Giant *GAL* gene clusters for the melibiose-galactose pathway in *Torulaspora*

Anjan Venkatesh, Anthony L. Murray, Aisling Y. Coughlan, Kenneth H. Wolfe*

UCD Conway Institute and School of Medicine, University College Dublin, Dublin 4, Ireland

*To whom correspondence should be addressed. kenneth.wolfe@ucd.ie

Abstract

In many yeast species the three genes at the center of the galactose catabolism pathway, *GAL1*, *GAL10* and *GAL7*, are neighbors in the genome and form a metabolic gene cluster. We report here that some yeast strains in the genus *Torulaspora* have much larger *GAL* clusters that include genes for melibiase (*MEL1*), galactose permease (*GAL2*), glucose transporter (*HGT1*), phosphoglucomutase (*PGM1*), and the transcription factor *GAL4*, in addition to *GAL1*, *GAL10*, and *GAL7*. Together, these 8 genes encode almost all the steps in the pathway for catabolism of extracellular melibiose (a disaccharide of galactose and glucose). We show that a progenitor 5-gene cluster containing *GAL 7-1-10-4-2* was present in the common ancestor of *Torulaspora* and *Zygotorulaspora*. It added *PGM1* and *MEL1* in the ancestor of *Torulaspora*. It underwent further expansion in the *T. pretoriensis* clade, involving the fusion of three progenitor clusters in tandem and the gain of *HGT1*. These giant *GAL* clusters are highly polymorphic in structure, and subject to horizontal transfers, pseudogenization and gene losses. We identify recent horizontal transfers of complete *GAL* clusters from *T. franciscae* into one strain of *T. delbrueckii*, and from a relative of *T. maleeae* into one strain of *T. globosa.* The variability and dynamic evolution of *GAL* clusters in *Torulaspora* indicates that there is strong natural selection on the *GAL* pathway in this genus.

Introduction

Physical clusters of genes that function in the same process or metabolic pathway are relatively rare in yeasts (Riley et al., 2016; Rokas et al., 2018), but in budding yeasts (Saccharomycotina) the known examples include gene clusters for the pathways *NIT* (nitrate assimilation (Ávila et al., 2002)), *PUL* (pulcherrimin synthesis (Krause et al., 2018)), *NAG* (N-acetyl glucosamine catabolism (Yamada-Okabe et al., 2001)), *LAC* (lactose utilization (Varela et al., 2019)), *DAL* (allantoin degradation (Wong and Wolfe, 2005)), *MAL* (maltose utilization (Viigand et al., 2018)), and *GAL* (galactose utilization (Slot and Rokas, 2010)). The *GAL* pathway is one of the most intensively studied systems in yeast genetics. The canonical *GAL* gene cluster was first characterized in *Saccharomyces cerevisiae*, where it consists of three genes (*GAL1*, *GAL10* and *GAL7*) that code for the pathway to convert intracellular β-D-galactose to glucose-1-phosphate (Fig. 1) (Douglas and Hawthorne, 1964; St John and Davis, 1981). The same three genes are clustered in the same order in *Kluyveromyces lactis* (Webster and Dickson, 1988) and most other species in the family Saccharomycetaceae. A similar cluster of *GAL 1-10-7*, interspersed with two genes of unknown function, occurs in *Candida albicans* and other species in the CUG-Ser1 clade (Slot and Rokas, 2010). In more divergent yeasts the *GAL* genes are generally not clustered, except for four genera (*Schizosaccharomyces*, *Nadsonia*, *Brettanomyces* and *Wickerhamomyces*) that gained clusters by horizontal transfer from donors in the CUG-Ser1 clade, and two genera (*Cryptococcus* and *Lipomyces*) in which *GAL* clusters appear to have formed independently (Slot and Rokas, 2010; Haase et al., 2020).

It is widely thought that clustering of metabolic genes evolves as a mechanism for co-regulating the expression of genes, and that clustering can be selected for if an intermediate metabolite in the pathway is toxic – as is the case for galactose-1-phosphate in the *GAL* pathway – so that is it important to coordinate synthesis and removal of the toxin (McGary et al., 2013). The local order of genes within clusters often varies among species (Wong and Wolfe, 2005; Slot and Rokas, 2010; Naseeb and Delneri, 2012), and it is common to find that genes that are in a cluster in one species are completely absent from the genome in others (Hittinger et al., 2004; Wolfe et al., 2015). It is also common to find that the metabolic pathways encoded by clustered genes show presence/absence polymorphism within a

species: for example, the *GAL* genes (including the *GAL 1-10-7* cluster but also the unclustered genes *GAL4*, *GAL2* and *GAL80*) are intact in some populations of *S. kudriavzevii* but pseudogenes in others (Hittinger et al., 2010).

We previously reported that the genome sequence of the type strain of *Torulaspora delbrueckii* (CBS1146^{T}) contains a large cluster of *GAL* genes, occupying 22 kb near a telomere of chromosome 5 (Wolfe et al., 2015). As well as *GAL10* (2 copies), *GAL1* (2 copies) and *GAL7* (1 copy), the cluster also contained predicted genes *MEL1* (melibiase), *GAL2* (galactose permease), *PGM1* (phosphoglucomutase), *GAL4* (transcription factor) and *HGT1* (high-affinity glucose transporter, orthologous to *K. lactis HGT1* (Billard et al., 1996)). The genes in this cluster appeared to code for additional steps in the *GAL* pathway, both upstream and downstream of the steps encoded by the canonical *GAL1-10-7* cluster (Fig. 1). In the extended pathway, extracellular melibiose (a disaccharide) is hydrolyzed into its constituent monosaccharides β-D-galactose and D-glucose by secreted Mel1 enzyme (melibiase, an α(1,6)-galactosidase). The monosaccharides are then imported across the plasma membrane by Gal2 (for galactose) and Hgt1 (for glucose). The galactose is processed by the Gal10, Gal1 and Gal7 enzymes to yield glucose-1-phosphate, which is then converted to glucose-6-phosphate by Pgm1. A second molecule of glucose-6-phosphate is made by importing the glucose and phosphorylating it by hexokinase (Hxk1) or glucokinase (Glk1). The two molecules of glucose-6-phosphate then enter the glycolytic pathway. Thus, the *T. delbrueckii* gene cluster appeared to contain genes for all the steps needed to convert melibiose into two molecules of glucose-6-phosphate, except for hexokinase/glucokinase; there are *HXK1* and *GLK1* genes in the *T. delbrueckii* genome but they are not in the cluster. The *T. delbrueckii* cluster also contains an ortholog of *S. cerevisiae GAL4*, the transcription factor that positively regulates expression of the other *GAL* genes (Hittinger et al., 2004).

In this study, we used genome sequences from additional species and strains of *Torulaspora*, generated in other studies (Galeote et al., 2018; Shen et al., 2018; Coughlan et al., 2020), to investigate the origin and evolution of *GAL* clusters in *Torulaspora* and related genera. We find that the large *GAL* cluster in the type strain of *T. delbrueckii* is atypical of this species, because all 14 other *T. delbrueckii* strains that we examined have no cluster, and we show that the cluster in the type strain of *T. delbrueckii* was acquired from

