and D. Hoffmeister, *Front Microbiol*, 2015, **6**, 127.
572 78. J. C. Navarro-Muñoz, N. Selem-Mojica, M. W. Mullowney, S. Kautsar, J. H. Tryon, E. I.
573 Parkinson, E. L. C. De Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker, A.
574 Roeters, W. Lokhorst, A. Fernandez-Guerra, L. T. Dias Cappelini, R. J. Thomson, W. W.
575 Metcalf, N. L. Kelleher, F. Barona-Gomez and M. H. Medema, *bioRxiv*, 2018.
576 79. T. C. Vesth, J. L. Nybo, S. Theobald, J. C. Frisvad, T. O. Larsen, K. F. Nielsen, J. B.
577 Hoof, J. Brandl, A. Salamov, R. Riley, J. M. Gladden, P. Phatale, M. T. Nielsen, E. K.
578 Lyhne, M. E. Kogle, K. Strasser, E. McDonnell, K. Barry, A. Clum, C. Chen, K. LaButti,
579 S. Haridas, M. Nolan, L. Sandor, A. Kuo, A. Lipzen, M. Hainaut, E. Drula, A. Tsang, J.
580 K. Magnuson, B. Henrissat, A. Wiebenga, B. A. Simmons, M. R. Makela, R. P. de Vries,
581 I. V. Grigoriev, U. H. Mortensen, S. E. Baker and M. R. Andersen, *Nat. Genet.*, 2018, **50**,
582 1688-1695.
583 80. C. Greco, N. P. Keller and A. Rokas, *Curr. Opin. Microbiol.*, 2019, **51**, 22-29.

584

585 **Table 1. Glossary of evolutionary terms**

Term	Definition
BGC *de novo* assembly	Refers to the process by which an entire BGC is evolutionarily assembled through the recruitment and relocation of native genes, duplicates of native genes, and horizontally acquired genes
BGC duplication	Refers to the generation of an additional (duplicate) copy of an entire BGC in the genome
BGC functional divergence	Refers to the process by which homologous BGCs, through the accumulation of genetic changes, gradually diverge in their functions (i.e., in the secondary metabolites they produce)
BGC horizontal or lateral transfer	Refers to the process by which an entire BGC from the genome of one organism is transferred and stably integrated into the genome of another through non-reproduction related mechanisms
Deletion	Type of mutation, which stems from the deletion of genetic material in the genome
Duplication	Refers to the generation of an additional (duplicate) copy of genetic material in the genome
Homology / homologous	In the context of genes, two genes are homologous if their origins can be traced to the same common ancestor. Homologous genes can originate via processes such as

	gene duplication (in which case they are paralogs) and vertical descent / speciation (in which case they are orthologs)
Horizontal / lateral gene transfer	Refers to the transfer and integration of genetical material from the genome of one organism to the genome of another through non-reproduction related mechanisms
Insertion	Type of mutation, which stems from the insertion of genetic material in the genome
Orthology / orthologous	In the context of genes, two genes are orthologous if they originated via vertical descent / speciation, i.e., if they stem from the same ancestral gene that was present in the last common ancestor of the strains / species being compared
Paralogy / paralogous	In the context of genes, two genes are paralogous if they originated via gene duplication
Point mutation	Type of mutation, which stems from the replacement of one nucleotide base pair by another
Rearrangement	Type of mutation, which stems from the rearrangement of genetic material in the genome

586