

1 **Functional and evolutionary characterization of a secondary metabolite gene cluster in budding**  
2 **yeasts**

3 David J. Krause<sup>a,b</sup>, Jacek Kominek<sup>a,b</sup>, Dana A. Opulente<sup>a,b</sup>, Xing-Xing Shen<sup>c</sup>, Xiaofan Zhou<sup>c</sup>, Quinn K.  
4 Langdon<sup>a</sup>, Jeremy DeVirgilio<sup>d</sup>, Amanda Beth Hulfachor<sup>a</sup>, Cletus P. Kurtzman<sup>d,1</sup>, Antonis Rokas<sup>c</sup>, and  
5 Chris Todd Hittinger<sup>a,b,2</sup>

6 <sup>a</sup>Laboratory of Genetics, Genome Center of Wisconsin, Wisconsin Energy Institute, J. F. Crow Institute  
7 for the Study of Evolution, University of Wisconsin-Madison, Madison, WI 53706

8 <sup>b</sup>DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53706

9 <sup>c</sup>Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235

10 <sup>d</sup>Mycotoxin Prevention and Applied Microbiology Research Unit, National Center for Agricultural  
11 Utilization Research, Agricultural Research Service, US Department of Agriculture, Peoria, IL

12 <sup>1</sup>Deceased

13 <sup>2</sup>Corresponding author: [cthittinger@wisc.edu](mailto:cthittinger@wisc.edu)

14 **Abstract**

15 Secondary metabolites are key in how organisms from all domains of life interact with their environment  
16 and each other. The iron-binding molecule pulcherrimin was described a century ago, but the genes  
17 responsible for its production in budding yeasts have remained uncharacterized. Here we used  
18 phylogenomic footprinting on 90 genomes across the budding yeast subphylum Saccharomycotina to  
19 identify the gene cluster associated with pulcherrimin production. Using targeted gene replacements in  
20 *Kluyveromyces lactis*, we characterized the four genes that make up the cluster, which likely encode two  
21 pulcherriminic acid biosynthesis enzymes, a pulcherrimin transporter, and a transcription factor involved  
22 in both biosynthesis and transport. The requirement of a functional putative transporter to utilize  
23 extracellular pulcherrimin-complexed iron demonstrates that pulcherriminic acid is a siderophore, a  
24 chelator that binds iron outside the cell for subsequent uptake. Surprisingly, we identified homologs of the  
25 putative transporter and transcription factor genes in multiple yeast genera that lacked the biosynthesis  
26 genes and could not make pulcherrimin, including the model yeast *Saccharomyces cerevisiae*. We deleted  
27 these previously uncharacterized genes and showed they are also required for pulcherrimin utilization in  
28 *S. cerevisiae*, raising the possibility that other genes of unknown function are linked to secondary  
29 metabolism. Phylogenetic analyses of this gene cluster suggest that pulcherrimin biosynthesis and  
30 utilization were ancestral to budding yeasts, but the biosynthesis genes, and subsequently the utilization  
31 genes, were lost in many lineages, mirroring other microbial public goods systems that lead to the rise of  
32 cheater organisms.

33 **Significance Statement**

34 Evolutionary and comparative genomics, combined with reverse genetics, have the power to identify and  
35 characterize new biology. Here we use these approaches in several non-traditional model species of  
36 budding yeasts to characterize the first budding yeast secondary metabolite gene cluster, a set of genes  
37 responsible for production and reutilization of the siderophore pulcherrimin. We also use this information

38 to assign roles in pulcherrimin utilization for two previously uncharacterized *S. cerevisiae* genes. The  
39 evolution of this gene cluster in budding yeasts suggests an ecological role for pulcherrimin akin to other  
40 microbial public goods systems.

41 /body

## 42 **Introduction**

43 The production of secondary metabolites is found in organisms from all domains of life (1–3). Fungi are  
44 particularly well known for their production of secondary metabolites, whose functions can include  
45 antibiotics, virulence factors, pigments, toxins, quorum-sensing molecules, and siderophores (chelators  
46 produced by organisms to capture extracellular iron) (3, 4). These molecules are generally synthesized by  
47 proteins encoded in gene clusters, typically containing a synthase protein belonging to one of several  
48 families: a non-ribosomal peptide synthetase (NRPS), polyketide synthase (PKS), terpene cyclase,  
49 dimethylallyl tryptophan synthetase, or cyclodipeptide synthase (CDPS) (2, 5–7). The absence of known  
50 genes encoding any of these functions in sequenced budding yeast genomes has precluded an  
51 understanding of secondary metabolism in the budding yeast subphylum Saccharomycotina (8).

52 The pigment pulcherrimin has long been known to be synthesized by a small number of budding yeast  
53 species, including various *Metschnikowia* spp. and some *Kluyveromyces* spp. (9–12). Pulcherrimin is a  
54 red iron-containing pigment composed of two cyclized and modified leucine molecules (13). The  
55 ecological role of pulcherrimin is not well understood, but several studies have shown it to mediate  
56 interspecific antagonistic interactions, possibly due to its ability to sequester ferric iron from the growth  
57 medium (14–16). The ability to sequester free iron from the environment from competitors could be  
58 counter-productive if the producer species were also unable to utilize the pulcherrimin-bound iron, raising  
59 the possibility that pulcherrimin producers can also re-utilize the compound as a siderophore.  
60 Pulcherrimin may not be the only secondary metabolite produced by budding yeasts; there is evidence for  
61 ferrichrome production by some *Lipomyces* and *Dipodascopsis* spp. (17), but the genes responsible for  
62 production of both secondary metabolites in yeasts are unknown. Biochemical work in the pulcherrimin-  
63 producing bacterium *Bacillus subtilis* has characterized a two-step biosynthetic pathway: two leucine  
64 molecules are cyclized via a CDPS, and the resulting diketopiperazine is oxidized by a cytochrome P450  
65 oxidase to generate pulcherriminic acid (PA), which can then spontaneously bind iron to form  
66 pulcherrimin (18–20). The formation of diketopiperazines, such as the cyclo-Leu-Leu precursor to PA, is  
67 often catalyzed by biochemical pathways containing NRPS or CDPS proteins (21). Thus, the absence of  
68 known homologs of these genes in budding yeasts raises the possibilities of highly divergent homologs,  
69 or novel biochemical pathways for pulcherrimin production.