95 *T. franciscae* recently by horizontal gene transfer. We also uncovered an extraordinary
96 diversity of allelic *GAL* gene cluster structures in *T. pretoriensis*, and a rich history of cluster
97 expansion, fusion, and degeneration.
98
99
100 **Results**
101
102 **Phylogeny and phenotypes**
103
104 We examined genome sequences from multiple strains of *T. delbrueckii*, *T. pretoriensis* and
105 *T. globosa*, and from single strains of other *Torulaspora* species, as well as outgroups
106 *Zygotorulaspora mrakii*, *Zygosaccharomyces rouxii*, *Kluyveromyces lactis* and *S. cerevisiae*.
107 The phylogeny of the species, and a summary of the major events we infer to have occurred
108 during *GAL* cluster evolution in *Torulaspora*, is shown in Figure 2. One gene in the well-
109 known *GAL* system of *S. cerevisiae*, *GAL3*, is a paralog of *GAL1* that was formed by the
110 whole-genome duplication (WGD). *Torulaspora* and all the other genera considered here
111 diverged from *S. cerevisiae* before the WGD occurred, so their *GAL1* genes are orthologous
112 to both *GAL1* and *GAL3* in *S. cerevisiae*. Another gene, *GAL80*, coding for a corepressor of
113 *GAL* gene expression, is absent from most *Torulaspora* species (Fig. 2).
114
115 A *GAL* cluster is present in at least some strains of all the *Torulaspora* species we studied.
116 We tested the ability of several strains to grow on solid media containing galactose,
117 melibiose, or glucose as a sole carbon source (Fig. 3). We found that the ability to grow on
118 galactose correlates with the presence of intact copies of the genes *GAL1*, *GAL10* and *GAL7*
119 in the genome, and the ability to grow on melibiose correlates with the presence of an
120 intact *MEL1* gene (Fig. 3). The starting point for our study was the large *GAL* cluster on
121 chromosome 5 of *T. delbrueckii* strain CBS1146^{T} (Wolfe et al., 2015), and we found that this
122 strain can grow on galactose whereas *T. delbrueckii* strain L09, which lacks the cluster,
123 cannot (Fig. 3). However, we were surprised to find that *T. delbrueckii* CBS1146^{T} cannot
124 grow on melibiose despite apparently having a *MEL1* gene. We realized that the open
125 reading frame we originally annotated as *MEL1* (*TDEL0E00170*) is truncated at the 5' end
126 relative to other *MEL1* genes. Comparison to a functional *MEL1* gene previously

127 characterized by Oda and Fukunaga (1999) from *T. delbrueckii* strain IFO1255 shows that
128 CBS1146^{T} has a TGG (Trp) -> TGA (stop) mutation at codon 38 which removes the region
129 coding for the secretion signal, so the *MEL1* gene of CBS1146^{T} is a pseudogene. A second
130 discrepancy between genotypes and phenotypes occurs in *T. pretoriensis* CBS2187^{T}, which
131 grows poorly on galactose despite containing *GAL1*, *GAL10* and *GAL7* genes (Fig. 3). This
132 discrepancy is discussed later.
133
134
135 **Synteny relationships**
136
137 Synteny comparisons among the *Torulaspora* species and outgroups revealed a complex
138 pattern of relationships and gene relocations (Fig. 2). For some loci, we refer to the
139 Ancestral gene numbering system of Gordon et al. (2009), which numbers genes
140 sequentially along the 8 chromosomes inferred to have existed just prior to the WGD, for
141 example locus Anc_8.123 is the 123rd gene along Ancestral chromosome 8. This numbering
142 system is also used in our Yeast Gene Order Browser (ygob.ucd.ie) (Byrne and Wolfe, 2005).
143
144 In the outgroup species shown at the bottom of Figure 2 (*S. cerevisiae*, *K. lactis*, *Z. rouxii*),
145 the only genes in the *GAL* pathway that are clustered are *GAL1*, *GAL10* and *GAL7*, and they
146 occur in the order *GAL 1-10-7*. This arrangement is conserved in *T. microellipsoides*,
147 including the flanking genes *SNQ2* and *RPT2* (Anc_3.216 to Anc_3.220). This cluster is at an
148 internal chromosomal site in these species, i.e. it is not subtelomeric. In the outgroups, the
149 other genes in the pathway are at conserved, dispersed, places in the genome (*PGM1* =
150 Anc_2.445; *GAL4* = Anc_6.279; *HGT1* = Anc_1.432; *GAL80* = Anc_1.500), and *MEL1* is not
151 present at all.
152
153
154 **Formation of a gene cluster for the complete *MEL-GAL-PGM* pathway in the common**
155 **ancestor of *Torulaspora* and *Zygotorulaspora***
156
157 In *Zygotorulaspora mrakii*, the cluster has expanded to 6 genes: it contains *GAL 7-1-10-4-2*
158 and a *PGM1* gene (Fig. 2). *Z. mrakii* also has an unlinked *MEL1* gene, which was previously

159 shown to be functional by Oda and Fujisawa (2000). The 6-gene cluster has gained genes for
160 the pathway steps upstream (*GAL2*) and downstream (*PGM1*) of the steps encoded by the
161 3-gene cluster, as well as gaining the transcription factor *GAL4*. It is interesting that the
162 order of the 3 genes has also changed, from *GAL 1-10-7* in the outgroups to *GAL 7-1-10* in
163 *Z. mrakii*. The *Z. mrakii* 6-gene cluster is located at an internal chromosomal site between
164 *EST3* (Anc_7.128) and *URM1* (Anc_7.129). The cluster therefore appears to have become
165 inserted between two genes that were ancestrally neighbors.
166
167 In *T. maleeae*, there is a 7-gene cluster with identical gene order to the 6-gene cluster of
168 *Z. mrakii*, plus *MEL1* (Fig. 2). This cluster appears to be at a subtelomeric location, and the
169 *EST3* and *URM1* genes (Anc_7.128/7.129) are adjacent in this species. Both *T. maleeae* and
170 *Z. mrakii* have two *PGM1* genes. The first, designated *PGM1_anc*, is at the ancestral *PGM1*
171 location (Anc_2.445). It is syntenic with the *PGM1* genes of other yeasts, including the
172 *PGM1/PGM2* gene pair of *S. cerevisiae*, which is a WGD pair. The second, designated
173 *PGM1_dup*, is a duplicated copy of *PGM1* located in the *GAL* cluster.
174
175 The gene order *GAL 7-1-10-4-2*, as seen in *Z. mrakii* and *T. maleeae*, is a pattern that recurs
176 throughout the *GAL* clusters of most *Torulaspora* species that will be described in the
177 following sections. However, *T. microellipsoides* has an ancestral-type cluster (*GAL 1-10-7*)
178 at the ancestral location (Anc_3.219), rather than the *GAL 7-1-10-4-2* pattern, even though
179 phylogenomic analysis (Shen et al., 2018) has indicated that the genus *Torulaspora* is
180 monophyletic and *Zygotorulaspora* is an outgroup to it. *T. microellipsoides* also has a *MEL1*
181 gene at an unlinked, non-telomeric location (Fig. 2).
182
183 The organization of *GAL* genes in *T. microellipsoides* resembles the outgroup species more
184 closely than it resembles other *Torulaspora* species, whereas the *Z. mrakii* organization
185 resembles *Torulaspora* species (Fig. 2). In phylogenetic trees of individual *GAL* genes,
186 *T. microellipsoides* is often placed outside *Z. mrakii* (Figure S1), in contrast to the
187 phylogenomic tree. Moreover, *GAL80* is present in *T. microellipsoides* but absent in *Z. mrakii*
188 and the other *Torulaspora* species (Fig. 2). Together, these results suggest that either the
189 phylogenomic tree is incorrect, or there has been horizontal transfer of a *GAL* cluster

190 between the *Z. mrakii* and *T. maleeae* branches, in either direction (Fig. 2), making the *GAL*
191 phylogeny different from the phylogeny of the rest of the genome.
192
193 In summary, the point of origin of the *GAL 7-1-10-4-2* cluster pattern is not fully clear, but it
194 appears to have been present in the common ancestor of the genera *Zygotorulaspora* and
195 *Torulaspora*. It is first seen with *PGM1* at one end, and later gained *MEL1* at the other end.
196
197
198 **Horizontal *GAL* cluster transfer into one strain of *T. globosa***
199
200 *T. globosa* is a sister species to *T. maleeae*. We sequenced the genomes of 12 strains of
201 *T. globosa* (Coughlan et al., 2020 and A.Y.C. and K.H.W., unpublished) and found that 11 of
202 them, including the type strain CBS764^T, have no *GAL* genes. However, one strain, *T. globosa*
203 NRRL YB-1481, has a *GAL* cluster, and the organization of this cluster is very similar to the
204 *T. maleeae* cluster (Fig. 2). Phylogenetic trees of *GAL 7*, *1*, *10*, *4*, *2* and *MEL1* all show that
205 the *T. globosa* NRRL YB-1481 genes group with the *T. maleeae* genes (Fig. S1). In plate tests,
206 *T. globosa* NRRL YB-1481 was able to grow on melibiose and galactose, whereas *T. globosa*
207 CBS764^T could not (Fig. 3).
208
209 Interestingly, the *GAL* cluster in *T. globosa* strain NRRL YB-1481 has formed at the ancestral
210 location of *PGM1* (Anc_2.445; Fig. 2). This strain has only one *PGM1* gene, in contrast to
211 *T. maleeae* and *Z. mrakii* which have two (*PGM1_anc* and *PGM1_dup*). Since most
212 *T. globosa* strains have no *GAL* genes, the most plausible scenario to explain the presence of
213 a cluster in NRRL YB-1481 is that it originated by horizontal transfer. In view of the relatively
214 low DNA sequence identity (74%) between the *T. globosa* NRRL YB-1481 and *T. maleeae*
215 clusters, the donor is more likely to have been an unidentified species related to
216 *T. maleeae*/*T. globosa*, rather than *T. maleeae* itself.
217
218 Although it is possible that recombination between the *PGM1* genes in the donor cluster
219 and the recipient *T. globosa* NRRL YB-1481 genome might have guided integration of the
220 cluster, this seems unlikely because the *T. maleeae* and *T. globosa PGM1* genes are
221 currently in opposite orientations relative to their neighbor *GAL7* (Fig. 2). Also, a

