Signatures of innovation and selection in the extremotolerant yeast *Kluyveromyces marxianus*

Kaylee E. Christensen[1,4], Aniliese Deal[1,2], Jun-Ting Johnson Wang[1,3], Abel Duarte[1], Judith L. Edwards[1], Jasmine L.N. Goodman[1], Zhenzhen Ma[1,5], Sasha I. Padilla[1], Edyta Szewczyk[1], Catherine Rha[1], Rachel B. Brem[1,*]

[1]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, 94720

[2]Graduate Group in Biophysics, University of California, Berkeley, Berkeley, CA, 94720

[3]Graduate Group in Comparative Biochemistry, University of California, Berkeley, Berkeley, CA, 94720

[4]Current address: Pacific Biosciences Research Center, University of Hawai'i at Mānoa, Honolulu, HI, 96822

[5]Current address: Department of Biology, Stanford University, Stanford, CA, 94305

*To whom correspondence should be addressed (rbrem@berkeley.edu)

Abstract

Organisms specialized to extreme environments can be the product of millions of years of evolutionary engineering and refinement. The underlying genetics can be quite distinct from the ones operating at earlier stages of trait innovation. In this work, we have developed the multi-stress resistant yeast *Kluyveromyces marxianus,* which diverged from its closest relative >20 million years ago, as a model for interspecies comparative biology and genomics. In growth assays of the *Kluyveromyces* genus, we found that *K. marxianus* exhibited unique tolerance of high heat and a subset of chemical stress conditions. We then generated and analyzed omic profiles from across the genus to find molecular features associated with—and potentially causal for—*K. marxianus* traits. Expression profiling revealed divergent lipid processing and membrane transport programs in *K. marxianus,* borne out in changes in lipid utilization in experimental assays. Sequence analyses found robust evidence for expansions in gene families in the *K. marxianus* genome, most notably among transmembrane transporters and in metabolic enzymes. In molecular-evolution tests, we identified adaptive protein variants throughout the *K. marxianus* genome among which plasma membrane transporters were over-represented. These data enable a model of the molecular mechanisms and evolutionary pressures underlying *K. marxianus* traits, including adaptive changes to transporters, lipid processing, and membrane functions mediating stress resistance.

Key Words

Trait variation, evolution, fungi, stress tolerance, cell membrane

Significance statement

Many traits of basic and applied interest arose long ago and manifest in the modern day as fixed in a given species; understanding how evolution built them, potentially over millions of years, remains a key challenge in the field. In this study, we report stress-resistance phenotypes that distinguish the yeast *Kluyveromyces marxianus* from its relatives, and we discover unique patterns of genetic and regulatory variation in membrane-protein genes, as well as unique properties of lipid metabolism, in this species. We propose a broadly applicable model in which evolution can tune membrane lipid composition and membrane-protein function to boost cellular fitness in challenging environments.

Introduction

A central goal of evolutionary biology is to understand trait diversity in the natural world. In many cases, a phenotype of ecological or industrial interest defines a given eukaryotic species, distinguishing it from its relatives. Elucidating the genetic underpinnings of any such trait across species boundaries poses unique technical challenges. As a consequence, against a backdrop of landmark case studies mapping genotype to phenotype over deep trait divergences (Weiss and Brem 2019; Smith et al. 2020; Masly and Azom 2022), principles remain at a premium. What kinds of phenotypes tend to fix in species as they split from one another? What are the cellular mechanisms and the underlying genes and variants? Addressing these questions requires detailed study of ancient traits from physiology to the molecular level.

64

65 Single-celled fungi can serve as a powerful model system for evolutionary genetics (Hittinger et
66 al. 2015). *Kluyveromyces,* a genus of budding yeasts, are well-suited to this paradigm given
67 their ~25-million-year divergence (Shen et al. 2018) and wide-ranging ecology. *Kluyveromyces*
68 have been isolated from a range of terrestrial and aquatic environments (Anderson et al. 1986;
69 Nagahama et al. 1999; Am-In et al. 2008; Araujo and Hagler 2011; Suzuki et al. 2014; Cernak
70 et al. 2018; Williams et al. 2019; Chacón-Vargas et al. 2020; Al-Qaysi et al. 2021; Baptista et al.
71 2021; Avchar et al. 2022; Avchar et al. 2022; Spurley et al. 2022; Verma et al. 2023), but are
72 best known for roles in food and beverage fermentations (Kurtzman et al. 2011), including
73 elegant evolutionary analyses of the origin of lactases in dairy-associated lineages (Varela,
74 Puricelli, Ortiz-Merino, et al. 2019; Friedrich et al. 2023). An extensive literature has focused on
75 the growth of one species in the genus, *K. marxianus,* at temperatures exceeding 45°C,
76 motivated by applications in the bioproduction industry (Lane et al. 2011; Lertwattanasakul et al.
77 2013; Mejía-Barajas et al. 2017; Madeira-Jr and Gombert 2018; Fu et al. 2019; Hoshida et al.
78 2020; Wu et al. 2020; Kosaka et al. 2022; Flores-Cosío et al. 2024). The evolutionary driver for
79 the latter trait is unknown, though *K. marxianus* has been isolated from self-heating agricultural
80 and domestic compost (Suzuki et al. 2014; Avchar et al. 2022; Verma et al. 2023). Likewise,
81 cellular and genetic mechanisms underlying unique *K. marxianus* traits have remained
82 incompletely understood (Nespolo et al. 2020); potential clues have emerged from comparison
83 of this species to distantly related yeasts, including reports of resistance to chemical stress
84 (Saini et al. 2017; Linder 2018; Illarionov et al. 2021) and expansions in sugar and potassium
85 importer genes (Varela, Puricelli, Montini, et al. 2019; Papouskova et al. 2024).

86

87 We set out to use a comparative approach to trace in detail how *K. marxianus* diverged from its
88 relatives. We first catalogued stress resistance characters that distinguished *K. marxianus* from

89 the rest of the species of its clade. We then applied omic profiling to find unique expression
90 programs and genomic features in *K. marxianus*. The results provided insights into the
91 underpinnings and the evolution of the *K. marxianus* trait syndrome.
92

93 ## Results

94

95 ### ***K. marxianus* exhibits a stress resistance syndrome unique in its genus**

96

97 With the ultimate goal of understanding evolutionary innovations in *Kluyveromyces*, we first
98 sought to survey stress resistance traits across the genus. We profiled the growth in liquid
99 culture of four species associated with terrestrial habitats (*K. marxianus, K. lactis, K.*
100 *wickerhamii*, and *K. dobzhanskii*) and three aquatic species (*K. aestuarii*, *K. siamensis*, and *K.*
101 *nonfermentans*) in a panel of media conditions (Figure 1 and Table S1). At 42°C, *K. marxianus*
102 grew far better than its relatives, to an extent beyond the pattern in a permissive control
103 condition (rich medium at 28°C; Figure 1A); genetically distinct *K. marxianus* isolates from
104 various substrates and geographical origins phenocopied each other with respect to high-
105 temperature growth except for one thermosensitive strain isolated from a fly (Figure S2, Table
106 S1), as expected if loss of thermotolerance were a rare and recent event in this predominantly
107 thermotolerant species (Lane et al. 2011). In ethanol, caffeine, the DNA intercalator propidium
108 iodide, and the DNA damage agent methyl methanesulfonate (MMS), *K. marxianus* likewise
109 outperformed the other species of the genus (Figure 1A). Our profiling also identified other
110 conditions in which *K. marxianus* did not stand out with respect to growth (Figure S1B). We
111 conclude that *K. marxianus* is unique in the genus for its resistance to a broad subset of abiotic
112 challenges, including high temperature.

113

114 **The uniqueness of *K. marxianus* resistance phenotypes depends on growth phase**

115

116 For many microbes, when cells face nutrient exhaustion, they shut down exponential (log-
117 phase) growth and enter the quiescent, extra-durable state called stationary phase. The
118 difference between *K. marxianus* and its relatives in stress resistance that we had noted across
119 a growth timecourse (Figure 1A) might also manifest in the stationary state or, alternatively, its
120 cellular mechanism could be tied to active cell division. To distinguish between these
121 possibilities, we focused on thermotolerance as a test case. We used a viability assay in which
122 we inoculated a given liquid culture into rich medium; incubated for growth at the permissive
123 temperature until nutrient exhaustion; subjected this stationary-phase culture to a short heat
124 shock; and used it to seed colonies for growth at the permissive temperature on solid medium
125 (Figure 1B). Results revealed only modest differences between species in heat shock
126 resistance under this stationary-phase heat treatment, with no particular advantage for *K.*
127 *marxianus* (Figure 1B). By contrast, when we treated exponential-phase cultures with heat
128 shock, *K. marxianus* exhibited much higher viability than did the other species of the genus
129 (Figure 1C), paralleling its advantage during chronic high-temperature exposure (Figure 1A). No
130 such differences were apparent in a permissive 28°C control experiment (Figure S3B). We
131 conclude that for high temperature, the *K. marxianus* advantage only manifests when cells are
132 actively dividing, and that cell death, rather than slow or arrested growth, mediates the defects
133 of the rest of the genus.

134

135 To extend our analysis of cell death as the underlying cellular basis of species differences in
136 stress, we next assayed viability in exponential-phase cultures across a panel of representative
137 stresses, focusing on the stress-sensitive species *K. lactis* as a point of comparison for *K.*

138 *marxianus* (Figure S1A). The latter exhibited a survival advantage in all stress conditions tested
139 (Figure S4). The difference was amplified by oxygen depletion, though both species performed
140 better in normoxia (Figure S4), as expected (Lertwattanasakul et al. 2013; Mejía-Barajas et al.
141 2017; Madeira-Jr and Gombert 2018; Fu et al. 2019; Hoshida et al. 2020; Kosaka et al. 2022).
142 Together, these trends from assays of viability establish the resistance character of *K.*
143 *marxianus* as a lineage-unique gain of protection against lethal failures of the proliferation
144 program under temperature and chemical stress.
145
146 **Unique expression and lipid utilization programs in *K. marxianus***
147
148 To learn more about the underpinnings of divergence across *Kluyveromyces*, we turned to an
149 expression profiling approach, for which we again chose *K. lactis* as a representative stress-
150 sensitive species for comparison against *K. marxianus*. We surveyed the transcriptomes of the
151 two species under exponential growth in representative treatments where we had observed
152 divergent growth phenotypes—high temperature, caffeine, MMS, and ethanol—and in the
153 permissive control of rich medium at 28°C, with stress conditions chosen to enable growth of
154 both species (Table S2 and see Methods). Principal component analyses of normalized RNA-
155 seq read counts made clear that the major source of variation across the data set was a marked
156 difference between *K. lactis* and *K. marxianus* transcriptomes in all conditions, including
157 controls (Figure 2A, Figure S5, and Table S2). Functional-enrichment tests revealed induction
158 of membrane-lipid trafficking and intracellular transport, clathrin endocytosis, and cell wall genes
159 driving the values of the major principal component in *K. marxianus* transcriptomes (Figure 2B
160 and Table S2). High expression of translation and ribosomal RNA processing genes drove the
161 values of the major principal component in *K. lactis* samples (Figure S5 and Table S2). Apart
162 from these effects, the second principal component reflected a suite of genes that *K. lactis*