70 Here we use a comparative genomic approach to identify a putative gene cluster involved in pulcherrimin  
71 production. Using targeted gene replacements in *Kluyveromyces lactis*, we characterize this four-gene  
72 cluster, finding genes responsible for the biosynthesis and re-utilization of pulcherrimin. We also find that  
73 several species, including *S. cerevisiae*, contain a partial gene cluster comprised of only the two  
74 utilization genes. Using targeted gene replacements in *S. cerevisiae*, we assign functions and standard  
75 names to these previously uncharacterized genes. Finally, we infer that the gene cluster was ancestral to  
76 all budding yeasts, but repeated gene loss led to its patchy species distribution, as well as to the rise of  
77 cheater organisms that exploit the public good of pulcherrimin without the cost of production.

## 78 RESULTS

### 79 **A secondary metabolite gene cluster is responsible for production of the siderophore pulcherrimin.**

80 Pulcherrimin biosynthesis in *B. subtilis* consists of the cyclization of two leucine molecules via a CDPS  
81 encoded by *yvmC*, followed by oxidation via a cytochrome P450 oxidase encoded by *cypX* (18–20).  
82 When we searched a custom BLAST database composed of all publicly available Saccharomycotina  
83 genome assemblies (<http://y1000plus.org/blast>) (22), we found no homologs for *yvmC* and only one  
84 significant hit ( $e < 0.001$ ) for *cypX*: the widely conserved sterol biosynthesis gene *ERG11* (23). Thus, the  
85 *B. subtilis* genes did not suggest candidates for the budding yeast pathway. We reasoned that a more  
86 divergent cytochrome P450 oxidase could be involved in pulcherrimin biosynthesis, so we searched the  
87 genome of the pulcherrimin producer *K. lactis* for proteins with a P450 conserved domain using NCBI's  
88 CDD tool and BLASTp. This approach yielded five P450 superfamily genes, one of which had no  
89 homologs in *S. cerevisiae*, and was annotated in GenBank as a “*gliC* homolog.” *gliC* encodes a  
90 cytochrome P450 oxidase involved in secondary metabolism in some species of filamentous fungi,  
91 including the biosynthesis of gliotoxin by *Aspergillus* spp. (24, 25). We searched the Saccharomycotina  
92 BLAST database for homologs of this protein and found significant hits ( $e < 0.001$ ) in the genomes of  
93 several *Kluyveromyces* spp., *Candida auris*, and *Metschnikowia fructicola*.

94 Among these species, we also found conservation of three genes surrounding the cytochrome P450  
95 oxidase homolog, hereafter termed the *PUL* (*PUL*cherrimin) gene cluster, with genes labeled *PUL1-4*  
96 (Figure 1, SI Appendix, Fig. S1). In addition to *K. lactis* (10), *Kluyveromyces aestuarii* and *M. fructicola*  
97 had been described as producing pulcherrimin (11, 12). We sequenced the genomes of two more budding  
98 yeast species: one previously described as producing an unknown red pigment, *Zygorulasporea mrakii*  
99 (26), from which we identified a *PUL* gene cluster; and one from the other recognized species of this  
100 genus, *Zygorulasporea florentina*, a non-pigmented strain that did not contain a complete *PUL* cluster.  
101 Thus, in each of the three genera, *Kluyveromyces*, *Metschnikowia*, and *Zygorulasporea*, we found that  
102 red-pigmented species always have complete *PUL* clusters, whereas species lacking the complete *PUL*  
103 cluster were never pigmented (Figure 1). A handful of species have complete *PUL* clusters but did not  
104 readily produce pulcherrimin during laboratory growth, nor did their species descriptions include a  
105 positive pulcherrimin trait. These species may be cryptic producers that make pulcherrimin under  
106 specific, yet untested, conditions (e.g. *C. auris*, and some *Kluyveromyces* spp.).

107 To directly test whether this gene cluster was responsible for pulcherrimin production, we generated  
108 targeted gene replacements in the genetically tractable, pulcherrimin-producing yeast *K. lactis* (27, 28).  
109 We replaced each of the four *PUL* genes individually with an antibiotic-resistance cassette and tested the  
110 mutant strains for pulcherrimin production on a defined, synthetic complete (SC) agar medium. Under  
111 these conditions, the wild-type strain produced a red pigment, a trait that was abolished in *pul1Δ*, *pul2Δ*,  
112 and *pul4Δ* mutants (Figure 2A). Complementation by reinsertion of each deleted gene restored  
113 pigmentation to all mutants (SI Appendix, Fig. S2). To further characterize these genes, we performed  
114 cross-feeding experiments between mutants. A *pul2Δ* strain constitutively expressing *PUL1* conferred  
115 pigmentation to a *pul1Δ* strain constitutively expressing *PUL2*, but not vice versa, strongly suggesting  
116 that *PUL1* is upstream of *PUL2* in the biochemical pathway (SI Appendix, Fig. S3). Further, while *pul4Δ*  
117 strains constitutively expressing *PUL1* or *PUL2* were not pigmented on their own, we observed the same  
118 cross-feeding relationship in this deletion strain, indicating that *PUL4* is not required for pulcherrimin  
119 biosynthesis when *PUL1* or *PUL2* are expressed heterologously (SI Appendix, Fig. S3). We also

120 attempted to express the *B. subtilis yvmC* and *cypX* genes in *pul1Δ* and *pul2Δ* strains, respectively, but  
121 they did not successfully complement the *K. lactis* mutants (SI Appendix, Fig. S4). Given these data and  
122 the putative annotations (Figure 1), we hypothesize that *PUL4* encodes a transcription factor that  
123 regulates *PUL1* and *PUL2*, which we hypothesize encode the enzymes responsible for the two-step  
124 pulcherrimin biosynthesis pathway (Figure 2D).