222 phylogenetic tree of *PGM1* sequences (Fig. 4) places the single, cluster-associated, *PGM1* of
223 *T. globosa* NRRL YB-1481 at the position expected for *PGM1_anc* gene: it is in a clade with
224 the single *PGM1* gene of *T. globosa* CBS764^T and *T. maleeae PGM1_anc*, and far away from
225 *T. maleeae PGM1_dup* which lies in a clade with *Z. mrakii PGM1_dup.*
226
227
228 **Horizontal *GAL* cluster transfer from *T. franciscae* into *T. delbrueckii***
229
230 *T. pretoriensis*, *T. franciscae* and *T. delbrueckii* form a clade of three species whose *GAL*
231 clusters, when present, are greatly expanded and contain numerous *GAL* pseudogenes as
232 well as functional genes. We analyzed data from multiple strains of *T. delbrueckii* and
233 *T. pretoriensis*, but we have only one genome sequence from *T. franciscae* (the type strain,
234 CBS2926^T).
235
236 In the set of 15 *T. delbrueckii* strains that we analyzed, none except CBS1146^T contains a
237 *GAL* cluster, which suggests that the cluster was gained by horizontal transfer. The CBS1146^T
238 cluster is identical in gene organization to a cluster in the type strain of *T. franciscae*, and
239 the two clusters have 97% DNA sequence identity over 22 kb. The similarity between these
240 two species is much higher than between either of them and *T. pretoriensis*, even though
241 *T. pretoriensis* is a sister species to *T. franciscae* (Fig. 2). We therefore infer that horizontal
242 transfer occurred from *T. franciscae* to *T. delbrueckii*. Curiously, although the cluster is near
243 a telomere in both species, the two species have opposite orientations of the cluster
244 relative to the telomere (Fig. 2).
245
246 The *MEL1* genes in the clusters in the type strains of both *T. franciscae* and *T. delbrueckii* are
247 pseudogenes, and these strains are unable to grow on melibiose but able to grow on
248 galactose (Fig. 3). In a previous study by Oda and Tonomura (1996), 12 of 28 *T. delbrueckii*
249 strains examined, including the type strain, were found to be able to grow on galactose.
250 Only one of the *T. delbrueckii* strains (IFO 1255) could grow on melibiose as well as
251 galactose and was shown to have an intact *MEL1* gene (Oda and Tonomura, 1996; Oda and
252 Fukunaga, 1999).
253

254

255 **Extensive structural polymorphism of *T. pretoriensis GAL* clusters**

256

257 We analyzed genome sequences from nine strains of *T. pretoriensis*, of which five have large
258 and variable *GAL* clusters, and the other four have none. The four strains without clusters
259 (CBS11100, CBS11121, CBS11123, CBS11124) are closely related to each other, so only
260 CBS11100 is shown in Figure 2. Among the five strains with clusters, there is extensive
261 structural polymorphism, with only two strains (CBS2187^T and CBS9333) having similar
262 organization. All the *GAL* clusters in *T. pretoriensis* strains appear to be near telomeres.

263

264 The most complex *GAL* cluster in *T. pretoriensis* is in strain UWOPS 83-1046.2 (Fig. 2; we
265 refer to this strain hereafter as UWOPS). It spans 42 kb and contains 8 intact genes and 8
266 pseudogenes related to galactose metabolism. It also contains 2 unrelated genes and 1
267 unrelated pseudogene, which appear to be of subtelomeric origin. These unrelated genes
268 occupy a region of 15 kb inside the cluster and divide it into two parts, left and right. The
269 right part is almost identical in gene organization to the large *GAL* cluster that was
270 transferred between *T. franciscae* and *T. delbrueckii* CBS1146^T, the only differences being
271 some genes that are pseudogenes in *T. pretoriensis* UWOPS but intact in *T. franciscae* and
272 *T. delbrueckii* CBS1146^T, or vice versa (*HGT1*, *MEL1*, and one copy each of *GAL1* and *GAL10*;
273 Fig. 2). Phylogenetic analysis of the genes in this region (Fig. 4; Fig. S1) shows that, in all
274 cases, *T. franciscae* and *T. delbrueckii* CBS1146^T form a clade with *T. pretoriensis* UWOPS
275 outside, which contradicts the expected species phylogeny (Fig. 2) and supports the
276 hypothesis of horizontal transfer between *T. franciscae* and *T. delbruckii*.

277

278 We tested the phenotypes of four *T. pretoriensis* strains (Fig. 3). As expected, only UWOPS
279 can grow on melibiose – it is the only strain with intact *MEL1*. On galactose, CBS11100
280 cannot grow (it has no *GAL* cluster), CBS5080 and UWOPS grow well, and the type strain
281 CBS2187^T grows more slowly. The poor growth of the type strain of *T. pretoriensis* on
282 galactose is consistent with previous studies. Oda and colleagues reported that
283 fermentation of galactose or melibiose by strain YK-1, which is a non-sedimenting derivative
284 of *T. pretoriensis* CBS2187^T (syn. IFO 10218), was undetectable after 2 days, whereas
285 *T. pretoriensis* CBS5080 (IFO 0022) and *T. franciscae* CBS2926^T (IFO 1360) fermented

286 galactose but not melibiose (Oda and Tonomura, 1993; Oda and Tonomura, 1996). Oda's
287 results are consistent with our results in Figure 3, except that we find that growth of
288 CBS2187^{T} on galactose is slow rather than absent. A possible reason for the poor growth is
289 that there is no *GAL2* galactose transporter gene anywhere in the *T. pretoriensis* CBS2187^{T}
290 genome; it is the only strain in Figure 3 that has the *GAL* enzyme genes without the
291 transporter gene.
292
293
294 **Cluster expansion by tandem triplication of progenitor *GAL 7-1-10-4-2* clusters**
295
296 Closer examination of the *T. pretoriensis GAL* clusters shows that they have an internal
297 structure that is based on tandem triplication of the *GAL 7-1-10-4-2* pattern mentioned
298 earlier. This structure is most clearly seen in *T. pretoriensis* UWOPS which has three copies
299 of the pattern: including pseudogenes, it has *GAL 7-1-10-4-2* in the left part of the cluster,
300 and *GAL 7-1-10-4* (without *GAL2*) followed by *GAL 7-1-10-2* (without *GAL4*) in the right part.
301 The other genes in the cluster (*HGT1*, *MEL1*, *PGM1*, and the unrelated genes between the
302 left and right parts) are located at the junctions between these three copies of the pattern.
303
304 This arrangement suggests that the large UWOPS cluster was formed by tandem fusion of
305 three smaller progenitor clusters that we designate L, R1 and R2, corresponding to the left
306 part and two sections of the right part of the current cluster (Fig. 2). We postulate that L
307 contained *GAL 7-1-10-4-2*, R1 originally contained *HGT1 – GAL 7-1-10-4-2*, and R2 originally
308 contained *MEL1 – GAL 7-1-10-4-2 – PGM1*. Subsequently, many of the triplicated *GAL* gene
309 copies became pseudogenes or relics (very short pseudogenes), and no trace remains of
310 *GAL2* in R1 or *GAL4* in R2. Notably, although there are many pseudogenes in the
311 *T. pretoriensis* clusters (of all strains), there are no pseudogenes that indicate that *HGT1*,
312 *MEL1*, or *PGM1* was ever duplicated within the clusters; all the duplications are of *GAL*
313 genes. Therefore we suggest that the triple-size cluster did not arise by triplicating a single
314 progenitor cluster, but instead arose by fusion of three progenitor clusters that were similar
315 (containing *GAL 7-1-10-4-2*) but already different regarding their content of *HGT1*, *MEL1* and
316 *PGM1*.
317