163 turned up and down depending on treatment, whereas the *K. marxianus* transcriptome was
164 much less volatile with respect to stress (Figure 2A, Figure S5, and Table S2), as expected if *K.*
165 *lactis* underwent more functional failures during growth in our panel of conditions. Most salient
166 in the set of stress-responsive genes in *K. lactis* was heightened induction of translation gene
167 sets over and above the effect we had noted in control conditions (Figure 2C, Figure S5, and
168 Table S2), dovetailing with previous studies (Lertwattanasakul et al. 2015). Functional-genomic
169 enrichment analysis of raw expression measurements yielded similar trends (Table S3).
170
171 Our observations of divergent expression of membrane-lipid trafficking genes in *K. marxianus*
172 (Figure 2B and Table S2) echoed previous comparative-biology studies of an unusual inability
173 by *K. marxianus* to take up sterol from growth medium (Dekker et al. 2021) and unusual effects
174 of sterol-metabolism mutations (Ren et al. 2024). We hypothesized that *K. marxianus* could
175 likewise exhibit divergent phenotypic responses to other classes of lipid. To explore this notion,
176 we used an experimental paradigm testing utilization by yeast of exogenous lipids when *de*
177 *novo* fatty acid synthesis is blocked by the drug cerulenin (Jung et al. 2013). Using oleate as the
178 lipid supplied in the medium in this assay, we found *K. marxianus* to be among the lowest
179 performers of the genus (Figure S6), a surprising contrast to its advantages in stress conditions
180 (Figure 1A). This minimal exogenous oleate processing by *K. marxianus*, even when it was
181 required for optimal growth, paralleled the behavior of *K. marxianus* in sterols (Dekker et al.
182 2021), suggestive of a hard-wired preference by this species for internal lipid sources.
183 Interestingly, the aquatic species *K. siamensis* and *K. nonfermentans* also failed to utilize
184 exogenous oleate in our assays (Figure S6), plausibly as a product of independent evolutionary
185 losses in their niche.
186
187 **Unique transporter and metabolic gene duplications in the *K. marxianus* genome**

188

189 We reasoned that genomic features distinguishing *K. marxianus* from its relatives could serve
190 as a source for clues to the underlying molecular basis of its traits. As a first strategy to find
191 genomic novelties in *K. marxianus*, we focused on species-specific copy number change in
192 paralogous gene families. From annotated reference genomes for six *Kluyveromyces* species
193 (*K. marxianus, K. lactis, K. wickerhamii*, *K. dobzhanskii*, *K. aestuarii*, and *K. nonfermentans*;
194 Figure S1A), plus outgroups from nearby clades in the Saccharomycotina subphylum (including
195 *S. cerevisiae* and *C. albicans* as well-annotated model organisms), we identified rapidly
196 changing gene families in each branch with tools of the CAFE 5 package (Mendes et al. 2021).
197 This approach identified 16 gene families with inferred gain or loss events along the *K.*
198 *marxianus* lineage at rates that were unusual relative to the average across the species set
199 (Figure 3A and Table S4), including families subject to tandem duplications and those with
200 genes across the *K. marxianus* chromosomes (Figure 3B and Table S4). Of these, 13 were
201 families with expansions in *K. marxianus* that could be classified based on domain annotations
202 as functioning in metabolism or localized to the plasma membrane, the latter including
203 transporters annotated as nutrient importers and xenobiotic efflux pumps (Figure 3A). These
204 data make clear that gene family expansions are prevalent in the *K. marxianus* genome, and
205 they highlight membrane proteins and metabolism as compelling candidate molecular players
206 underlying *K. marxianus* phenotypes.

207

208 **Widespread phylogenetic evidence for positive selection on amino acid alleles in *K.***
209 ***marxianus***

210

211 As a further means to identify genomic correlates of *K. marxianus* phenotypes, we pursued
212 molecular-evolution analyses of amino acid variation on a per-gene basis. We reasoned that

213 some phenotypes unique to *K. marxianus* could have evolved as a product of adaptation to
214 challenging ecological niches, and that at relevant loci, we could detect evidence for positive
215 selection on protein-coding sequences. For a phylogenetic test toward this end, at each gene in
216 turn, we aligned one-to-one orthologs from reference genomes for our six focal *Kluyveromyces*
217 species, to which we fit models in the PAML (Yang 2007) and HyPhy (Pond et al. 2005)
218 packages earmarking codons subject to positive selection along the *K. marxianus* lineage.
219 Emerging from these tests were >500 and 80 genes with evidence for positive selection from
220 PAML and HyPhy analyses, respectively (Figure 4A-B), overlapping in 51 cases (Figure 4B and
221 Table S5). We considered the latter genes, given their signal in multiple independent analysis
222 frameworks, as representing the strongest cases for positive selection in *K. marxianus*
223 according to phylogenetics. This hit list included the potassium transporter *TRK1*, whose *K.*
224 *marxianus* ortholog is known to be hyperactive relative to that of model yeast (Papouskova et al.
225 2024), and other genes annotated in transmembrane transport as well as metabolism, cell wall
226 structure, and protein production (Figure 4B and Table S5). Control analyses revealed far less
227 signal for positive selection on the lineage leading to *K. lactis* (Figure 4A and Table S5),
228 supporting the inference of a history of particularly pervasive adaptation at the amino acid level
229 in *K. marxianus*.
230
231 For most of the many gene hits from our phylogenetic tests focused on *K. marxianus*, a small
232 number of codons per gene were classified as targets of positive selection (Table S5). We
233 expected that if the derived alleles at each such site had undergone a selective sweep through
234 *K. marxianus* populations, we could detect it across strains of this species. To test this, we
235 carried out short-read resequencing of the genomes of five wild *K. marxianus* isolates (Table
236 S1) and used the resulting reads alongside our reference genome of the Km31 *K. marxianus*
237 strain as input to a pipeline for variant calling, pseudogenome construction, and sequence

238 alignment. We then inspected these alignments at each site where the *K. marxianus* allele had
239 been flagged as arising under selection according to our phylogenetic analysis of reference
240 genomes. At 99% of the latter, we detected complete conservation across *K. marxianus* strains
241 (Figure 4C and Table S5). These data establish that alleles in Km31 are representative of the
242 population more broadly at these loci, providing further support for our inference of their derived
243 alleles maintained under selection.
244
245 **Enrichment for amino acid variation between *Kluyveromyces* species in transporter**
246 **genes**
247
248 To complement our phylogenetic analyses, we next applied population-based molecular-
249 evolution tests that used our *K. marxianus* resequencing data and a population-genomic
250 resource for *K. lactis* (Friedrich et al. 2023). We implemented McDonald-Kreitman molecular
251 evolution paradigms (McDonald and Kreitman 1991) to assess evidence for differences in
252 selective pressure between the species on a given protein sequence, as follows. For each
253 coding region we tabulated sites of non-synonymous and synonymous divergence between *K.*
254 *marxianus* and *K. lactis*, and likewise for polymorphic variants within the species (Table S6).
255 Using these counts in a non-polarized McDonald-Kreitman test on a per-gene basis, we
256 identified 11 genes with a significant excess of interspecies amino acid divergence across the
257 coding region (relative to polymorphism and for synonymous variation; Figure 4D and Table
258 S6). Thus, relative to our phylogenetic methods (Figure 4A), the McDonald-Kreitman scheme
259 was under-powered in this system, as expected if most loci only had a few amino acid targets of
260 adaptation between *K. marxianus* and *K. lactis* (Table S5). But among the gene hits that did
261 emerge from the McDonald-Kreitman test, each represented an inference of positive selection
262 from population data that complemented our phylogenetic results. Two such loci had also been

263 top-scoring in both of our phylogenetic frameworks, the annotated glycolytic enzyme gene *FBA1*
264 and *FAT3,* annotated in lipid transport (Figure 4C-D). Allele-swap experiments revealed a small
265 (1.09-fold) but significant advantage accruing from the *K. lactis* allele of *FAT3* in the *K.*
266 *marxianus* background in assays of lipid utilization, consonant with the better performance by
267 wild-type *K. lactis* in the latter (Figure 4E and Figure S6); no such allele-swap effect was
268 detectable in standard-media controls (Figure S7). This result validated our inference from
269 sequence analyses (Figure 4C-D) of differential function of *FAT3* between the species, and
270 attested to a modifying role for *FAT3* in a complex genetic architecture of lipid processing
271 preferences.
272
273 For more holistic insights into *Kluyveromyces* evolution using our population-genomic data
274 resource, we developed a group-wise version of the McDonald-Kreitman paradigm, in which we
275 tabulated the neutrality index (NI) for a given gene as non-synonymous divergence normalized
276 for polymorphism and for synonymous variation, and we evaluated significance of this metric
277 across cohorts of genes of similar function via resampling (Figure 4F and Table S7). For
278 metabolic gene groups with extreme NIs in this test, focused analyses indicated that the signal
279 was driven by low rates of protein evolution within species, reflecting a history of purifying
280 selection (Table S7). By contrast, for five Gene Ontology terms, the individual components of
281 the NI metric followed the opposite pattern: non-synonymous variation between species was
282 high in the genes of the group and polymorphism levels were on par with the genomic average
283 (Figure 4E and Table S7). Of particular note among these gene groups enriched for low NI was
284 a transmembrane transport gene cohort (Figure 4F), echoing our observation of membrane
285 protein gene family expansions in *K. marxianus* (Figure 3) and the annotated transporters,
286 including *TRK1* and *FAT3*, that had been top-scoring in our single-gene phylogenetic and
287 McDonald-Kreitman analyses (Figure 4D and Table S5). Taken together, our molecular-

evolution analyses attest to a complex history of selection on loci throughout the *K. marxianus* genome, with strong concordance across our test approaches for evidence for unique selective pressures on the coding regions of transporter genes.