125 **The *PUL* cluster also contains genes responsible for pulcherrimin-complexed iron utilization.** While  
126 the *pul3Δ* mutant still produced pulcherrimin, it grew poorly on SC (Figure 2). Several observations led  
127 us to hypothesize that the *pul3Δ* mutant was unable to re-utilize the pulcherrimin it produced: growth on  
128 SC often gave rise to  $\text{Pul}^-$  suppressors that grew normally but did not produce pulcherrimin, and a *pul1Δ*  
129 *pul3Δ* double mutant grew normally but did not produce pulcherrimin (SI Appendix, Fig. S5). Since the  
130 *pul3Δ* mutant strains grew poorly when they produced pulcherrimin, we hypothesized that the *PUL3* gene  
131 was required to bring pulcherrimin-complexed iron back into the cell, perhaps by a mechanism similar to  
132 the siderophore-iron transporters characterized in *S. cerevisiae*, which are encoded by genes that include  
133 *ENB1*, *SIT1*, *ARN1*, and *ARN2* (29).

134 To test the hypothesis that *PUL3* is required for the uptake of iron using pulcherrimin, we devised a  
135 growth assay where strains were grown in liquid SC, with and without the addition of exogenous  
136 pulcherriminic acid (PA). PA is pulcherrimin that has been extracted away from iron using strong base  
137 and readily binds iron in growth media. Thus, addition of PA to growth medium sequesters available iron  
138 in the form of pulcherrimin. Growth rates for wild-type *K. lactis* in PA-treated medium were not  
139 significantly different from growth rates in untreated control medium (SI Appendix, Fig. S6). In PA-  
140 treated medium, wild-type, *pul1Δ*, and *pul2Δ* mutants grew normally, implying that *PUL1* and *PUL2* play  
141 a role specific to pulcherrimin biosynthesis and not re-utilization (Figure 2). Although the *pul3Δ* mutant  
142 grew normally in YPD liquid, it had a lower maximum cell density in SC than the wild-type strain, in line  
143 with the observations on solid medium (SI Appendix, Fig. S6). The *pul3Δ* strain grew negligibly  
144 compared to wild-type in PA-treated medium (Student's t-test,  $p = 0.001$ ; Figure 2B), demonstrating that  
145 *PUL3* is required for utilization of pulcherrimin-complexed iron.

146 Along with the role of *PUL4* in pulcherrimin biosynthesis (Figure 2D), the *pul4Δ* mutant also showed a  
147 significant lag during growth in PA-treated medium compared to wild-type cells (Student's t-test,  $p =$   
148  $0.04$ ). After the lag, the *pul4Δ* strain did not show a significant growth decrease compared to wild-type  
149 cells (Student's t-test,  $p = 0.15$ , Figure 2B). These results suggest that *PUL4* also positively regulates  
150 *PUL3*, but there are likely other ways for the cells to activate *PUL3* expression in *K. lactis* (Figure 2B).

151 ***S. cerevisiae* and some other yeast species have pulcherrimin-complexed iron utilization genes.** In  
152 addition to the complete *PUL* clusters present in all pulcherrimin-producing yeast species, we also found  
153 that 12 species contain partial *PUL* gene clusters composed of *PUL3* and *PUL4*, or only *PUL4* (Figure 1).  
154 We hypothesized that organisms containing *PUL3*, all of which also contain *PUL4*, might be able to  
155 utilize exogenous pulcherrimin, but not make their own. One such species is the model yeast *S.*  
156 *cerevisiae*, and its *PUL3* and *PUL4* homologs are the adjacent and previously uncharacterized genes  
157 *YNR062C* and *YNR063W*, respectively (Figure 3A). To test whether these genes also play a role in  
158 pulcherrimin-complexed iron utilization in *S. cerevisiae*, we individually deleted *YNR062C* and  
159 *YNR063W* and grew both mutant strains in PA-treated media. Wild-type *S. cerevisiae* grew similarly in  
160 PA-treated and untreated growth media (SI Appendix, Fig. S7), but both the *pul3Δ* and *pul4Δ* mutant

161 strains failed to grow in PA-treated medium (Figure 3B). *PUL4* has been shown to bind DNA in *S.*  
162 *cerevisiae* via protein-binding microarrays (30), and we therefore hypothesize, as in *K. lactis*, that *PUL4*  
163 encodes a transcription factor that regulates expression of the putative transporter *PUL3*.

164 To determine whether the presence or absence of *PUL3* and *PUL4* homologs was correlated with growth  
165 in PA-treated medium for other, less genetically tractable yeast species, we examined several  
166 monophyletic genera where species varied in the presence of *PUL3* and *PUL4*. Five genera contain  
167 examples of species that contain *PUL3* and *PUL4* and species that lack both genes, or at least *PUL3*. For  
168 each genus, we grew one species containing *PUL3* and *PUL4* and one species lacking these genes in  
169 media with and without PA treatment. Species lacking *PUL3* and *PUL4* showed significantly slower  
170 growth in PA-treated medium, relative to growth in untreated medium, when compared to closely related  
171 species containing the genes (Student's t-test,  $p = 0.011$ ; Figure 3C), suggesting that these genes enable  
172 pulcherrimin-mediated iron uptake across a broad range of budding yeast taxa.