318 The clusters in the other *T. pretoriensis* strains are smaller than in UWOPS but still
319 consistent with the hypothesis of cluster expansion by tandem fusion of progenitors. Strain
320 CBS2785 has an overall organization similar to UWOPS, but it has lost *MEL1* and adjacent
321 parts of R1 and R2. It has also sustained an inversion of *GAL1-10-4* in the L part, probably in
322 conjunction with the formation of an extra relic of *GAL7* that is also in inverted orientation.
323 Strain CBS5080 has parts L and R2 but not R1, and it also has additional *HGT1* and *GAL1*
324 genes to the right of R2. Strains CBS2187^{T} and CBS9333 have only part L and an additional
325 *GAL1* gene; they lack *MEL1*, *HGT1* and *PGM1* in the cluster and have only one *PGM1* gene in
326 their genomes (at the ancestral locus Anc_2.445). The phylogenies of most genes and
327 pseudogenes in the *T. pretoriensis* clusters generally support the relationships shown in
328 Figure 2, which are based on synteny as well as phylogenetic considerations (Fig. S1). It is
329 impossible to infer the complete history of the *T. pretoriensis* clusters, but we can conclude
330 that (i) at least three progenitor clusters fused in tandem to form them, and (ii) they are
331 undergoing extensive within-species structural rearrangement and turnover.
332
333
334 **Vestigial *GAL* clusters and extra unclustered *GAL10* and *HGT1* genes**
335
336 The large *GAL* cluster in *T. delbrueckii* originated by horizontal transfer from *T. franciscae*.
337 Among our sequenced strains, it is only present in CBS1146^{T} and is located near a telomere
338 of chromosome 5. However, in addition, all 15 *T. delbrueckii* strains (including CBS1146^{T})
339 also contain an intact *GAL10* gene near a telomere of chromosome 7 (Fig. 2). It is located
340 beside four pseudogenes in the arrangement *HGT1 – GAL 7-1-10-4*, where *GAL10* is the only
341 intact gene, so it appears to be a remnant of a primordial *GAL* cluster that has almost
342 disappeared. Its structure is the same as the R1 primordial cluster inferred in *T. pretoriensis*.
343
344 Similarly, most strains of *T. pretoriensis* have an extra copy of *GAL10*, located near *HGT1* and
345 a telomere (Fig. 2). This *GAL10* gene is present even in strains such as CBS11100 that cannot
346 utilize galactose. Therefore, many strains of both *T. delbrueckii* and *T. pretoriensis* contain
347 *GAL10* but no other *GAL* genes. This situation has also been seen in other yeasts (Haase et
348 al., 2020) but its physiological significance is unknown.
349

350 An extra vestigial telomeric *GAL* cluster is also seen in *T. maleeae*, containing an intact *HGT1*
351 gene and pseudogenes of *GAL7* and *GAL1* (Fig. 2). Thus, in both *T. maleeae* and
352 *T. pretoriensis*, high-affinity glucose transporter function is provided by an *HGT1* gene that is
353 neither located at the ancestral *HGT1* locus (Anc_1.432), nor in an active *GAL* cluster
354 containing intact *GAL1* and *GAL7*, but in a remnant of a degraded cluster at a telomeric
355 location that sometimes also includes *GAL10*. Notably, in the only *T. pretoriensis* strain that
356 includes an intact *HGT1* in its *GAL* cluster (CBS5080), there are no additional telomeric *HGT1*
357 or *GAL10* genes (Fig 2).
358
359
360 **Discussion**
361
362 The *GAL* clusters of *Torulaspora* species are remarkably large and heterogeneous. There are
363 polymorphisms both for presence/absence of the cluster, and for gene order within the
364 cluster. Formation of pseudogenes is common. As a result, *Torulaspora* strains vary in their
365 ability to grow using galactose or melibiose as the sole carbon source.
366
367 The cluster first expanded from a canonical 3-gene *GAL 1-10-7* structure by adding *GAL2* and
368 *GAL4*, around the time of the common ancestor of *Torulaspora* and *Zygotorulaspora*. The
369 synteny relationships in Figure 2 suggest that a duplicate copy of *PGM1* was then recruited
370 into the *GAL 7-1-10-4-2* cluster, followed later by relocation of *MEL1* and then *HGT1*.
371 However, the phylogeny of *PGM1* sequences (Fig. 4) shows that there must have been
372 multiple separate incorporations of *PGM1* into the cluster, because the *PGM1_dup* genes in
373 the giant *GAL* clusters of the *T. pretoriensis/T. delbrueckii*/*T. franciscae* clade originated
374 independently of the *PGM1_dup* genes in the smaller clusters of *T. maleeae* and *Z. mrakii*.
375 Including the integration of a *GAL* cluster beside *PGM1_anc* in *T. globosa* NRRL Y-1481,
376 there were three separate, parallel, events of incorporation of *PGM1* into *Torulaspora GAL*
377 clusters – pointing to strong selection to incorporate it. In two *Lachancea* species a *GAL*
378 cluster including *GAL1, GAL7* and *GAL2* has formed beside *PGM1* at its ancestral location
379 (Kuang et al., 2018), similar to what we observe in *T. globosa* NRRL Y-1481. *PGM1* is a
380 bottleneck gene, coding for an enzyme that integrates metabolic flux from several pathways
381 including glycogen synthesis, trehalose synthesis and the pentose phosphate pathway as

382 well as the *GAL* pathway, and in the genera *Saccharomyces* and *Lachancea*, regulation of
383 *PGM1* by *GAL4* has been gained and lost multiple times (Kuang et al., 2018). We find that in
384 the species with two *PGM1* genes (Fig. 4), the *PGM1_dup* genes in the cluster contain
385 multiple putative Gal4 binding sites (CGG-N_{11}-CCG) in their upstream regions, whereas the
386 *PGM1_anc* genes do not. In *T. globosa* NRRL YB-1481, *PGM1* is not duplicated but has Gal4
387 sites in the upstream region that it shares with *GAL7* (Fig. 2). Thus, in all the clusters in the
388 *Torulaspora* clade, a *PGM1* gene has come under the regulation of *GAL4.*
389
390 Unexpectedly, our results indicate that duplication and fusion of whole clusters, rather than
391 duplication of individual genes, was the major mechanism of evolution of *GAL* clusters. In
392 *T. pretoriensis*, three primordial clusters fused to form one giant cluster and many of the
393 genes later became pseudogenes. Tandem fusion of clusters may have provided an
394 opportunity to experiment with shuffling the gene order, by allowing different gene copies
395 to become pseudogenes. For example, in the *T. pretoriensis* clusters, the intact gene
396 upstream of *GAL1* can be *GAL10, GAL2*, *GAL4*, or *MEL1* (Fig. 2). Haase et al. (2020) recently
397 identified a similar fusion of two *GAL* clusters (one ancestral and one horizontally
398 transferred) in *Nadsonia fulvescens*.
399
400 The *Torulaspora GAL* clusters include up to eight different functional genes, comprising the
401 whole *MEL-GAL-PGM* pathway except for hexokinase/glucokinase (Fig. 1). Since the sugar
402 kinases also function in the pathway for catabolism of glucose monomers imported into the
403 cell by hexose transporters, the eight genes in the cluster constitute the complete set of
404 genes that need to be activated in the presence of melibiose or galactose, and repressed in
405 their absence. In *K. lactis*, *HGT1* was originally described as a high-affinity glucose
406 transporter, but it can also transport galactose and is induced by galactose (Baruffini et al.,
407 2006).
408
409 To build clusters with eight functional genes by random genomic rearrangements, natural
410 selection on the *GAL* metabolic pathway must be exceptionally strong in *Torulaspora*.
411 However, we have no explanation for why selection to form clusters is stronger in
412 *Torulaspora* than in other budding yeast genera. It seems likely that regulatory changes,
413 involving duplication of *PGM1*, loss of *GAL80*, and movement of *GAL4* into the cluster were