Discussion

Against the backdrop of years of work mapping the genetic basis of trait variation within fungal species (Liti and Louis 2012; Taylor et al. 2017), dissecting fixed trait divergences between species remains a key challenge for mycologists. In landmark cases, individual loci have been identified that underlie such lineage-defining phenotypes in fungi (Nagy et al. 2017; Wang et al. 2024). For the vast majority of traits varying across fungal species barriers, however, the mechanistic underpinnings remain unknown. In this work, we have established a framework for discovery of candidate genes and pathways underlying traits unique to *K. marxianus*. Our results have revealed attributes that distinguish *K. marxianus* from the rest of its genus: stress resistance in growth assays, divergent expression programs and lipid processing phenotypes, and sequence evidence for gene-content and amino-acid variation. These observations serve to complement classic trait compendia (Kurtzman et al. 2011) and the modern literature (Lane and Morrissey 2010; Lane et al. 2011; Madeira-Jr and Gombert 2018). In terms of genetics, the alleles we have found to be unique in *K. marxianus* represent candidates for the determinants that nature would have used to build the trait syndrome in this species. We expect that as these loci are validated, they will prove sufficient to reconstitute traits from *K. marxianus* in an exogenous background, as opposed to the handful of previous studies that have reported genes necessary for thermotolerance in *K. marxianus* (Falcão Moreira et al. 1998; Wu et al. 2020; Montini et al. 2022).

312

313 Most pervasive in our omic profiles was the trend of induction of membrane-gene expression,
314 expanded membrane-gene families, and derived amino acid alleles in single-copy transporter
315 genes, in the *K. marxianus* genome relative to the rest of its clade. The latter include gains by *K.*
316 *marxianus* in sugar and potassium transporter repertoires, which were previously analyzed in
317 comparisons with distant fungal relatives (Varela, Puricelli, Montini, et al. 2019; Papouskova et
318 al. 2024) and which we can now date to the divergence of *K. marxianus* from its sister species.
319 A causal phenotypic role of this remodeling of the transporter repertoire in the stress resistance
320 by *K. marxianus* remains to be established. It is tempting to speculate that *K. marxianus* has
321 tuned many of its transporters to tighten selectivity against extracellular toxins or to improve
322 function in the face of heat or chemical stress. Additionally, given the evidence from our work
323 and that of others (Dekker et al. 2021) for heightened membrane trafficking and limited
324 exogenous lipid usage in *K. marxianus*, we propose that evolution has tuned intracellular lipid
325 processing in this species, with molecular mechanisms to restrict exogenous uptake that involve
326 *FAT3*. Future analyses will be necessary to test directly for a relationship between lipid
327 processing changes and stress resistance in *K. marxianus*, potentially including specialized
328 programs to tune membrane fluidity as seen in other yeast systems (Henderson and Block
329 2014; Lairón-Peris et al. 2021; Ribeiro et al. 2022; Sokolov et al. 2022). Any such unique
330 attributes of *K. marxianus* membrane lipids could be interconnected with the genetic and
331 expression changes we have observed at transporter genes, *e.g.* to adjust protein anchoring or
332 to compensate for changes to membrane barrier function that would impact passive nutrient
333 acquisition (Ferenci 2016). That said, we observed signals of evolutionary innovation by *K.*
334 *marxianus* in genes of many other functions, and globally, we picture a multifaceted molecular
335 mechanism for its resistance traits.

336

337 Why, ecologically, would *K. marxianus* have evolved stress tolerance? The ecological pressures
338 that would have driven changes in this lineage remain poorly established. Though well-
339 recognized as an industrial dairy yeast (Prado et al. 2015; Gethins et al. 2016; Ortiz-Merino et
340 al. 2018; Varela, Puricelli, Ortiz-Merino, et al. 2019; Yazdi et al. 2022), *K. marxianus* was likely
341 domesticated for this purpose fairly recently, as inferred for *K. lactis* (Friedrich et al. 2023). And
342 though *K. marxianus* can also be isolated from rotting fruit (Williams et al. 2019; Al-Qaysi et al.
343 2021), it may not have specialized narrowly to such substrates: a neighboring clade that
344 includes *Saccharomyces cerevisiae* produces more ethanol than do any of the *Kluyveromyces*
345 species (Nespolo et al. 2020), suggesting a weaker drive in the latter toward this putative
346 strategy to kill off microbial competitors in high-sugar environments. Among microbial-ecology
347 findings, the strongest clue to the origin of unique phenotypes in *K. marxianus* lies in its
348 prevalence in decomposing plant material other than fruit, including leaf litter and pine needles
349 (Spurley et al. 2022), cow dung (Avchar et al. 2022), industrially processed cocoa beans
350 (Baptista et al. 2021), agave (Chacón-Vargas et al. 2020), and sugar cane (Anderson et al.
351 1986; Cernak et al. 2018) as well as industrial and domestic compost (Suzuki et al. 2014;
352 Avchar et al. 2022; Verma et al. 2023). Plant defense compounds in decomposing material are
353 a compelling candidate driver for the rise of resistance traits in *K. marxianus* (Baptista and
354 Domingues 2022) since they can persist for days in a leaf or stem after abscission (Burton
355 1982; Goodspeed et al. 2013; Shimada et al. 2015). Furthermore, some evidence for low pH
356 deep in compost piles (Rainey and Covey 2016) suggests that this condition could also have
357 contributed to *K. marxianus* evolution. In this scenario, the phenotypic syndrome of *K.*
358 *marxianus* would involve not only resistance to the stresses of the decaying plant environment,
359 but also the rapid utilization of simple sugars liberated early in plant breakdown, mediated by
360 the new copies and new alleles of metabolic genes we have noted in the *K. marxianus* genome.

361 As *K. marxianus* lacks the capacity to degrade lignocellulosic plant cell walls, these would be
362 left to other fungal saprotrophs acting later in decomposition.
363
364 We thus envision an ecological model in which *K. marxianus* originally evolved as a plant decay
365 specialist, with extant strains coming to occupy a broad range of niches, including soil and
366 human and animal material (Rij 2013; Ortiz-Merino et al. 2018; Zhang et al. 2021). Given that
367 we and others have seen conservation of resistance traits across *K. marxianus* strains (Lane et
368 al. 2011; Madeira-Jr and Gombert 2018; Lappe-Oliveras et al. 2023), we infer that unique
369 capacities arose early in *K. marxianus,* sometime after *K. marxianus* originated as a species,
370 ~25 million years ago (Shen et al. 2018), and persisted during its radiation into modern
371 populations.
372

373 # Methods

374

375 **1. Strains**

376

377 Strains used in phenotyping experiments, population-genomic analyses, and mutant
378 phenotyping are listed in Table S1.
379

380 **2. Liquid growth phenotyping of wild-type *Kluyveromyces***

381

382 Liquid culture experiments in stress and standard-condition controls in Figure 1A used *K.*
383 *marxianus* Km31 (Cernak et al. 2018); *K. lactis* yLB72 (Sorrells et al. 2015); and the strains of
384 *K. dobzhanskii, K. wickerhamii, K. nonfermentans, K. aestuarii*, and *K. siamensis* listed in Table

385 S1. For a given species and condition, we streaked a frozen stock onto yeast peptone dextrose
386 (YPD) agar plates, incubated for two days at 28°C, picked a single colony, inoculated into 5 mL
387 liquid YPD, and incubated with shaking at 28°C for 16-18 hours. We call this the first pre-
388 culture. We diluted, into fresh liquid YPD, an inoculum of the first pre-culture at the volume
389 required to attain an optical density (OD_{600}) of 0.05 for a total culture volume of 5 mL, and we
390 incubated this second pre-culture with shaking at 28°C for ~7 hours until it reached $OD_{600} = 0.2$.
391 For unstressed controls in Figure 1A, we diluted an inoculum of the second pre-culture into
392 fresh liquid YPD at the volume required to attain $OD_{600} = 0.2$ for a total culture volume of 5 mL.
393 150 μL of this experimental culture was transferred to each of 12 wells of a 96-well plate, sealed
394 with a Breathe-Easy membrane (Diversified Biotech), for incubation at 28°C and growth
395 measurements across a 24-hour timecourse in a Tecan M200 PRO plate reader (parameter
396 time point measurement, OD_{595}). For heat stress treatment in Figure 1A, setup and growth
397 assays were as above except that incubation of the experimental culture was at 42°C. For
398 chemical stress treatments in Figure 1A, setup and growth assays were as above except that
399 for the experimental culture, before transfer to the 96-well plate, ethanol (Koptec) was added to
400 attain a concentration of 6.5% (v/v); caffeine (Sigma-Aldrich) was added to attain 25 mM;
401 methane methylsulfonate (MMS; Thermo Scientific) was added to attain 0.03% (v/v); or
402 propidium iodide was added to attain 1mM, and growth timecourses were 48 to 72 hours. For
403 chemical stress treatments and growth measurements in Figures S1 and S2, for a given
404 species and condition, streaking and the first pre-culture were as above. We diluted an
405 inoculum of the first pre-culture into fresh liquid YPD at the volume required to attain $OD_{600} = 0.2$
406 for a total culture volume of 5 mL, with 1 M NaCl (Fisher), 3% glycerol (v/v; Fisher), 80 mM
407 $CaCl_2$ (Sigma), 250 mM rapamycin (Research Products International), 6 mM H_2O_2 (Fisher), or
408 16% (v/v) Triton X-100 (Fisher), followed by incubation and growth measurements in the 96-well
409 format as above.

410

411 **3. Assays of cell survival from *K. marxianus* and *K. lactis* stationary and log-phase liquid**

412 **culture**

413

414 For survival assays of cells in exponential phase in Figures 1B,C and S3, for *K. marxianus*

415 Km31 and separately for *K. lactis* yLB72 we streaked a frozen stock onto yeast peptone

416 dextrose (YPD) agar plates, incubated for two days at 28°C, picked a single colony, inoculated

417 into 5 mL liquid YPD, and incubated with shaking at 28°C for 16-18 hours. We diluted, into fresh

418 liquid YPD, an inoculum of this first pre-culture at the volume required to attain OD_{600} = 0.05 for

419 a total culture volume of 5 mL, and we incubated this second pre-culture with shaking at 28°C

420 for ~7 hours until it reached OD_{600} = 0.2. We took this second pre-culture directly into

421 experimental treatment as follows. For unstressed controls, serial dilutions of the culture were

422 spotted onto agar plates and incubated for 2 days at 28°C. For heat shock, 1 mL of the culture

423 was incubated in a heat block at 47°C for 30 minutes and then transferred to ice for 2 minutes,

424 followed by spotting and incubation as above.

425

426 For survival assays of cells in stationary phase in Figures 1B and S3, for *K. marxianus* Km31

427 and separately for *K. lactis* yLB72 we streaked a frozen stock onto yeast peptone dextrose

428 (YPD) agar plates, incubated for two days at 28°C, picked a single colony, inoculated into 5 mL

429 liquid YPD, and incubated with shaking at 28°C for 96 hours. Heat shock and spotting (and

430 unstressed controls) were as above.