173 **The rarity of pulcherrimin production is the result of pervasive gene loss.** The sparse distribution of  
174 the *PUL* gene cluster within budding yeasts implied either that the cluster was ancestral to the subphylum  
175 Saccharomycotina and lost in most extant lineages or that it was gained via several horizontal gene  
176 transfer events. Horizontal gene transfer, including transfer of entire metabolic gene clusters, has been  
177 observed in both budding yeasts and fungi in general, but it is considered to be quite rare (31–34). We  
178 identified significant ( $e < 0.001$ ) BLAST hits in the GenBank non-redundant (nr) protein sequence  
179 database for all four *PUL* genes in the Pezizomycotina and Basidiomycota, ranging between 22–30%  
180 identity at the amino acid level. In other fungi, the gene neighborhoods of *PUL1* homologs were often  
181 predicted to encode cytochrome P450 oxidases, MFS transporters, and transcription factors, indicating  
182 that the *PUL* cluster may exist in other fungi (SI Appendix, Fig. S8). When we constructed maximum-  
183 likelihood (ML) phylogenies for each of the fungal *PUL* genes, all four gene trees supported a single  
184 Saccharomycotina clade with 100% bootstrap support, showing a shared origin for all Saccharomycotina  
185 *PUL* gene copies (SI Appendix, Figs. S9-S12). Thus, there is no evidence that individual species within  
186 Saccharomycotina obtained the *PUL* genes from a known source outside of the subphylum.

187 We also tested for HGT within the subphylum by using one-sided Kishino-Hasegawa (KH) (35) and  
188 Approximately Unbiased (AU) tests (36) to test for significant differences between likelihood values for  
189 the gene trees and corresponding trees constrained on the genome-wide species topology (Figure 1). None  
190 of these tests identified significant differences between the gene trees and species tree (SI Appendix,  
191 Table S3), meaning the phylogenies are fully compatible with accepted yeast relationships and suggest no  
192 HGT within the subphylum. In the absence of evidence for HGT, and given the rarity of HGT in fungi,  
193 we infer that the *PUL* gene cluster was present in the common ancestor of Saccharomycotina. Its patchy  
194 distribution is best explained by parallel loss events during evolution, a feature commonly observed in  
195 secondary metabolic gene clusters (33, 34), rather than rare acquisitions via HGT.

## 196 DISCUSSION

197 **Non-traditional model systems reveal novel biology.** We have discovered and characterized the first  
198 secondary metabolite gene cluster in budding yeasts. This gene cluster is responsible for the production of  
199 the siderophore pulcherrimin to capture iron from the environment. The rarity of pulcherrimin production  
200 among yeast species made the identification and characterization of the underlying genes dependent upon  
201 knowledge of the genome sequences of diverse yeast species for comparative genomics and, fortunately, a

202 genetically tractable system in *K. lactis*. Our study underscores the importance of efforts to sequence the  
203 genomes of diverse species (37–39), as well as to explore the biology of both established and non-  
204 traditional model systems. In fact, understanding the pulcherrimin system in *K. lactis* was essential to  
205 characterizing *PUL3* and *PUL4* (*YNR062C* and *YNR063W*) in *S. cerevisiae*, two genes that had previously  
206 evaded characterization by several high-throughput screens of gene function (40, 41).

207 Unlike the other 3 genes required for pulcherrimin biosynthesis, the predicted coding sequence of *PUL1*  
208 lacks obvious sequence similarity to known families. The biochemical function of Pul1 remains  
209 unverified, as direct complementation of *pul1Δ* by expressing bacterial *yvmC* was unsuccessful, and  
210 attempts at recombinant protein expression for biochemical characterization so far have been  
211 unsuccessful. The lack of complementation observed from the bacterial genes could result from lack of  
212 recognition of yeast tRNA-Leu, different cofactor requirements, or trivial differences that prevent protein  
213 expression, stability, or function. The possibility also remains that the biochemistry of the yeast  
214 pulcherrimin biosynthesis pathway differs from that of bacteria, and further work will be required to  
215 elucidate this.

216 We have also identified homologs of *PUL1* in filamentous fungi, where their gene neighborhoods  
217 frequently resemble the *PUL* clusters characterized in this study and many other fungal secondary  
218 metabolite gene clusters. It is unknown whether these *PUL1*-containing clusters in diverse fungi produce  
219 pulcherrimin, but it seems likely that they are involved in secondary metabolite production. Searching for  
220 secondary metabolite gene clusters is often centered around identifying genes encoding NRPS or PKS  
221 using computational programs, such as antiSMASH (42) and SMURF (43), and genes encoding CDPS  
222 have not been comprehensively incorporated into these tools. Incorporating known CDPS and the novel  
223 *PUL1* gene into future scans could help identify new secondary metabolite gene clusters (42, 44).

224 **The ecological role of pulcherrimin.** The rarity of pulcherrimin production among budding yeast species  
225 has also made understanding its ecological role difficult. However, by characterizing the genes in the  
226 *PUL* locus, we have identified several cryptic producers that have full *PUL* clusters but do not produce  
227 pulcherrimin in laboratory conditions, including some *Kluyveromyces* spp. and the emerging pathogen *C.*  
228 *auris*. Also, several species contain partial clusters that confer the ability to utilize pulcherrimin without  
229 producing it. Ultimately, we infer that the ancestor of all budding yeasts contained a complete *PUL*  
230 cluster, but the genes were lost in parallel in most lineages during yeast evolution.