414 central to expansion of the cluster. Previous work has shown that Gal4 became the major
415 regulator of the *GAL* pathway relatively recently, displacing Rtg1/Rtg3 in an ancestor of the
416 family Saccharomycetaceae (Choudhury and Whiteway, 2018; Haase et al., 2020). In the
417 *Torulaspora/Zygotorulaspora* clade, the further step of moving the *GAL4* gene into the
418 cluster has occurred. Relocation of *GAL4* into the cluster would have enabled the Gal4
419 protein to evolve in concert with its binding sites in the promoters of the nearby *GAL* genes.
420 Moreover, in the *Torulaspora/Zygotorulaspora* species (except *T. microellipsoides*), Gal4 has
421 lost the C-terminal domain for interaction with the co-repressor Gal80 (Choudhury and
422 Whiteway, 2018), and the *GAL80* gene is absent from their genomes (Fig. 2). In each cluster,
423 multiple putative Gal4 binding sites are present upstream of each intact *GAL* gene (except
424 *GAL4*) as well as *PGM1* and *HGT1*, but not *MEL1*. These regulatory changes may have made
425 the cluster almost independent of other loci in the genome, and hence made it more
426 amenable to transfer among species.
427
428
429 **Methods**
430
431 Yeast strains were obtained from the Westerdijk Fungal Biodiversity Institute (CBS strains),
432 the USDA Agricultural Research Service (NRRL strains), Lallemand Inc. (L09), and M.-A.
433 Lachance (UWOPS 83-1046.2).
434
435 For growth tests, yeast strains were streaked onto agar plates made with YPD (2% dextrose)
436 (Formedium, catalog CCM0110), YNB (yeast nitrogen base; Sigma-Aldrich, 51483) with 2%
437 D-(+)-galactose (Sigma-Aldrich, G0625), or YNB with 2% D-(+)-melibiose (Sigma-Aldrich,
438 63630). Plates were incubated at 30° C for 48 hours before photographing.
439
440 For sequencing *T. globosa* strain NRRL YB-1481, cultures were grown under standard rich-
441 medium conditions. DNA was harvested from stationary-phase cultures by homogenization
442 with glass beads followed by phenol-chloroform extraction and ethanol precipitation.
443 Purified DNA was concentrated with the Genomic DNA Clean and Concentrator-10 (Zymo
444 Research, catalog D4010). Sequencing was done by BGI Tech Solutions (Hong Kong) using
445 Illumina HiSeq 4000 (paired end, 2 x 150 bp reads), and assembled using SPAdes version

446 3.11.1 (Bankevich et al., 2012). Coverage was approximately 85x. All other genome
447 sequences are from sources cited in Coughlan et al. (2020).
448
449 *GAL* clusters were annotated manually. In the *T. franciscae* genome assembly, the large
450 cluster was initially split into three contigs due to high similarity between the two *GAL10*
451 genes. Its organization was inferred by manually merging scaffold 86, scaffold 87, and contig
452 C4393.
453
454 Genes were inferred to be located in subtelomeric regions if the gene is near the end of a
455 chromosome-sized scaffold, or if DNA sequences neighboring the gene are repeat
456 sequences that occur only near the ends of multiple very large scaffolds, or if several
457 neighbors of the gene are members of gene families that are often found in subtelomeric
458 regions (Brown et al., 2010) and do not have Ancestral gene numbers (Gordon et al., 2009).
459
460 Phylogenetic trees were constructed from MUSCLE alignments of amino acid sequences,
461 using PhyML as implemented in version 5.0 of SeaView (Gouy et al., 2010). Approximate
462 translations of pseudogenes were made by manual annotation.
463
464
465

466 **Figure Legends**

467

468 **Figure 1.** The yeast biochemical pathway for catabolism of extracellular melibiose (Holden
469 et al., 2003). Colored backgrounds indicate genes that are located in clusters in *Torulaspora*
470 species. Gal10 has two distinct functions, mutarotase and epimerase, performed by two
471 domains of the protein. Hgt1 has been reported to transport galactose as well as glucose in
472 *K. lactis* (Baruffini et al., 2006).

473

474 **Figure 2.** Synteny relationships among *GAL* genes and clusters in *Torulaspora* species and
475 outgroups. Genes are labeled with their *GAL* gene number (7, 1, 10, 4, 2, or 80), or M
476 (*MEL1*), P (*PGM1*), or H (*HGT1*). Dashed borders on gene symbols indicate pseudogenes.
477 Gray backgrounds highlight groups of adjacent genes with the progenitor cluster gene order
478 *GAL 7-1-10-4-2* or subsets thereof. Large gray boxes indicate groups of genes that are at
479 syntenic locations in different strains/species, and are indicated as being either telomeric or
480 internal to chromosomes. Ancestral gene locations refer to the numbering system of
481 Gordon et al. (2009) and are internal to chromosomes. Different P symbols are used to
482 distinguish between *PGM1* genes at the ancestral location (*PGM1_anc*, dark brown), and
483 duplicate *PGM1* genes in *GAL* clusters (*PGM1_dup*, light brown). Tel indicates a region
484 inferred to be close to a telomere (subtelomeric), and zigzag symbols in *T. pretoriensis*
485 indicate intervening regions of 10-15 kb with no genes related to *GAL* metabolism. The tree
486 topology is from the phylogenomic analysis of Shen et al. (2018) with *T. globosa* added as in
487 (Saluja et al., 2012; Kaewwichian et al., 2020).

488

489 **Figure 3.** Growth of *Torulaspora* strains on galactose, melibiose, and glucose (YPD) media.
490 Plates were incubated at 30° C for 48 hours before photographing. The lower panel
491 indicates the presence or absence of intact genes in each genome.

492

493 **Figure 4.** Phylogenetic tree of *PGM1* (phosphoglucomutase) genes. For species/strains with
494 two *PGM1* genes, the copies are labeled *PGM1_anc* and *PGM1_dup*; other species/strains
495 have only one gene. Asterisks beside gene names indicate *PGM1* genes that are located in
496 *GAL* clusters, and asterisks on the tree indicate inferred points of incorporation of *PGM1*

genes into *GAL* clusters. Approximate likelihood ratio test (aLRT) branch support values are shown.

Figure S1. Phylogenetic trees constructed from (a) *GAL7*, (b) *GAL1*, (c) *GAL10*, (d) *GAL4*, (e) *GAL2*, (f) *HGT1*, and (g) *MEL1* sequences. Approximate likelihood ratio test (aLRT) branch support values are shown.

Acknowledgments

This work was supported by the European Research Council (789341) and Science Foundation Ireland (13/IA/1910).

511 **References**

512

513 Ávila, J., González, C., Brito, N., Machín, M.F., Pérez, D., and Siverio, J.M. (2002). A second
514 Zn(II)(2)Cys(6) transcriptional factor encoded by the *YNA2* gene is indispensable for
515 the transcriptional activation of the genes involved in nitrate assimilation in the
516 yeast *Hansenula polymorpha*. Yeast *19*, 537-544.

517 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M.,
518 Nikolenko, S.I., Pham, S., Prjibelski, A.D.*, et al.* (2012). SPAdes: a new genome
519 assembly algorithm and its applications to single-cell sequencing. J Comput Biol *19*,
520 455-477.

521 Baruffini, E., Goffrini, P., Donnini, C., and Lodi, T. (2006). Galactose transport in
522 *Kluyveromyces lactis*: major role of the glucose permease Hgt1. FEMS Yeast Res *6*,
523 1235-1242.

524 Billard, P., Menart, S., Blaisonneau, J., Bolotin-Fukuhara, M., Fukuhara, H., and Wesolowski-
525 Louvel, M. (1996). Glucose uptake in *Kluyveromyces lactis*: role of the *HGT1* gene in
526 glucose transport. J Bacteriol *178*, 5860-5866.

527 Brown, C.A., Murray, A.W., and Verstrepen, K.J. (2010). Rapid Expansion and Functional
528 Divergence of Subtelomeric Gene Families in Yeasts. Curr Biol *20*, 895-903.

529 Byrne, K.P., and Wolfe, K.H. (2005). The Yeast Gene Order Browser: combining curated
530 homology and syntenic context reveals gene fate in polyploid species. Genome Res
531 *15*, 1456-1461.

532 Choudhury, B.I., and Whiteway, M. (2018). Evolutionary Transition of *GAL* Regulatory Circuit
533 from Generalist to Specialist Function in Ascomycetes. Trends Microbiol *26*, 692-702.