431

432 **4. Assays of cell survival from *K. marxianus* and *K. lactis* anaerobic and aerobic liquid**

433 **cultures**

434

435 As *Kluyveromyces* cannot tolerate strictly anaerobic conditions (Kiers et al. 1998; Ishtar Snoek
436 and Yde Steensma 2006; Merico et al. 2007; Dekker et al. 2021), for Figure S4 we used a
437 limited-oxygen experimental paradigm (Hoshida et al. 2020) as follows. For *K. marxianus* Km31
438 and separately for *K. lactis* yLB72, for each treatment we isolated a single colony on solid plates
439 and carried out first and second pre-cultures in liquid YPD at 28°C under aerobic conditions as
440 in Section 2. For unstressed controls in anaerobic conditions, we diluted an inoculum of the
441 second pre-culture into fresh liquid YPD at the volume required to attain $OD_{600} = 0.05$ for a total
442 culture volume of 5 mL. 150 μL of this experimental culture was transferred to each of 12 wells
443 of a 96-well plate, which was placed (with lid) into an AnaeroPouch (ThermoScientific) with a
444 AnaeroPouch - Anaero gas pack. The sealed pouch and plate were incubated for 72 hours at
445 28°C. Then we collected 30 μL of culture from each well and combined them all into one master
446 culture, 100 μL of which we serially diluted into spots on agar plates, followed by incubation for
447 2 days at 28°C. Aerobic controls were treated analogously except that the 96-well plate was
448 inoculated without a pouch. Viability assays of chemically stressed cells in anaerobic conditions
449 and aerobic controls were as above except that for the experimental culture, before transfer to
450 the 96-well plate, ethanol, caffeine, or MMS were added as in Section 2. Viability assays of
451 heat-treated cells in anaerobic conditions and aerobic controls were as above except that the
452 sealed pouch and plate were incubated for 24 hours at 39°C.
453
454 **5. Transcriptional profiling**
455
456 *5.1. Culture protocol and experimental design.*
457
458 For a given replicate of RNA-seq of *K. marxianus* Km31 and separately for *K. lactis* yLB72 in
459 unstressed control conditions, we isolated a single colony on solid plates and carried out first

460 and second pre-cultures in liquid YPD at 28°C as in Section 2. For unstressed controls, we
461 diluted an inoculum of the second pre-culture into fresh liquid YPD at the volume required to
462 attain OD_{600} = 0.05 for a total culture volume of 50 mL. This experimental culture was incubated
463 at 28°C with shaking until it reached OD_{600} = 0.4-0.85. From this culture we took an aliquot
464 whose volume contained 2 OD units of cells, pelleted its cells by centrifugation for 5 minutes at
465 1,500 x g, and resuspended in 1 mL of RNAlater (ThermoFisher) followed by storage at 4°C for
466 up to 48 hours. For RNA-seq of heat-stressed cells, culture, pelleting, and storage were as
467 above except that the experimental culture was incubated at 39°C; for RNA-seq of cells under
468 chemical stress, culture, pelleting, and storage were as above except that the experimental
469 culture was done in YPD with 22 mM caffeine, 0.01% MMS, or 3% ethanol. In each case, we
470 used incubation times for the respective condition and species that allowed the culture to reach
471 the target optical density (*i.e.* mid-exponential phase). For expression-based reannotation of *K.*
472 *marxianus* Km31 and *K. lactis* yLB72 (see below), we used the above pipeline for a preliminary
473 expression profiling experiment which we call the training batch: we cultured one biological
474 replicate of Km31 and yLB72 at 37°C and one of Km31 and yLB72 in unstressed control
475 conditions. For comprehensive transcriptional profiling analyses in Figure 2 and Figure S5, we
476 used the above pipeline to generate cell pellets in two production batches. In one batch we
477 grew three biological replicates of Km31 and three of yLB72 in 39°C heat stress, and three
478 unstressed cultures of each species. In another batch we grew three biological replicates of
479 each species in caffeine, three in MMS, and three in ethanol, and three unstressed controls.
480
481 *5.2. RNA isolation and sequencing*.
482
483 For RNA isolation of a given sample pellet stored in RNAlater (see Section 5.1), we centrifuged
484 for 5 minutes at 1,500 x g, and after decanting of the supernatant, we resuspended the pellet in

485 200 μL of nuclease-free water followed by addition of 750 μL of BashingBead buffer from the
486 Direct-Zol RNA MicroPrep RNA kit (Zymo Research). We then transferred our 950 μL of
487 suspension into BashingBead tubes, vortexed for 20 minutes, and centrifuged for 5 minutes at
488 12,000 x g. We took ~600 μL of supernatant from the top layer being careful not to touch the
489 pellet, transferred to a new tube, and mixed with an equal volume of 100% ethanol. This mixture
490 was used as input to the rest of the kit, following the manufacturer's instructions. RNA library
491 preparation and sequencing of ~20M 150 bp paired-end reads per sample was carried out by
492 Novogene, Inc.
493
494 *5.3. Genome re-annotation and ortholog calls for* K. marxianus *and* K. lactis
495
496 To generate a reference genome for *K. marxianus* strain Km31, we proceeded as follows. We
497 first used bwa (bio-bwa.sourceforge.net) to map the RNA-seq reads from the *K. marxianus*
498 cultures of the training batch of our transcriptional profiling (see Section 5.1) to a published
499 genome of the DMB1 strain of *K. marxianus* (Suzuki et al. 2014). We then used these
500 mappings, requiring mapping qualities of >20 and mapped lengths >20, to call single-nucleotide
501 polymorphisms with BCFtools (Li 2011; Danecek et al. 2021) mpileup with the -I flag to ignore
502 indels, and filtering for exclusion of variants with: a depth of less than 10 reads, or more than
503 10% of reads with a mapping quality of 0 supporting the variant. At the called polymorphisms,
504 we computationally introduced Km31 alleles into the DMB1 genome sequence with the
505 consensus tool in BCFTools to generate a first-pass pseudogenome for Km31. Next, we used
506 this pseudogenome and the training batch RNA-seq reads as input into transcript assembly with
507 Trinity (Grabherr et al. 2011) and annotation with PASA (Haas et al. 2003), and we took the
508 output as our final annotated genome for *K. marxianus* Km31 (Supplementary Data 1). We
509 generated a reference genome for *K. lactis* strain yLB72 as above, except that we used RNA-

510 seq from the *K. lactis* training batch and the published genome of *K. lactis* strain NRRL Y-1140
511 (Dujon et al. 2004) (Supplementary Data 1). To call orthologous gene pairs between *K.*
512 *marxianus* Km31 and *K. lactis* yLB72, we used our final annotations from the two species as
513 input into OrthoFinder (Emms and Kelly 2019). We retained all one-to-one orthology calls, *i.e.*
514 cases in which a single annotated gene from *K. marxianus* was called orthologous to a single
515 gene from *K. lactis*. We also sought to valorize cases in which OrthoFinder, with the goal of
516 finding orthologs in species B of a gene $G_{1,A}$ in species A, did not distinguish between *n*
517 candidate orthologous genes $G_{1...n,B}$ from species B. We evaluated each candidate $G_{i,B}$ for
518 synteny: if the next gene up- or downstream of $G_{i,B}$ in the species B genome was called a one-
519 to-one ortholog of the up- or downstream genes of $G_{1,A}$ in the species A genome, we retained
520 this case as a bona fide ortholog pair. Our final set comprised 3578 ortholog pairs between *K.*
521 *marxianus* Km31 and *K. lactis* yLB72. *Saccharomyces* orthologs were taken from (Sorrells et al.
522 2018).
523
524 *5.4. Production-batch transcriptional profiling data analysis*
525
526 For comprehensive transcriptional profiling analyses in Figure 2 and Figure S5, we used our
527 RNA-seq data from cultures of *K. marxianus* Km31 and *K. lactis* yLB72 from production batches
528 (see Section 5.1) as follows. For the RNA-seq reads from a given replicate of a given species
529 (*K. marxianus* or *K. lactis*) and condition (39°C, ethanol, caffeine, MMS, or unstressed control),
530 we mapped to the respective pseudogenome (see Section 5.3) using the TopHat (Kim et al.
531 2013) aligner with default parameters, and we used these mappings as input into HTseq
532 (Anders et al. 2015) to generate raw read counts per gene. For a given species and stress, we
533 used the counts from all replicates from the stress, and the unstressed control from the
534 respective batch, as input into edgeR (Robinson et al. 2010) to generate normalized counts per

535 gene and replicate and a normalized $\log_2$ fold-change between stress and control conditions as
536 an average across replicates. Principal component analysis (PCA) of the RNA-seq data was
537 done using the PCA module of scikit-learn (Pedregosa et al. 2011). The edgeR normalized
538 counts for each gene in each treatment were scaled to unit variance with scikit-learn before
539 running the PCA. Principal component weights are listed in Supplementary Table 2.
540
541 To identify Gene Ontology (GO) terms associated with principal components, we assigned GO
542 annotations to each gene using the *K. lactis* annotation from FungiDB
543 (https://fungidb.org/fungidb/app/record/dataset/NCBITAXON_284590). For each GO term with
544 at least three annotated genes in our dataset, we tested whether genes within that term had
545 significantly extreme PC1 or PC2 weights compared to the null expectation. We used a
546 bootstrap resampling approach to generate the null distribution: for each GO term containing n
547 genes, we randomly sampled n genes (with replacement) from the complete set of genes and
548 calculated the mean PC weight of this random sample. This process was repeated 10,000 times
549 to construct an empirical null distribution representing the expected mean PC weight for a
550 random gene set of that size. We calculated two-tailed *p*-values by comparing the observed
551 mean PC weight to this null distribution. To correct for multiple testing across all GO terms, we
552 applied the Benjamini-Hochberg false discovery rate (FDR) correction using the multipletests
553 function from statsmodels (Seabold and Perktold 2010), with a significance threshold of $\alpha =$
554 0.05.
555
556 To identify gene groups in which a given stress treatment evoked significant gene induction or
557 repression relative to the unstressed control, in *K. marxianus* Km31 and separately in *K. lactis*
558 yLB72 (Table S3), we carried out resampling-based enrichment tests as follows. We used
559 assignments of each gene to Gene Ontology terms for *K. lactis* from FungiDB