231 Pulcherrimin has been previously described as mediating antagonistic interactions toward non-producing  
232 yeast species, a property proposed to be linked to the sequestration of free iron from the environment (14,  
233 15). However, without the ability to re-utilize this pulcherrimin-bound iron, producers would inhibit their  
234 own growth, as observed in the *K. lactis pul3Δ* mutant. Therefore, in this study, we have expanded the  
235 ecological and functional role of pulcherrimin from solely an antagonistic compound to also include a  
236 role as an iron-binding siderophore. The pulcherrimin iron-chelation system is peculiar in budding yeasts;  
237 most characterized bacterial siderophores are used to scavenge for scarce iron in the environment (45),  
238 but many yeasts produce pulcherrimin even under high-iron conditions. Thus, pulcherrimin's role in  
239 budding yeasts may not be in iron-scavenging, but rather iron-monopolizing.

240 **Gene loss and the public goods problem.** We have inferred that the *PUL* gene cluster was present in the  
241 common ancestor of Saccharomycotina, and the resulting sparse distribution of species containing the  
242 *PUL* genes, including the presence of partial *PUL* gene clusters, was the result of many gene loss events.

243 We identified four different genotypes: 1) the presence of all four *PUL* genes, 2) the presence of *PUL3*  
244 and *PULA*, 3) the presence of *PULA* only, and 4) the absence of *PUL* genes. We hypothesize that gene  
245 loss has occurred, and continues to occur, in order from 1-4 (Figure 4). This ordered loss would be  
246 consistent with the observation of a growth defect in the *K. lactis pul3Δ* mutant. While many of these loss  
247 events likely occurred in the deep branches of the phylogeny, comparisons with close relatives suggest  
248 several recent loss events between species, which may or may not reflect entire species given the use of a  
249 single strain for these analyses. The *PUL* genes are localized to subtelomeres, which are often associated  
250 with gene content variation, as well as sources of evolutionary novelty (46, 47).

251 In characterizing the phenotypes resulting from gene loss events, we propose a simple model. The  
252 minimal cost of pulcherrimin production is the production of two leucine molecules and the likely  
253 expenditure of ATP to cyclize them. There is no apparent cost to utilizing pulcherrimin, except for  
254 production of the Pul3 putative transporter and Pul4 transcription factor. Therefore, we predict that cells  
255 that lose the biosynthesis genes avoid the costs of pulcherrimin production, but still reap the potential  
256 benefits of pulcherrimin production from neighboring cells. Similar public goods dynamics are common  
257 in microbes, in which cells producing a good are exploited by non-producing cells that consume the good  
258 without the costs of production. Similar scenarios of gene loss in siderophore biosynthesis genes, with  
259 retention of the utilization genes, have been observed and described in this context in *Vibrio* and  
260 *Pseudomonas* spp. (48, 49). By combining the public goods problem with the patterns of gene loss  
261 without regain, we propose an evolutionary model where the initial loss of the biosynthesis genes allows a  
262 lineage to exploit closely related producers and rise in frequency. However, if they outcompete the  
263 producers too strongly and replace them, there is no more pulcherrimin to exploit, and the remaining  
264 utilization genes will eventually be lost. As more lineages lose pulcherrimin utilization, the rare lineages  
265 that have retained production may regain the upper hand. These dynamics may explain the many  
266 independent *PUL* gene cluster losses during budding yeast evolution.

267 A similar public goods scenario has been studied in laboratory experiments using *S. cerevisiae*, where  
268 *suc2* mutants are able to exploit the invertase production of cells with functional *SUC2* genes by utilizing  
269 the byproducts of sucrose digestion in mixed culture (50). However, a study of wild isolates did not  
270 identify any naturally occurring ‘cheaters’ that lacked invertase production (51). Here we showed that  
271 genetic variation in the *PUL* locus exists between species, but we have only investigated a single strain  
272 for each species. Further investigations of the biology of distinct strains within species will be required to  
273 test evolutionary hypotheses at more recent timescales. For example, a single strain of *K. marxianus*,  
274 UFS-Y2791 (52), contains homologs of *PUL1* and *PUL2*, differing from the type strain, which lacks  
275 them.

276 **The power of exploring biodiversity.** The clustered arrangement of genes encoding secondary  
277 metabolite functions is common in other fungi, and the *PUL*cherrimin gene cluster in budding yeasts is  
278 exceptional mainly because it is found in a group of fungi previously thought to lack such clusters. As the  
279 abundance of genome-scale data continues to increase in the subphylum Saccharomycotina (37), we  
280 predict that the genetic pathways for additional metabolites will be discovered and characterized in these  
281 diverse yeasts as it becomes clear that budding yeasts harbor many traits that the model yeast *S. cerevisiae*  
282 lacks. Exploring novel biology in non-traditional model organisms will continue to shed light on  
283 uncharacterized aspects of even well-studied model organisms, enhance our understanding of yeast  
284 ecology, and reveal both novel and general biology.

## 285 MATERIALS AND METHODS

286 Detailed materials and methods can be found in the *SI appendix*.

287 **Strains, media, and oligonucleotides used in this study.** All genetic work was performed in a *MATa*  
288 *ku80-Δ* NHEJ-deficient background derived from *K. lactis* CBS2359 (27, 28) and in a *MATa his3Δ1*  
289 *leu2Δ0 lys2Δ0 ura3Δ0* background of *S. cerevisiae* (BY4741) (53). Most yeast species used were obtained  
290 from the ARS Culture Collection (NRRL) in Peoria, IL. Media formulations can be found in the *SI*  
291 *appendix*. A complete list of strains and oligonucleotides used can be found in SI Appendix Tables S1  
292 and S2. Genetic manipulations in *K. lactis* and *S. cerevisiae* were performed according to standard  
293 protocols.