534 Coughlan, A.Y., Lombardi, L., Braun-Galleani, S., Martos, A.A., Galeote, V., Bigey, F., Dequin,
535 S., Byrne, K.P., and Wolfe, K.H. (2020). The yeast mating-type switching
536 endonuclease HO is a domesticated member of an unorthodox homing genetic
537 element family. Elife *9*, e55336.

538 Douglas, H.C., and Hawthorne, D.C. (1964). Enzymatic Expression and Genetic Linkage of
539 Genes Controlling Galactose Utilization in Saccharomyces. Genetics *49*, 837-844.

540 Galeote, V., Bigey, F., Devillers, H., Ortiz-Merino, R.A., Dequin, S., Wolfe, K.H., and
541 Neuveglise, C. (2018). Genome Sequence of *Torulaspora microellipsoides* CLIB
542 830(T). Genome Announc *6*, e00615-00618.

543 Gordon, J.L., Byrne, K.P., and Wolfe, K.H. (2009). Additions, losses, and rearrangements on
544 the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces*
545 *cerevisiae* genome. PLoS Genet *5*, e1000485.

546 Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical
547 user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol
548 *27*, 221-224.

549 Haase, M.A.B., Kominek, J., Opulente, D.A., Shen, X.-X., LaBella, A.L., Zhou, X., DeVirgilio, J.,
550 Hulfachor, A.B., Kurtzman, C.P., Rokas, A.*, et al.* (2020). Repeated horizontal gene
551 transfer of GALactose metabolism genes violates Dollo's law of irreversible loss.
552 BioRxiv.

553 Hittinger, C.T., Goncalves, P., Sampaio, J.P., Dover, J., Johnston, M., and Rokas, A. (2010).
554 Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature
555 *464*, 54-58.

556 Hittinger, C.T., Rokas, A., and Carroll, S.B. (2004). Parallel inactivation of multiple *GAL*
557 pathway genes and ecological diversification in yeasts. Proc Natl Acad Sci USA *101*,
558 14144-14149.

559 Holden, H.M., Rayment, I., and Thoden, J.B. (2003). Structure and function of enzymes of
560 the Leloir pathway for galactose metabolism. J Biol Chem *278*, 43885-43888.

561 Kaewwichian, R., Khunnamwong, P., Am-In, S., Jindamorakot, S., and Limtong, S. (2020).
562 *Torulaspora nypae sp. nov.*, a novel yeast species isolated from nipa (*Nypa fruticans*
563 Wurmb.) inflorescence sap in southern Thailand. Int J Syst Evol Microbiol *70*, 1112-
564 1116.

565 Krause, D.J., Kominek, J., Opulente, D.A., Shen, X.X., Zhou, X., Langdon, Q.K., DeVirgilio, J.,
566 Hulfachor, A.B., Kurtzman, C.P., Rokas, A.*, et al.* (2018). Functional and evolutionary
567 characterization of a secondary metabolite gene cluster in budding yeasts. Proc Natl
568 Acad Sci USA *115*, 11030-11035.

569 Kuang, M.C., Kominek, J., Alexander, W.G., Cheng, J.F., Wrobel, R.L., and Hittinger, C.T.
570 (2018). Repeated Cis-Regulatory Tuning of a Metabolic Bottleneck Gene during
571 Evolution. Mol Biol Evol *35*, 1968-1981.

572 McGary, K.L., Slot, J.C., and Rokas, A. (2013). Physical linkage of metabolic genes in fungi is
573 an adaptation against the accumulation of toxic intermediate compounds. Proc Natl
574 Acad Sci USA *110*, 11481-11486.

575 Naseeb, S., and Delneri, D. (2012). Impact of chromosomal inversions on the yeast *DAL*
576 cluster. PLoS One *7*, e42022.

577 Oda, Y., and Fujisawa, T. (2000). Nucleotide sequence of alpha-galactosidase *MEL* gene from
578 *Zygosaccharomyces mrakii*. Curr Microbiol *41*, 220-222.

579 Oda, Y., and Fukunaga, M. (1999). Isolation and characterization of *MELt* gene from
580 *Torulaspora delbrueckii* IFO 1255. Yeast *15*, 1797-1801.

581 Oda, Y., and Tonomura, K. (1993). Selection of a Novel Baking Strain from the *Torulaspora*
582 Yeasts. Biosci Biotech Bioch *57*, 1320-1322.

583 Oda, Y., and Tonomura, K. (1996). α-Galactosidase from the yeast *Torulaspora delbrueckii*
584 IFO 1255. Journal of Applied Bacteriology, *80*, 203–208.

585 Riley, R., Haridas, S., Wolfe, K.H., Lopes, M.R., Hittinger, C.T., Goker, M., Salamov, A.A.,
586 Wisecaver, J.H., Long, T.M., Calvey, C.H.*, et al.* (2016). Comparative genomics of
587 biotechnologically important yeasts. Proc Natl Acad Sci USA *113*, 9882-9887.

588 Rokas, A., Wisecaver, J.H., and Lind, A.L. (2018). The birth, evolution and death of metabolic
589 gene clusters in fungi. Nat Rev Microbiol *16*, 731-744.

590 Saluja, P., Yelchuri, R.K., Sohal, S.K., Bhagat, G., Paramjit, and Prasad, G.S. (2012).
591 *Torulaspora indica* a novel yeast species isolated from coal mine soils. Antonie Van
592 Leeuwenhoek *101*, 733-742.

593 Shen, X.X., Opulente, D.A., Kominek, J., Zhou, X., Steenwyk, J.L., Buh, K.V., Haase, M.A.B.,
594 Wisecaver, J.H., Wang, M., Doering, D.T.*, et al.* (2018). Tempo and mode of genome
595 evolution in the budding yeast subphylum. Cell *175*, 1533-1545 e1520.

596 Slot, J.C., and Rokas, A. (2010). Multiple *GAL* pathway gene clusters evolved independently
597 and by different mechanisms in fungi. Proc Natl Acad Sci USA *107*, 10136-10141.

598 St John, T.P., and Davis, R.W. (1981). The organization and transcription of the galactose
599 gene cluster of *Saccharomyces*. J Mol Biol *152*, 285-315.

600 Varela, J.A., Puricelli, M., Ortiz-Merino, R.A., Giacomobono, R., Braun-Galleani, S., Wolfe,
601 K.H., and Morrissey, J.P. (2019). Origin of Lactose Fermentation in *Kluyveromyces*
602 *lactis* by Interspecies Transfer of a Neo-functionalized Gene Cluster during
603 Domestication. Curr Biol *29*, 4284-4290 e4282.

604 Viigand, K., Posnograjeva, K., Visnapuu, T., and Alamae, T. (2018). Genome Mining of Non-
605 Conventional Yeasts: Search and Analysis of MAL Clusters and Proteins. Genes
606 (Basel) *9*.

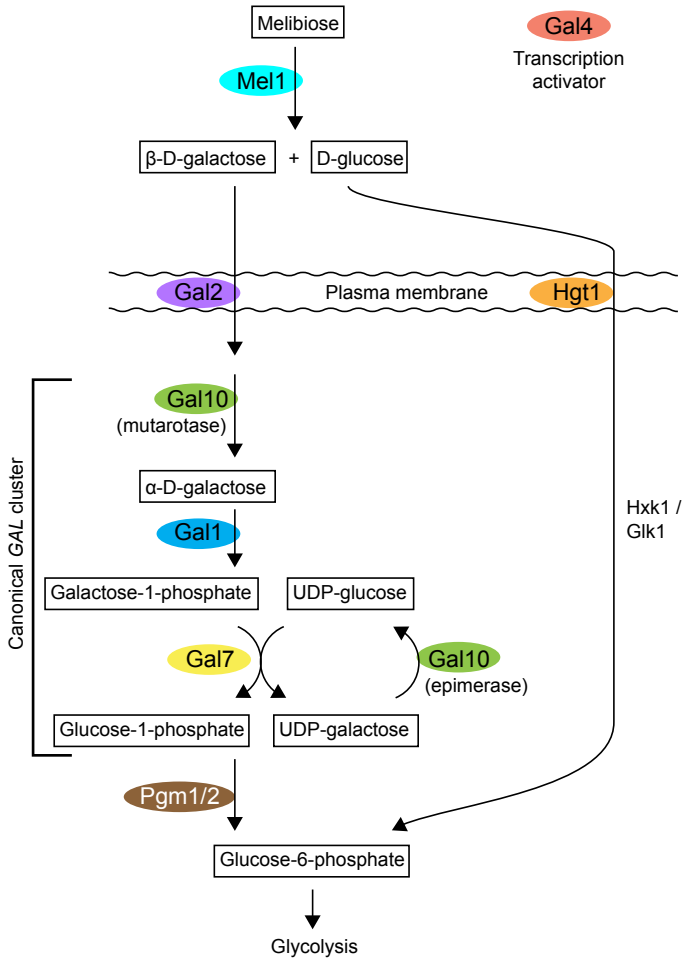
607 Webster, T.D., and Dickson, R.C. (1988). Nucleotide sequence of the galactose gene cluster
608 of *Kluyveromyces lactis*. Nucleic Acids Res *16*, 8192-8194.