560 (https://fungidb.org/fungidb/app/record/dataset/NCBITAXON_284590), and we propagated
561 these assignments to orthologous genes in *K. marxianus* from Section 5.3. For a given species
562 and stress, for each Gene Ontology term in turn, for the n genes from the term with expression
563 observations, if $n > 5$ we collated the genes' $\log_2$ fold-changes in expression between stress
564 and the untreated control, each of which was a mean across replicates (from edgeR; see
565 above). We took the mean of these fold-changes across the genes of the term, m_{true}. We then
566 picked n random genes from the genome and tabulated their mean expression fold-change for
567 the respective species and stress, $m_{resample}$. Repeating the latter resampling 1000 times, we
568 calculated an empirical p-value reporting the enrichment of the magnitude of expression fold-
569 change as the percentage of resamples for which $(|m_{resample}| \geq (|m_{true}|)$. We used the p-values for
570 all terms tested for the respective species and stress as input into multiple testing correction
571 with the Benjamini-Hochberg method.
572
573 **6. Molecular-evolution analyses**
574
575 *6.1. Gene duplication analysis*
576
577 As input to the gene family copy number analysis for Figure 3 and Table S4, genomes and
578 annotations were as follows. For *K. marxianus* and *K. lactis* we used the reference genomes
579 and annotations of Km31 and yLB72, respectively, from Section 5.3. For *K. wickerhamii* we
580 used the reference genome and annotation of UCD54-210 from (Sorrells et al. 2018). For the *K.*
581 *nonfermentans* strain NRRL Y-27343 (GenBank acc. no. GCA_030569915.1), the *K. aestuarii*
582 strain NRRL YB-4510 (GenBank acc. no. GCA_003707555.2), the *K. siamensis* strain CBS
583 10860 (GenBank acc. no. GCA_030579315.1), and the *K. dobzhanskii* strain NRRL Y-1974
584 (GenBank acc. no. GCA_003705805.3) we used YGAP (Proux-Wéra et al. 2012) to generate *de*

585 *novo* annotations from previously generated reference genomes (Supplementary Data 1). As
586 outgroup species we used the genomes and annotations for *Ashbya aceri* (GenBank acc. no.
587 GCA_000412225.2), *Candida albicans* (GenBank acc. no. GCA_000182965.3), *Eremothecium*
588 *cymbalariae* (GenBank acc. no. GCA_000235365.1), *Eremothecium gossypii* (GenBank acc.
589 no. GCA_000091025.4), *Eremothecium sinecaudum* (GenBank acc. no. GCA_001548555.1),
590 *Lachancea dasiensis* (GenBank acc. no. GCA_900074725.1), *Lachancea fantastica* (GenBank
591 acc. no. GCA_900074735.1), *Lachancea fermentatii* (GenBank acc. no. GCA_900074765.1),
592 *Lachancea lanzarotensis* (GenBank acc. no. GCA_000938715.1), *Lachancea meyersii*
593 (GenBank acc. no. GCA_900074715.1), *Lachancea mirantina* (GenBank acc. no.
594 GCA_900074745.1), *Lachancea nothofagi* (GenBank acc. no. GCA_900074755.1), *Lachancea*
595 *quebecensis* (GenBank acc. no. GCA_002900925.1), *Lachancea thermotolerans* (GenBank
596 acc. no. GCA_000142805.1), and *Saccharomyces cerevisiae* (GenBank acc. no.
597 GCA_000146045.2). We used proteomes for each species as input into OrthoFinder (Emms
598 and Kelly 2019) to find groups of orthologs. We then used the time-weighted tree and gene
599 family counts from OthoFinder as input into CAFE 5 (Mendes et al. 2021) to analyze gene
600 family contractions and expansions. Gene families that did not exist at the root were included.
601 We adjusted *p*-values with the Benjamini-Hochberg test to correct for multiple testing. We used
602 CafePlotter (https://github.com/moshi4/CafePlotter) to visualize significant gene family
603 expansions and contractions on the phylogenetic tree. We then assigned functions for each
604 gene family by running all proteins from all species for each gene family through InterProScan
605 (Jones et al. 2014). We took the most common InterPro term in the output of the scan as the
606 inferred function of the gene family in Figure 3.
607
608 *6.2. Phylogenetic analyses*
609

610 As input to phylogenetic analyses of one strain for each *Kluyveromyces* species for Figure 4
611 and Table S5, we used genomes and annotations for the *Kluyveromyces* genus excluding *K.*
612 *siamensis* as in Section 7.1. We called orthologs across these six species with OrthoFinder
613 (Emms and Kelly 2019), and for a given orthogroup we generated protein alignments using
614 MUSCLE (Edgar 2004). We then used this alignment and its nucleotide sequences as input into
615 PAL2NAL (Suyama et al. 2006) using the -nogap flag to remove gaps to generate a codon
616 alignment. For each orthogroup in turn, we ran the PAML (Yang 2007) codeml package run
617 using the branch-site model (Zhang et al. 2005) to evaluate the fit to the sequence data of a
618 model in which regions of the coding sequence are under positive selection for amino acid
619 variation along a focal (foreground) branch and not on any other branch from the tree (the
620 background branches). We used as input to PAML the codon alignment for the orthogroup and
621 the topology of the phylogenetic tree relating the species (as in Figure 1A), omitting *K.*
622 *siamensis*. The calculation proceeded in two steps. First, we fit a null model in which all
623 branches of the tree had the same protein evolutionary rate (PAML codeml run with options
624 verbose = 1, seqtype = 1, clock = 0, model = 2, NSsites = 2, fix_omega = 1, omega = 1),
625 yielding a protein evolutionary rate and log-likelihood of the data under the model. Next, we re-
626 fit using an alternative model in which the branch leading to *K. marxianus* was the foreground
627 (PAML codeml run with options verbose = 1, seqtype = 1, clock = 0, model = 2, NSsites = 2,
628 fix_omega = 0). Outputs were the codon sites assigned to the class with an excess of amino
629 acid changes in *K. marxianus* relative to silent changes, and their posterior probabilities under
630 Bayes Empirical Bayes inference (Yang et al. 2005); the codon sites assigned to classes with
631 relaxed and purifying selection, respectively, along the background lineages, and their posterior
632 probabilities; and the log-likelihood of the data under the model. The *p*-value assessing the
633 improvement in fit to the data of the alternative model relative to the null was calculated from a
634 likelihood ratio test (a χ^2 test on twice the difference of likelihoods with one degree of freedom).

635 *p*-values were corrected for multiple testing by the Benjamini/Hochberg method using the
636 multipletests function from statsmodels (Seabold and Perktold 2010). In Figure 5A we tabulated
637 all genes with adjusted $p < 0.05$ from this test that had more than 1% of codon sites assigned to
638 the class with an excess of amino acid changes in *K. marxianus* (proportion in class 2a > 0.01
639 in codeml). Separately, we repeated the above analyses using the branch leading to *K. lactis* as
640 the foreground. Conservation of sites called significant by PAML using *K. marxianus* as the
641 foreground were calculated using Biopython (Cock et al. 2009) to count the conservation of
642 each site among our sequenced *K. marxianus* strains (Table S1 and see Section 7.3).
643 *Saccharomyces* orthologs were taken from (Sorrells et al. 2018).
644
645 To complement our PAML analysis, we also ran HyPhy aBSREL (Pond et al. 2005) on the
646 PAL2NAL alignments across *Kluyveromyces* orthologs as described above. aBSREL uses a
647 branch-site model similar to PAML to detect positive selection on individual branches. We ran
648 aBSREL indicating either *K. marxianus* or *K. lactis* as the foreground. We pulled the *p*-value as
649 calculated by aBSREL for each orthogroup and corrected for multiple testing by the
650 Benjamini/Hochberg method using the multipletests function from statsmodels (Seabold and
651 Perktold 2010). *Saccharomyces* orthologs were taken from (Sorrells et al. 2018). We compared
652 HyPhy aBSREL hits against PAML branch-site hits to confidently list genes acted on positive
653 selection in *K. marxianus.*
654
655 *6.3. Resequencing and analysis of* K. marxianus *and* K. lactis *population genomes*
656
657 For resequencing, we cultured, in 10 mL of liquid YPD, each of five *K. marxianus* strains of non-
658 domesticated provenance beside Km31, and two non-domesticated *K. lactis* strains beside
659 yLB72 (Table S1), and we isolated DNA from each with the Quick-DNA™ Fungal/Bacterial

660 Miniprep Kit (Zymo Research). Genomic DNA sequencing libraries were made and sequenced
661 on an Illumina NovaSeq 6000 by the UC Berkeley Vincent J. Coates Genomic Sequencing
662 Laboratory. We used Minimap2 (Li 2018) to map the *K. marxianus* reads to our *K. marxianus*
663 Km31 reference genome (see Section 5.3). Separately, we mapped the reads from each of 41
664 *K. lactis* strains previously sequenced (Friedrich et al. 2023), and those from our two sequenced
665 *K. lactis* strains, to our *K. lactis* yLB72 reference genome. For each strain we sorted mappings
666 using the SAMtools fixmate and markdup functions, and we called single-nucleotide
667 polymorphisms and generated a pseudogenome as in Section 5.3, except the variant filtering
668 included only a quality score (QUAL) threshold of above 20. Called variants are provided in
669 Supplementary Data 2. For each pair of orthologous genes between *K. marxianus* and *K. lactis*
670 from section 5.3, we aligned the nucleotide sequences of the coding regions from all strains of
671 the two species using MUSCLE (Edgar 2004). We then used this alignment and its associated
672 nucleotide sequences as input into PAL2NAL (Suyama et al. 2006) using the -nogap flag to
673 remove gaps to generate a codon alignment.
674
675 For Figure 4D,E and Table S6,7 we carried out per-gene McDonald-Kreitman analyses
676 (McDonald and Kreitman 1991) on the PAL2NAL alignments using the mktest function from the
677 CodonAlignment class from Biopython. *p*-values were corrected for multiple testing by the
678 Benjamini/Hochberg method using the multipletests function from statsmodels (Seabold and
679 Perktold 2010). For a given gene, this approach tabulates the rates of non-synonymous
680 polymorphism within species (P_n), synonymous polymorphism (P_s), non-synonymous
681 divergence between species (D_n), and synonymous divergence (D_s), and the neutrality index NI
682 = $(P_n/P_s)/(D_n/D_s)$. Because low counts of amino acid polymorphism can drive significance in the
683 McDonald-Kreitman framework, we developed a filtering scheme in which we tabulated the