294 **Pulcherriminic acid extraction and treatment.** Extraction of pulcherriminic acid (PA) was performed  
295 similarly to previously described protocols (9) using a strain of *Metschnikowia* sp., yHQL305. This strain  
296 is an avid pulcherrimin producer that we isolated from Garry oak (*Quercus garryana*) leaves from the  
297 Standish-Hickey State Recreational Area in California using established protocols (54). Extraction details  
298 can be found in the *SI appendix*.

299 **Growth experiments and analysis.** Growth experiments were performed in 96-well plates on a  
300 FLUOstar Omega plate reader (BMG Labtech). Experiments using only *S. cerevisiae* and *K. lactis* were  
301 performed at 30°C, while experiments using several yeast species were performed at 22°C to permit the  
302 growth of species that prefer cooler temperatures. Three biological replicates were performed of all  
303 growth experiments. Growth data were analyzed using the grofit package in R under the logistic model of  
304 growth (55); plots were generated with ggplot2 (56). Spline-fitted values were used when growth was  
305 negligible, causing the model to fail. Representative growth curves shown in the *Results* section are the  
306 biological replicate whose growth rate is closest to the mean growth rate for the set of replicates.  
307 Statistical comparisons of grofit parameters were performed using paired Student's t-tests with all pairs  
308 grown concurrently on the same plate using the same reagent batch.

309 **Sequence alignment, phylogeny construction, and topology testing.** Phylogenetic analyses were  
310 performed according to standard methods. Details can be found in the *SI Methods*.

311 **Data availability.** Genome assemblies for *Z. mrakii*, *Z. florentina*, *K. nonfermentans*, and *Ko.*  
312 *pseudopastoris* can be found under NCBI accession numbers PPHZ000000000, PPJY000000000,  
313 PPKX000000000, and QYLQ000000000.

## 314 ACKNOWLEDGMENTS

315 We thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing  
316 Illumina and Sanger sequencing facilities and services, Russell Wrobel for helpful suggestions on  
317 oligonucleotide design, Gilles Fischer for providing *L. thermotolerans* Y-8284, Kevin Myers and Audrey  
318 Gasch for providing *S. cerevisiae* strains, Jue Wang for providing *Bacillus subtilis* genomic DNA,  
319 Maitreya Dunham for providing the pIL68 plasmid, and William Vagt for collecting the sample from  
320 which yHQL305 was isolated. This material is partly based upon work supported by the U.S. Department  
321 of Energy, Office of Science, Office of Biological and Environmental Research under Award Number  
322 DE-SC0018409, and work funded by the DOE Great Lakes Bioenergy Research Center (DOE BER  
323 Office of Science DE-FC02-07ER64494), as well as the National Science Foundation under Grant Nos.

324 DEB-1442148 (to CTH and CPK), DEB-1442113 (to AR), and DEB-1253634 (to CTH). CTH is a Pew  
325 Scholar in the Biomedical Sciences and Vilas Faculty Early Career Investigator, supported by the Pew  
326 Charitable Trusts and the Vilas Trust Estate, respectively. QKL was supported by a National Science  
327 Foundation Graduate Research Fellowship (DGE-126259) and the National Institutes of Health  
328 Predoctoral Training Program for the UW-Madison Laboratory of Genetics (5T32GM007133). Mention  
329 of trade names or commercial products in this publication is solely for the purpose of providing specific  
330 information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.  
331 USDA is an equal opportunity provider and employer.

## 332 **References**

- 333 1. Charlesworth JC, Burns BP (2015) Untapped resources: biotechnological potential of peptides and  
334 secondary metabolites in Archaea. *Archaea Vanc BC* **2015**: 282035.
- 335 2. Donadio S, Monciardini P, Sosio M (2007) Polyketide synthases and nonribosomal peptide  
336 synthetases: the emerging view from bacterial genomics. *Nat Prod Rep* **24**:1073–1109.
- 337 3. Macheleidt J, et al. (2016) Regulation and role of fungal secondary metabolites. *Annu Rev Genet*  
338 **50**:371–392.
- 339 4. Keller NP, Turner G, Bennett JW (2005) Fungal secondary metabolism — from biochemistry to  
340 genomics. *Nat Rev Microbiol* **3**:937–947.
- 341 5. Finking R, Marahiel MA (2004) Biosynthesis of nonribosomal peptides. *Annu Rev Microbiol* **58**:453–  
342 488.
- 343 6. Weissman KJ, Leadlay PF (2005) Combinatorial biosynthesis of reduced polyketides. *Nat Rev*  
344 *Microbiol* **3**:925–936.
- 345 7. Gondry M, et al. (2009) Cyclodipeptide synthases are a family of tRNA-dependent peptide bond–  
346 forming enzymes. *Nat Chem Biol* **5**:414–420.
- 347 8. Bushley KE, Turgeon BG (2010) Phylogenomics reveals subfamilies of fungal nonribosomal peptide  
348 synthetases and their evolutionary relationships. *BMC Evol Biol* **10**:26.
- 349 9. Kluver AJ, van der Walt JP, van Triet AJ (1953) Pulcherrimin, The pigment of *Candida pulcherrima*.  
350 *Proc Natl Acad Sci USA* **39**:583–593.
- 351 10. Kurtzman C, Fell JW, Boekhout T (2011) *The Yeasts: A Taxonomic Study*. (Elsevier).
- 352 11. Lachance MA (2016) *Metschnikowia*: half tetrads, a regicide and the fountain of youth. *Yeast* **33**:563–  
353 574.
- 354 12. Araujo FV, Hagler AN (2011) *Kluyveromyces aestuarii*, a potential environmental quality indicator  
355 yeast for mangroves in the State of Rio de Janeiro, Brazil. *Braz J Microbiol* **42**:954–958.
- 356 13. MacDonald JC (1965) Biosynthesis of pulcherriminic acid. *Biochem J* **96**:533–538.