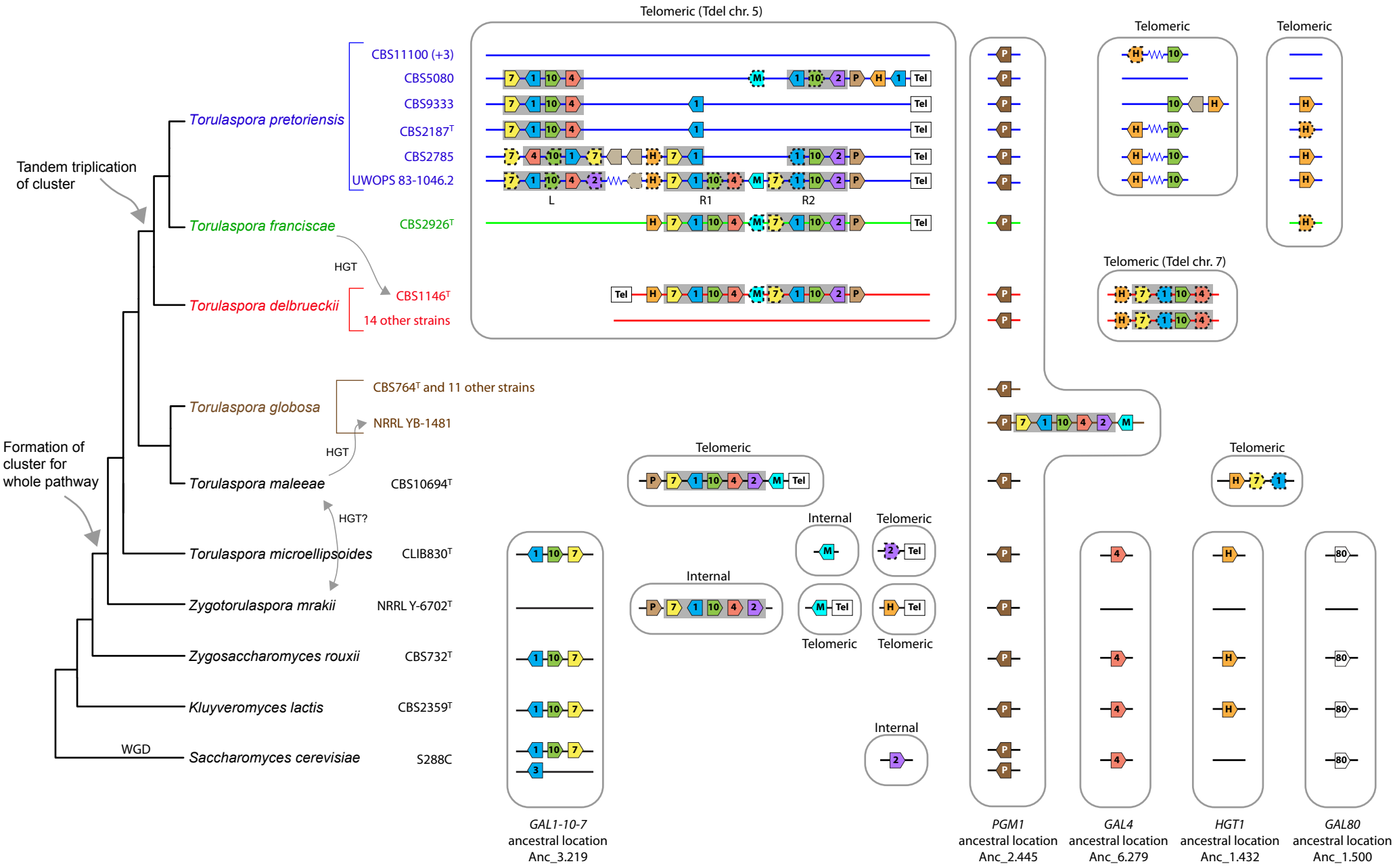
609 Wolfe, K.H., Armisen, D., Proux-Wera, E., OhEigeartaigh, S.S., Azam, H., Gordon, J.L., and
610 Byrne, K.P. (2015). Clade- and species-specific features of genome evolution in the
611 Saccharomycetaceae. FEMS Yeast Res *15*, fov035.

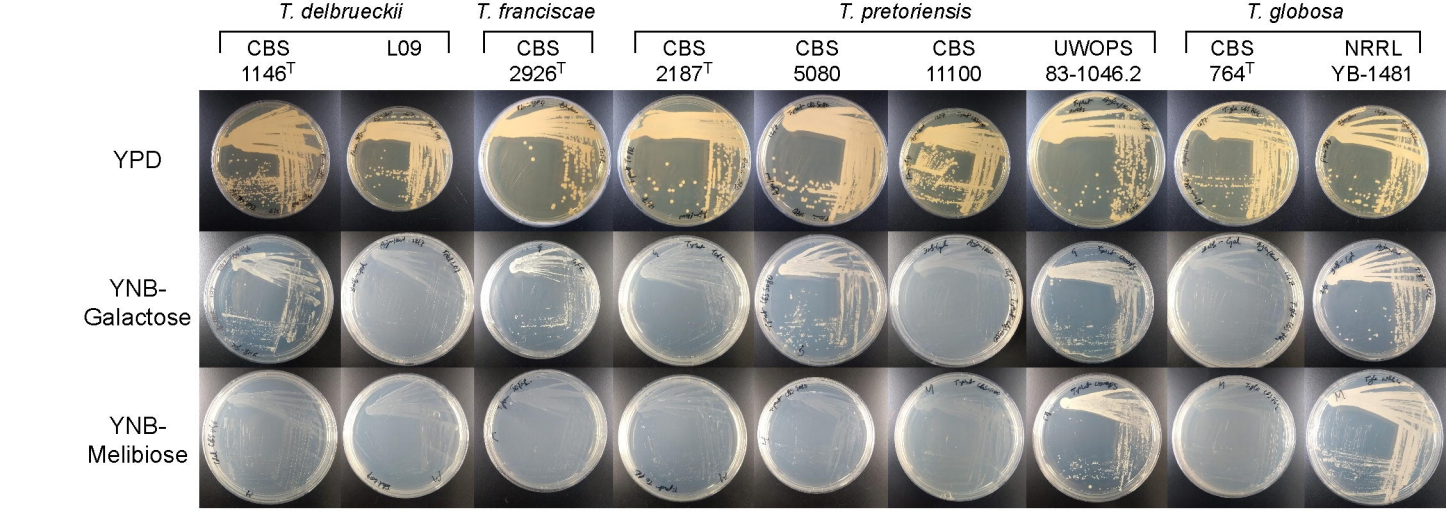
612 Wong, S., and Wolfe, K.H. (2005). Birth of a metabolic gene cluster in yeast by adaptive gene
613 relocation. Nature Genet *37*, 777-782.

614 Yamada-Okabe, T., Sakamori, Y., Mio, T., and Yamada-Okabe, H. (2001). Identification and
615 characterization of the genes for N-acetylglucosamine kinase and N-
616 acetylglucosamine-phosphate deacetylase in the pathogenic fungus *Candida*
617 *albicans*. Eur J Biochem *268*, 2498-2505.