684 average *a* and standard deviation σ of P_n/P_s across all genes for which we had sequence data.
685 We only retained for Figure 5D and NI genomic analyses those genes with $P_n/P_s > a - 1.5\sigma$.
686
687 For genomic analyses of P_n/P_s and D_n/D_s in the McDonald-Kreitman framework in Table S7, we
688 used Gene Ontology term assignments and resampling (using 10,000 resamples) as in Section
689 5.4, except that we only included Gene Ontology terms containing 10 genes or more after
690 filtering, and for P_n/P_s we tested for enrichment of low values of the metric across a given term.
691 Genomic analyses on NI proceeded as above, except on the filtered dataset of genes with P_n/P_s
692 $> a - 1.5\sigma$ as explained above and testing for enrichment of low values of the metric across the
693 term. *p*-values were corrected for multiple testing by the Benjamini/Hochberg method using the
694 multipletests function from statsmodels (Seabold and Perktold 2010).
695
696 **7. *FAT3* allelic swap mutant generation and cerulenin with oleic acid growth phenotyping**
697
698 To search for *K. marxianus* phenotypes that *FAT3* could be implicated in, we assayed the
699 genus in media with cerulenin and oleate according to the Tecan 96-well plate experimental
700 design described in Section 2, except media contained 500 μM oleate (Sigma) and 0.5% Tween
701 40 (Sigma) with or without 25 μM cerulenin (from 20 mg/mL stock in ethanol; Cayman
702 Chemical). We gathered final OD readings as an efficiency read out and tested for significance
703 with a Mann-Whitney U test from SciPy.
704
705 We made Δ*DNL4* mutants of *K. marxianus* Km31 and *K.* lactis yLB72 as described in (Rajkumar
706 and Morrissey 2022). We then made CRISPR/Cas9 plasmids to target each species' *FAT3*
707 allele in the Δ*DNL4* mutants. We chose gRNAs for *K. marxianus* based on the dataset

708 generated from sgRNAcas9 (Xie et al. 2014) in (ref above), and chose gRNAs for *K. lactis* using
709 CCTop (Stemmer et al. 2015). Alongside the CRISPR plasmids we made repair fragments by
710 PCR, with 1kb regions of the species' native gDNA flanking the opposite species' *FAT3* allele to
711 allow for homologous recombination. We transformed the stitched amplicon and CRISPR
712 plasmid into *K. marxianus* and *K. lactis* ΔDNL4 mutants by electroporation (Hsu et al. 2021) and
713 selected transformants by plating onto agar plates with 50 µg/mL of Hygromycin B (Invitrogen
714 by Thermo Fisher Scientific). Nanopore sequencing of the ~4kb region including the swapped
715 allele, the flanking homology arms, and the genomic region around it confirmed positive swaps.
716
717 We assayed the growth in cerulenin and oleate of *K. marxianus* Km31, *K. lactis* yLB72, and the
718 *FAT3* allelic swaps in either direction to assess the efficiency of *FAT3*. We streaked out *K. lactis*
719 (Δ*DNL4*), *K. marxianus* (Δ*DNL4*), two replicate clones of *K. lactis* (Δ*DNL4*; *ΔFAT3*::*K. marxianus*
720 *FAT3*), and two replicate clones of *K. marxianus* (Δ*DNL4*; *ΔFAT3*::*K. lactis FAT3*) from -80°C
721 frozen stocks onto YPD plates and incubated them at 28°C for 2 days to allow for colony
722 formation. We inoculated one colony per strain into 3 ml of YPD and incubated at 28°C
723 overnight. We then back-diluted the cultures into 10 ml of YPD to achieve 0.05 OD_{600}/ml. We let
724 the cultures grow at 28°C with shaking for 6-7 hours to reach the early log-phase (0.2-0.4
725 OD_{600}). We back-diluted the early log-phase cultures into 3 ml of oleate media (YPD with 0.5%
726 Tween-40, 0.7% KH_2PO_4, 25 µM cerulenin, and 500 µM oleate) or control media (YPD with
727 0.5% Tween-40 and 0.7% KH_2PO_4). We incubated the cultures at 28°C with shaking for 24
728 hours and recorded the OD600 as the final readout for cell density. We ran three technical
729 replicates per assay with at least 4 independent assays, from streaking to measurements of
730 OD_{600}.
731

Acknowledgements

The authors thank Adam Arkin for his generosity with computational resources; Jamie Cate and Alexander Johnson for strains; and Jacob Steenwyk, Jeremy Roop, and members of the Brem lab for helpful discussions. This work was supported by NIH 2R01GM120430 to R.B.B. and an NSF Graduate Research Fellowship to A.D.

Data Availability

Raw RNA-seq reads are available on GEO under series record GSE302710. Genome sequencing data is available on SRA with accession PRJNA1291269.

743 ## Figure captions

744

745 **Figure 1. *K. marxianus* shows a unique heat and chemical stress resistance phenotype**
746 **that only manifests during its growing stage. A,** Each panel reports the results of a growth
747 profiling experiment across the genus in the indicated condition. In a given panel, each trace
748 reports a growth timecourse for the indicated species, with the solid line displaying the mean
749 across technical replicates and the faint outline showing the standard error. OD, optical density.
750 **B**, Each row of serial dilutions reports viability after heat shock of a sample of the indicated
751 species in the indicated liquid growth phase. Schematics show experimental design of heat-
752 shocking cells either at stationary or exponential phase. See Figure S3B for unshocked
753 controls.

754

755 **Figure 2. Overall regulatory volatility by *K. lactis* relative to *K. marxianus* and distinct**
756 **regulatory patterns that separate the species. A,** Shown are results of principal component
757 (PC) analysis of *K. lactis* (orange) and *K. marxianus* (blue) transcriptomes across treatments.
758 Each point reports the values of the top two PCs for one replicate transcriptome of the indicated
759 species subjected to the indicated stress treatment or a 28°C unstressed control. In a given axis
760 label, the value in parentheses reports the proportion of variance across the set of
761 transcriptomes explained by the indicated PC. **B,C,** Each panel reports the effect of stress on
762 expression of genes of the indicated Gene Ontology term in *K. marxianus* and *K. lactis*. In a
763 given panel each cell reports, for the indicated gene, normalized read counts in the indicated
764 stress treatment and an unstressed control, standardized by row using z-scores.

765

766 **Figure 3. Gene family expansions in *K. marxianus*. A,** The color of each cell reports the
767 change in copy number, in the indicated genome of a *Kluyveromyces* or related species relative
768 to its parent node in the phylogeny, of a gene family (orthogroup, OG) with significant expansion
769 or contraction in the *K. marxianus* genome as inferred with CAFE 5. Transposon and reverse
770 transcriptase families with significant change values are not listed, see Table S4 for results for
771 all analyzed orthogroups. **B,** Each panel reports the positions of genes in the *K. marxianus*
772 genome in one indicated gene family from **A** with annotations as major facilitator superfamily
773 (MFS) transporters. In a given panel, each row represents a chromosome in *K. marxianus*, and
774 each arrow reports a gene of the respective family and its transcriptional direction (not to scale).
775
776 **Figure 4. Phylogenetic analysis reveals accelerated amino acid evolution in *K.***
777 ***marxianus*. A,** Each bar shows the number of genes with significant evidence for positive
778 selection on the indicated lineage (adjusted $p > 0.05$) in a phylogenetic PAML branch-site test
779 and HyPhy aBSREL test. **B,** Venn diagram of genes called significant for positive selection in *K.*
780 *marxianus* in PAML branch-site and HyPhy aBSREL tests. **C,** Amino acid alignment of an
781 example region of a gene, KLLA0B02607g (*S. cerevisiae* ortholog FAT3/*YKL187C*), with
782 evidence for positive selection from the PAML test in **A** and from the McDonald-Kreitman test of
783 **D**. Highlighted columns show inferred targets of positive selection from the PAML test in **B**.
784 Highlighting indicates the amino acid's BLOSUM 62 matrix score with respect to the consensus
785 amino acid. **D,** Each row reports a gene with evidence for adaptive protein variation between *K.*
786 *marxianus* and *K. lactis* according to the McDonald-Kreitman test (adjusted $p < 0.05$). NI,
787 neutrality index. Genes highlighted in green also showed significant evidence for positive
788 selection on the branch leading to *K. marxianus* in phylogenetic tests (PAML branch-site
789 adjusted $p < 0.05$ and HyPhy aBSREL adjusted $p < 0.05$; Table S5). Not shown here are genes
790 with p-value < 0.05 as a product of exceptionally high or low amino acid polymorphism; see

791 Methods and Table S5. **E,** Each bar shows the cell density (as OD_{600}/mL after 24 hours of *K.*

792 *marxianus* and *K. lactis* and their respective FAT3 allele swaps in media with cerulenin and

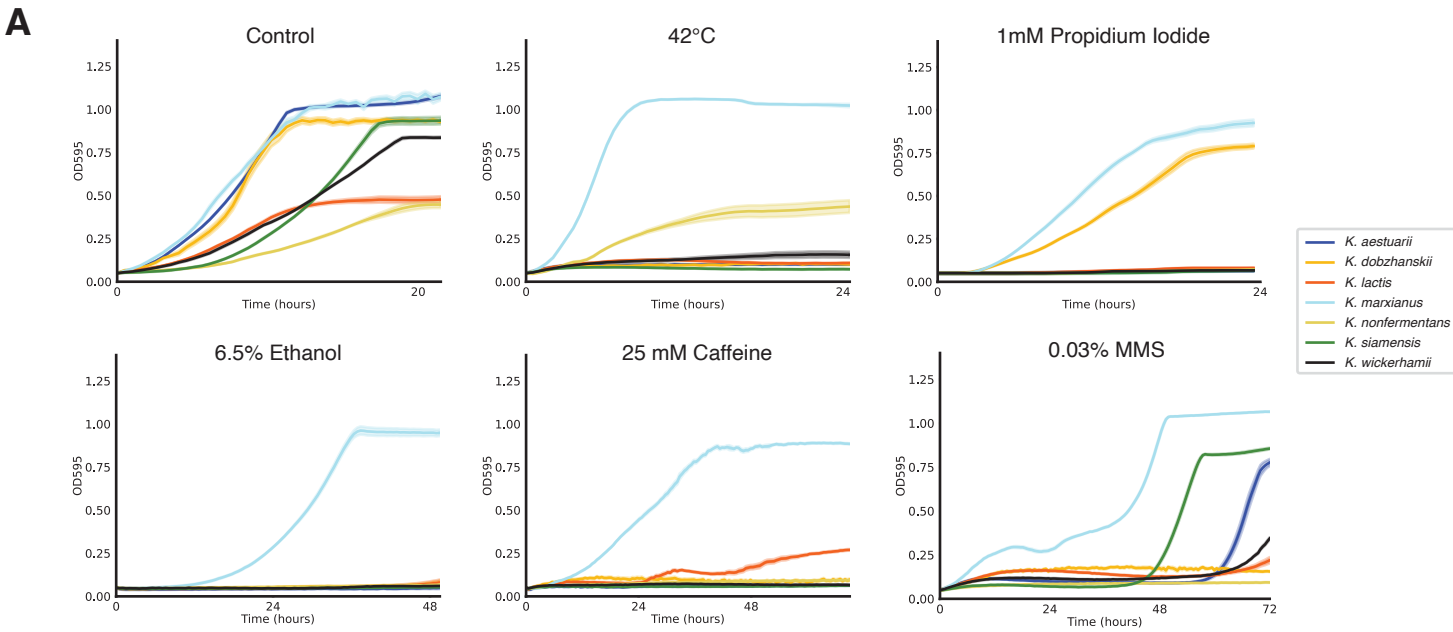
793 oleate, where cerulenin blocks lipid uptake and oleate is the only available lipid for utilization.