- 357 14. Türkel S, Ener B (2009) Isolation and characterization of new *Metschnikowia pulcherrima* strains as  
358 producers of the antimicrobial pigment pulcherrimin. *Z Naturforschung C J Biosci* **64**:405–410.
- 359 15. Sipiczki M (2006) *Metschnikowia* strains isolated from botrytized grapes antagonize fungal and  
360 bacterial growth by iron depletion. *Appl Environ Microbiol* **72**:6716–6724.
- 361 16. Oro L, Ciani M, Comitini F (2014) Antimicrobial activity of *Metschnikowia pulcherrima* on wine  
362 yeasts. *J Appl Microbiol* **116**:1209–1217.
- 363 17. Van Der Walt JP, Botha A, Eicker A (1990) Ferrichrome production by Lipomycetaceae. *Syst Appl*  
364 *Microbiol* **13**:131–135.
- 365 18. Bonnefond L, et al. (2011) Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA  
366 synthetase paralog. *Proc Natl Acad Sci USA* **108**:3912–3917.
- 367 19. Randazzo P, Aubert-Frambourg A, Guillot A, Auger S (2016) The MarR-like protein PchR (YvmB)  
368 regulates expression of genes involved in pulcherriminic acid biosynthesis and in the initiation of  
369 sporulation in *Bacillus subtilis*. *BMC Microbiol* **16**:190.
- 370 20. Cryle MJ, Bell SG, Schlichting I (2010) Structural and biochemical characterization of the  
371 cytochrome P450 CypX (CYP134A1) from *Bacillus subtilis*: a cyclo-L-leucyl-L-leucyl dipeptide oxidase.  
372 *Biochemistry (Mosc)* **49**:7282–7296.
- 373 21. Gu B, He S, Yan X, Zhang L (2013) Tentative biosynthetic pathways of some microbial  
374 diketopiperazines. *Appl Microbiol Biotechnol* **97**:8439–8453.
- 375 22. Shen XX, et al. (2016) Reconstructing the backbone of the Saccharomycotina yeast phylogeny using  
376 genome-scale data. *G3* **6**:3927–3939.
- 377 23. Karst F, Lacroute F (1977) Ergosterol biosynthesis in *Saccharomyces cerevisiae*: mutants deficient in  
378 the early steps of the pathway. *Mol Gen Genet* **154**:269–277.
- 379 24. Gardiner DM, Howlett BJ (2005) Bioinformatic and expression analysis of the putative gliotoxin  
380 biosynthetic gene cluster of *Aspergillus fumigatus*. *FEMS Microbiol Lett* **248**:241–248.
- 381 25. Gardiner DM, Cozijnsen AJ, Wilson LM, Pedras MSC, Howlett BJ (2004) The sirodesmin  
382 biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans*. *Mol Microbiol*  
383 **53**:1307–1318.
- 384 26. Capriotti A (1958) *Zygosaccharomyces mrakii nova* species. *Arch Für Mikrobiol* **30**:387–392.
- 385 27. Hittinger CT, Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic  
386 switch. *Nature* **449**:677–681.
- 387 28. Kooistra R, Hooykaas PJJ, Steensma HY (2004) Efficient gene targeting in *Kluyveromyces lactis*.  
388 *Yeast* **21**:781–792.
- 389 29. Philpott CC (2006) Iron uptake in fungi: a system for every source. *Biochim Biophys Acta Mol Cell*  
390 *Res* **1763**:636–645.

- 391 30. Zhu C, et al. (2009) High-resolution DNA-binding specificity analysis of yeast transcription factors.  
392 *Genome Res* **19**:556-566.
- 393 31. Fitzpatrick DA (2012) Horizontal gene transfer in fungi. *FEMS Microbiol Lett* **329**:1–8.
- 394 32. Novo M, et al. (2009) Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence  
395 of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc Natl Acad Sci USA* **106**:16333–16338.
- 396 33. Wisecaver JH, Slot JC, Rokas A (2014) The evolution of fungal metabolic pathways. *PLoS Genet*  
397 **10**:e1004816.
- 398 34. Wisecaver, JH & Rokas A (2015) Fungal metabolic gene clusters-caravans traveling across genomes  
399 and environments. *Front Microbiol* **6**:161.
- 400 35. Goldman N, Anderson JP, Rodrigo AG, Olmstead R (2000) Likelihood-based tests of topologies in  
401 phylogenetics. *Syst Biol* **49**:652–670.
- 402 36. Shimodaira H, Goldman N (2002) An approximately unbiased test of phylogenetic tree selection. *Syst*  
403 *Biol* **51**:492-508
- 404 37. Hittinger CT, et al. (2015) Genomics and the making of yeast biodiversity. *Curr Opin Genet Dev*  
405 **35**:100–109.
- 406 38. Wu D, et al. (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*  
407 **462**:1056–1060.
- 408 39. Grigoriev IV, et al. (2011) Fueling the future with fungal genomics. *Mycology* **2**:192–209.
- 409 40. Costanzo M, et al. (2016) A global genetic interaction network maps a wiring diagram of cellular  
410 function. *Science* **353**:aaf1420.
- 411 41. Usaj M, et al. (2017) TheCellMap.org: A web-accessible database for visualizing and mining the  
412 global yeast genetic interaction network. *G3* **7**:1539–1549.
- 413 42. Blin K, et al. (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster  
414 boundary identification. *Nucleic Acids Res.* **45**:W36-W41
- 415 43. Khaldi N, et al. (2010) SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal*  
416 *Genet Biol* **47**:746-741.
- 417 44. Dejong CA, et al. (2016) Polyketide and nonribosomal peptide retro-biosynthesis and global gene  
418 cluster matching. *Nat Chem Biol* **12**:1007.
- 419 45. Neilands JB (1995) Siderophores: structure and function of microbial iron transport compounds. *J*  
420 *Biol Chem* **270**:26723–26726.
- 421 46. Brown CA, Murray AW, Verstrepen KJ (2010) Rapid expansion and functional divergence of  
422 subtelomeric gene families in yeasts. *Curr Biol.* **20**:895-903.