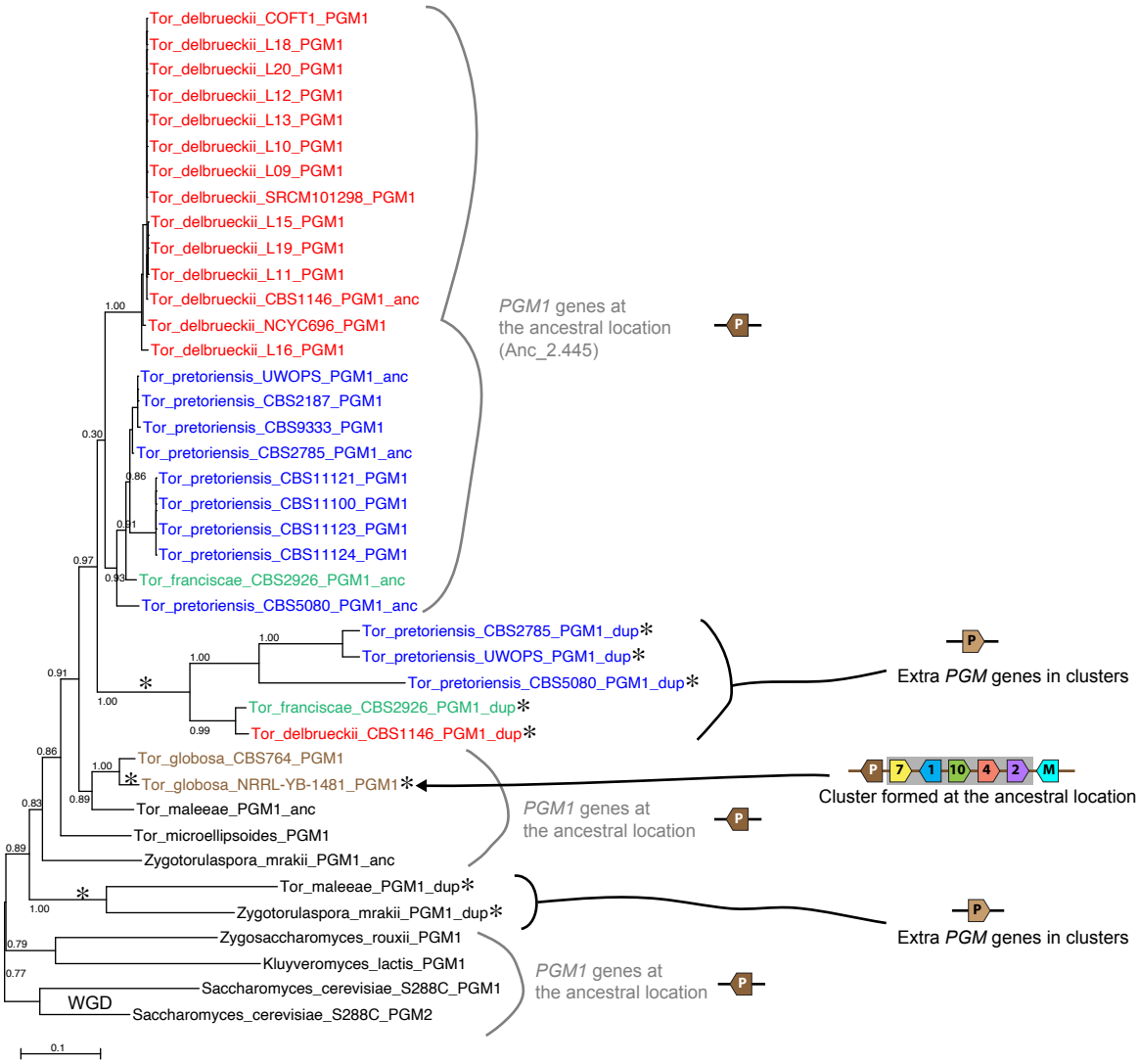
618


Melibiose
Gal4
Transcription activator
Mel1
β-D-galactose
+
D-glucose
Gal2
Plasma membrane
Hgt1
Gal10
(mutarotase)
α-D-galactose
Gal1
Galactose-1-phosphate
UDP-glucose
Gal7
Gal10
(epimerase)
Glucose-1-phosphate
UDP-galactose
Pgm1/2
Glucose-6-phosphate
Glycolysis
Hxk1 / Glk1
Canonical GAL cluster


Telomeric (Tdel chr. 5)
Telomeric
Telomeric
Telomeric (Tdel chr. 7)
Telomeric
Telomeric
Internal
Telomeric
Internal
Telomeric
Telomeric
Internal
Tandem triplication of cluster
Formation of cluster for whole pathway
Torulaspora pretoriensis
Torulaspora franciscae
Torulaspora delbrueckii
Torulaspora globosa
Torulaspora maleeae
Torulaspora microellipsoides
Zygotorulaspora mrakii
Zygosaccharomyces rouxii
Kluyveromyces lactis
Saccharomyces cerevisiae
CBS11100 (+3)
CBS5080
CBS9333
CBS2187T
CBS2785
UWOPS 83-1046.2
CBS2926T
CBS1146T
14 other strains
CBS764T and 11 other strains
NRRL YB-1481
CBS10694T
CLIB830T
NRRL Y-6702T
CBS732T
CBS2359T
S288C
HGT
HGT
HGT?
WGD
L
R1
R2
GAL1-10-7 ancestral location Anc_3.219
PGM1 ancestral location Anc_2.445
GAL4 ancestral location Anc_6.279
HGT1 ancestral location Anc_1.432
GAL80 ancestral location Anc_1.500



	T. delbrueckii CBS 1146^T	*T. delbrueckii* L09	*T. franciscae* CBS 2926^T	*T. pretoriensis* CBS 2187^T	*T. pretoriensis* CBS 5080	*T. pretoriensis* CBS 11100	*T. pretoriensis* UWOPS 83-1046.2	*T. globosa* CBS 764^T	*T. globosa* NRRL YB-1481
GAL1-10-7 genes	+	−	+	+	+	−	+	−	+
GAL2 gene	+	−	+	−	+	−	+	−	+
MEL1 gene	−	−	−	−	−	−	+	−	+


Tor_delbrueckii_COFT1_PGM1
Tor_delbrueckii_L18_PGM1
Tor_delbrueckii_L20_PGM1
Tor_delbrueckii_L12_PGM1
Tor_delbrueckii_L13_PGM1
Tor_delbrueckii_L10_PGM1
Tor_delbrueckii_L09_PGM1
Tor_delbrueckii_SRCM101298_PGM1
Tor_delbrueckii_L15_PGM1
Tor_delbrueckii_L19_PGM1
Tor_delbrueckii_L11_PGM1
Tor_delbrueckii_CBS1146_PGM1_anc
Tor_delbrueckii_NCYC696_PGM1
Tor_delbrueckii_L16_PGM1
Tor_pretoriensis_UWOPS_PGM1_anc
Tor_pretoriensis_CBS2187_PGM1
Tor_pretoriensis_CBS9333_PGM1
Tor_pretoriensis_CBS2785_PGM1_anc
Tor_pretoriensis_CBS11121_PGM1
Tor_pretoriensis_CBS11100_PGM1
Tor_pretoriensis_CBS11123_PGM1
Tor_pretoriensis_CBS11124_PGM1
Tor_franciscae_CBS2926_PGM1_anc
Tor_pretoriensis_CBS5080_PGM1_anc
PGM1 genes at the ancestral location (Anc_2.445)
Tor_pretoriensis_CBS2785_PGM1_dup
Tor_pretoriensis_UWOPS_PGM1_dup
Tor_pretoriensis_CBS5080_PGM1_dup
Tor_franciscae_CBS2926_PGM1_dup
Tor_delbrueckii_CBS1146_PGM1_dup
Extra PGM genes in clusters
Tor_globosa_CBS764_PGM1
Tor_globosa_NRRL-YB-1481_PGM1
Cluster formed at the ancestral location
Tor_maleeae_PGM1_anc
Tor_microellipsoides_PGM1
Zygotorulaspora_mrakii_PGM1_anc
PGM1 genes at the ancestral location
Tor_maleeae_PGM1_dup
Zygotorulaspora_mrakii_PGM1_dup
Extra PGM genes in clusters
Zygosaccharomyces_rouxii_PGM1
Kluyveromyces_lactis_PGM1
Saccharomyces_cerevisiae_S288C_PGM1
Saccharomyces_cerevisiae_S288C_PGM2
WGD
PGM1 genes at the ancestral location
0.1