794 The starred comparison between *K. marxianus* and its klFAT3 mutant indicate a statistically

795 significant difference ($p = 0.0223$). **F**, Each row reports a gene ontology term with significant

796 enrichment for low NI (adjusted $p < 0.05$) in the McDonald-Kreitman test (Table S5).

Figure 1


A
Control
42°C
1mM Propidium Iodide
6.5% Ethanol
25 mM Caffeine
0.03% MMS
OD595
Time (hours)
K. aestuarii
K. dobzhanskii
K. lactis
K. marxianus
K. nonfermentans
K. siamensis
K. wickerhamii

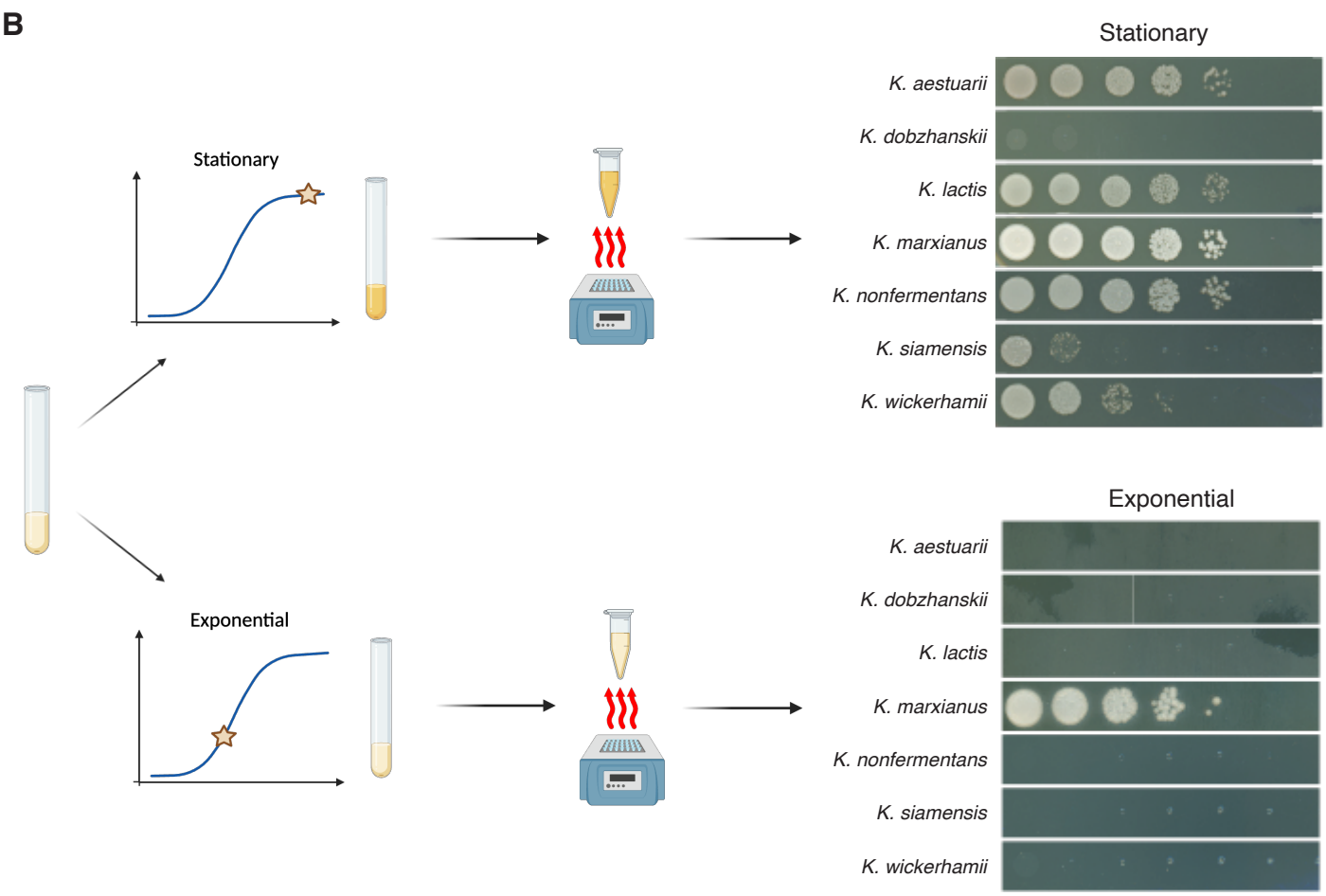

B
Stationary
Exponential
K. aestuarii
K. dobzhanskii
K. lactis
K. marxianus
K. nonfermentans
K. siamensis
K. wickerhamii

Figure 2

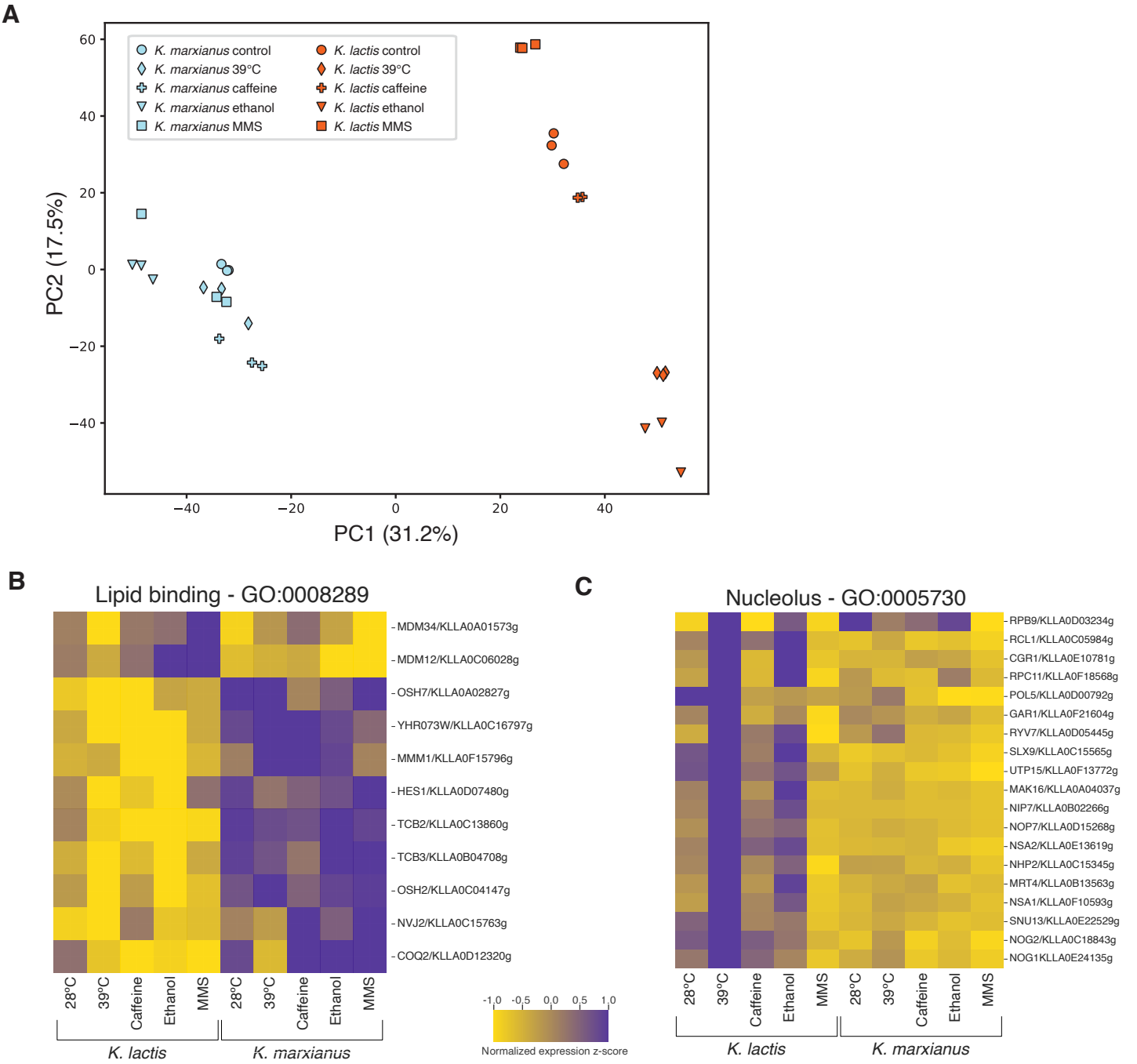
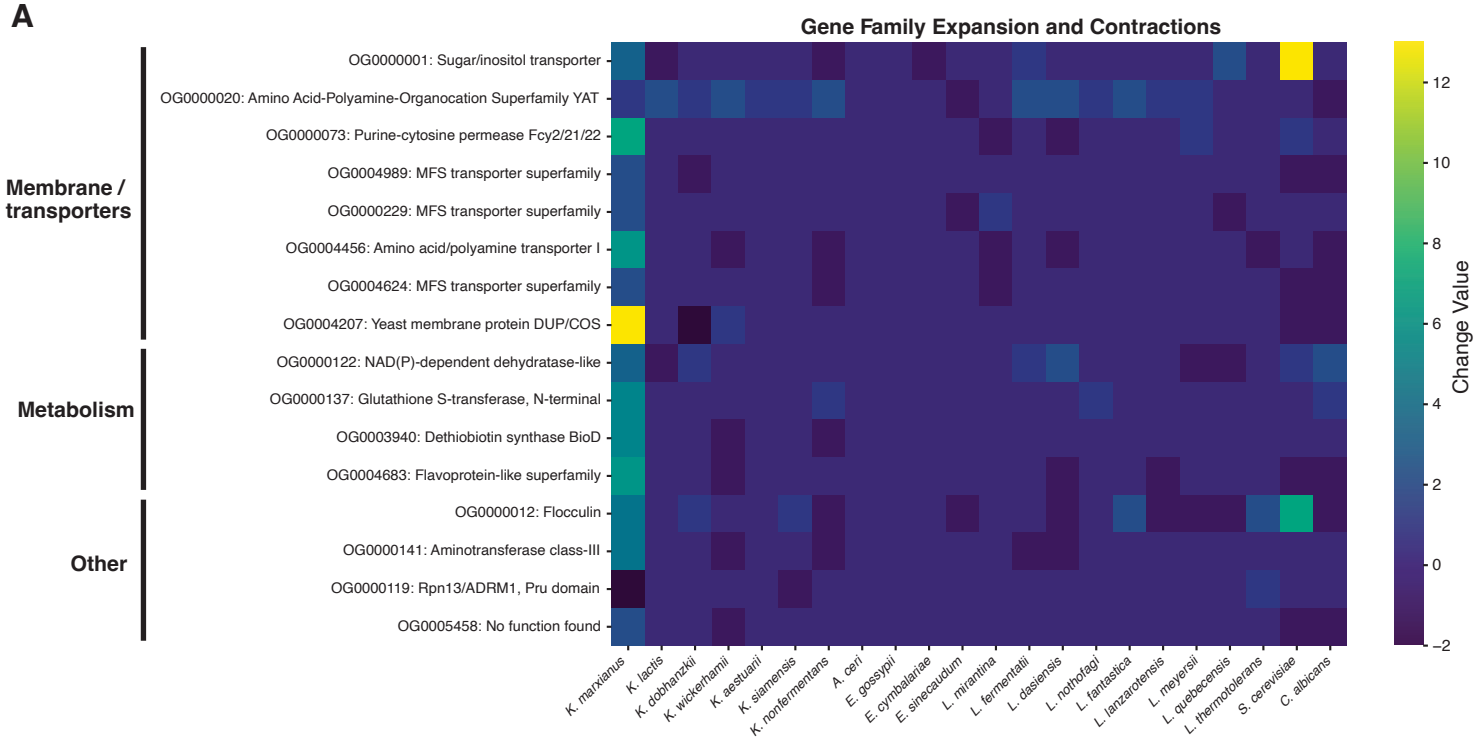