- 423 47. Snoek T, Voordeckers K, Verstrepen KJ (2014) Subtelomeric regions promote evolutionary  
424 innovation of gene families in yeast. *Subtelomeres* (Springer, Berlin, Heidelberg).
- 425 48. Cordero OX, Ventouras LA, DeLong EF, Polz MF (2012) Public good dynamics drive evolution of  
426 iron acquisition strategies in natural bacterioplankton populations. *Proc Natl Acad Sci USA* **109**:20059–  
427 20064.
- 428 49. Butaitė E, Baumgartner M, Wyder S, Kümmerli R (2017) Siderophore cheating and cheating  
429 resistance shape competition for iron in soil and freshwater *Pseudomonas* communities. *Nat Commun*  
430 **8**:414.
- 431 50. Greig D, Travisano M (2004) The Prisoner’s Dilemma and polymorphism in yeast *SUC* genes. *Proc R*  
432 *Soc B Biol Sci* **271**:S25–S26.
- 433 51. Bozdag GO, Greig D (2014) The genetics of a putative social trait in natural populations of yeast. *Mol*  
434 *Ecol* **23**:5061–5071.
- 435 52. Schabort DTWP, Letebele PK, Steyn L, Kilian SG, du Preez, JC. (2016) Differential RNA-seq, multi-  
436 network analysis and metabolic regulation analysis of *Kluyveromyces marxianus* reveals a  
437 compartmentalised response to xylose. *PLOS ONE* **11**:e0156242.
- 438 53. Brachmann CB, et al. (1998) Designer deletion strains derived from *Saccharomyces cerevisiae*  
439 S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications.  
440 *Yeast* **14**:115–132.
- 441 54. Sylvester K, et al. (2015) Temperature and host preferences drive the diversification of  
442 *Saccharomyces* and other yeasts: a survey and the discovery of eight new yeast species. *FEMS Yeast Res*  
443 **15**:fov002
- 444 55. grofit: Fitting biological growth curves with R | Kahm | Journal of Statistical Software.  
445 doi:10.18637/jss.v033.i07
- 446 56. Wickham H (2009) *Ggplot2: Elegant Graphics for Data Analysis*. (Springer Publishing Company,  
447 Incorporated).

#### 448 **Figure Legends**

449 Figure 1 – The *PUL* gene cluster and its phylogenetic distribution. The genome-scale cladogram was  
450 constructed using alignments of 1,037 conserved BUSCO genes (22) with newly sequenced genomes in  
451 bold. Collapsed clades with multiple (*N*) species are shown as triangles, and the complete tree, with  
452 bootstrap values, is shown in SI Appendix, Fig. S1. Red filled boxes indicate pulcherrimin production has  
453 been observed in that species, and filled boxes for *PUL1-4* indicate presence or absence of the  
454 corresponding *PUL* gene. Gene arrow diagrams show the arrangement of the four *PUL* genes in  
455 representative species that produce pulcherrimin. The gene table contains putative annotations for the four  
456 genes prior to this study.

457 Figure 2 – Characterization of the *PUL* genes. A) Cells grown on defined medium produced a red  
458 pigment when the complete *PUL*cherrimin pathway was present, but they lacked pigmentation when key

459 genes were deleted. Complementation tests can be found in SI Appendix, Fig. S2. B) Representative  
460 growth curves of individual *pul* mutants in PA-treated medium, where cells must uptake pulcherrimin to  
461 grow. Complementation tests can be found in SI Appendix, Fig. S5B. C) Summary table of pulcherrimin  
462 production and growth in PA-treated medium. D) Model of the putative roles of the *PUL* genes in  
463 siderophore production and reuptake based on our genetic data and previous biochemistry data (18–20);  
464 however, the yeast proteins currently lack direct biochemical assays. Blue lines represent regulation of  
465 *PUL* gene expression by Pul4.

466 Figure 3 – A) Gene arrow diagram depicting the location and arrangement of the partial *PUL* gene cluster  
467 in *S. cerevisiae*. B) Representative growth curves of *pul* mutants of *S. cerevisiae* in PA-treated medium,  
468 where cells must uptake pulcherrimin to grow. C) PA-treated:untreated growth rate ratios for pairs of  
469 congeneric species containing or lacking *PUL3* and *PUL4* homologs. Boxplots contain the mean growth  
470 rate ratio from three biological replicates for each of the five species in each group. The ‘YES’ group is  
471 composed of species containing *PUL3* and *PUL4*, and the ‘NO’ group is comprised of closely related  
472 species lacking one or both genes. *p*-value is the result of a paired Student’s t-test.

473 Figure 4 – A) Repeated loss of the *PUL* cluster in Saccharomycotina evolution. Branch colors indicate the  
474 hypothesized presence or absence of each *PUL* gene for the lineage, reconstructed using Dollo parsimony  
475 (i.e. only allowing gene loss because there is no evidence of gene gain by HGT). Black names indicate the  
476 absence of *PUL* genes; blue names indicate the presence of *PUL3* and *PUL4* but not *PUL1* and *PUL2*;  
477 and red names indicate all four *PUL* genes are present. B) Model for the order of *PUL* gene loss and how  
478 resulting phenotypes interact relevant to the public goods problem.