A
K. marxianus control
K. marxianus 39°C
K. marxianus caffeine
K. marxianus ethanol
K. marxianus MMS
K. lactis control
K. lactis 39°C
K. lactis caffeine
K. lactis ethanol
K. lactis MMS
PC2 (17.5%)
PC1 (31.2%)
60
40
20
0
−20
−40
−40
−20
0
20
40
B
Lipid binding - GO:0008289
MDM34/KLLA0A01573g
MDM12/KLLA0C06028g
OSH7/KLLA0A02827g
YHR073W/KLLA0C16797g
MMM1/KLLA0F15796g
HES1/KLLA0D07480g
TCB2/KLLA0C13860g
TCB3/KLLA0B04708g
OSH2/KLLA0C04147g
NVJ2/KLLA0C15763g
COQ2/KLLA0D12320g
28°C
39°C
Caffeine
Ethanol
MMS
K. lactis
K. marxianus
-1.0
-0.5
0.0
0.5
1.0
Normalized expression z-score
C
Nucleolus - GO:0005730
RPB9/KLLA0D03234g
RCL1/KLLA0C05984g
CGR1/KLLA0E10781g
RPC11/KLLA0F18568g
POL5/KLLA0D00792g
GAR1/KLLA0F21604g
RYV7/KLLA0D05445g
SLX9/KLLA0C15565g
UTP15/KLLA0F13772g
MAK16/KLLA0A04037g
NIP7/KLLA0B02266g
NOP7/KLLA0D15268g
NSA2/KLLA0E13619g
NHP2/KLLA0C15345g
MRT4/KLLA0B13563g
NSA1/KLLA0F10593g
SNU13/KLLA0E22529g
NOG2/KLLA0C18843g
NOG1KLLA0E24135g
28°C
39°C
Caffeine
Ethanol
MMS
K. lactis
K. marxianus

Figure 3

A


Gene Family Expansion and Contractions
Membrane / transporters
Metabolism
Other
OG0000001: Sugar/inositol transporter
OG0000020: Amino Acid-Polyamine-Organocation Superfamily YAT
OG0000073: Purine-cytosine permease Fcy2/21/22
OG0004989: MFS transporter superfamily
OG0000229: MFS transporter superfamily
OG0004456: Amino acid/polyamine transporter I
OG0004624: MFS transporter superfamily
OG0004207: Yeast membrane protein DUP/COS
OG0000122: NAD(P)-dependent dehydratase-like
OG0000137: Glutathione S-transferase, N-terminal
OG0003940: Dethiobiotin synthase BioD
OG0004683: Flavoprotein-like superfamily
OG0000012: Flocculin
OG0000141: Aminotransferase class-III
OG0000119: Rpn13/ADRM1, Pru domain
OG0005458: No function found
K. marxianus
K. lactis
K. dobhanzkii
K. wickerhamii
K. aestuanii
K. siamensis
K. nonfermentans
A. ceri
E. gossypii
E. cymbalariae
E. sinecaudum
L. mirantina
L. fermentati
L. dasiensis
L. nothofagi
L. fantastica
L. lanzarotensis
L. meyersii
L. quebecensis
L. thermotolerans
S. cerevisiae
C. albicans
Change Value
12
10
8
6
4
2
0
−2

B

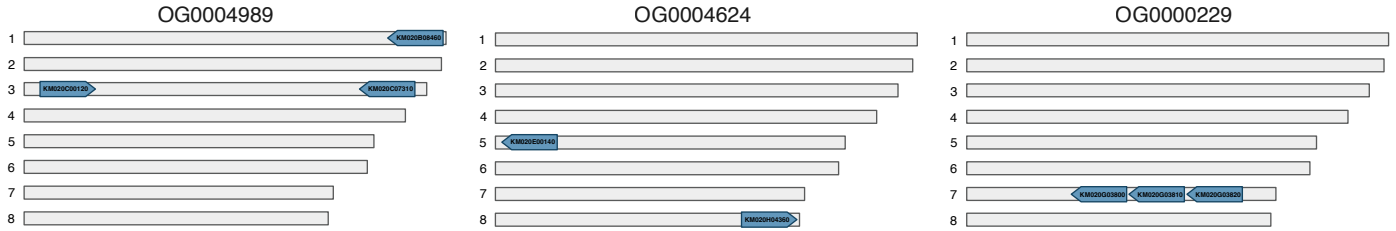
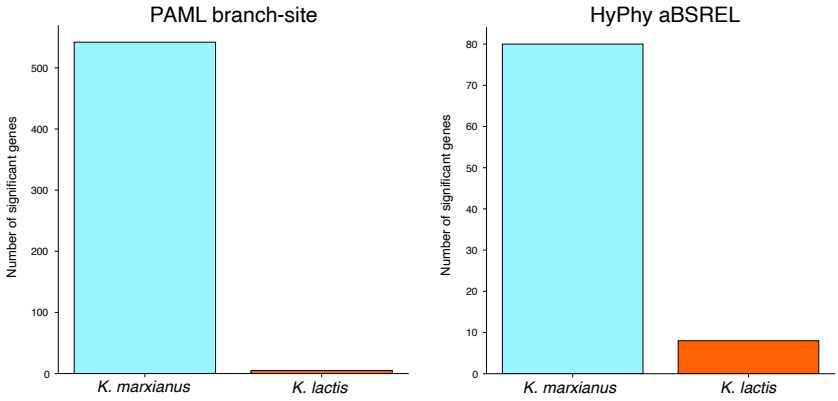
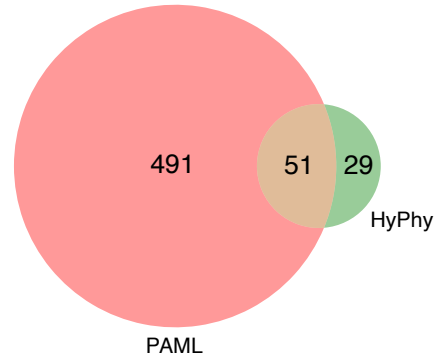

OG0004989
OG0004624
OG0000229
1
2
3
4
5
6
7
8

Figure 4

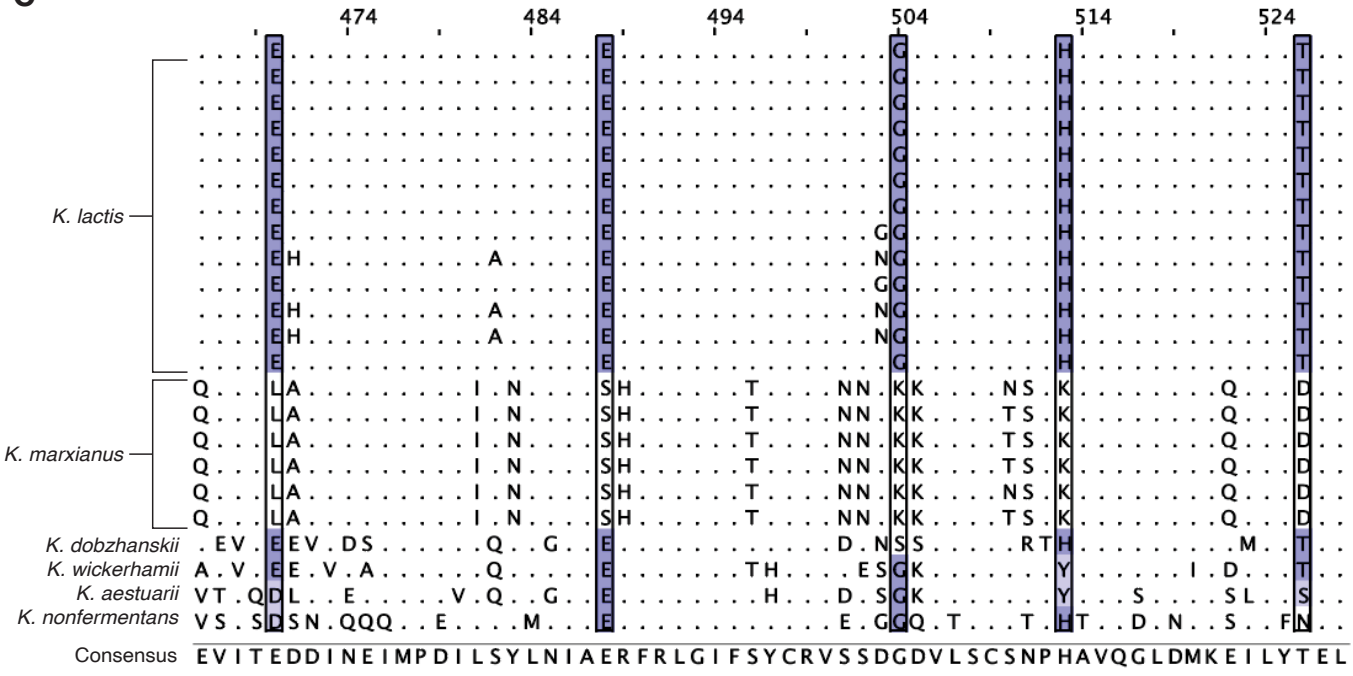
A



B



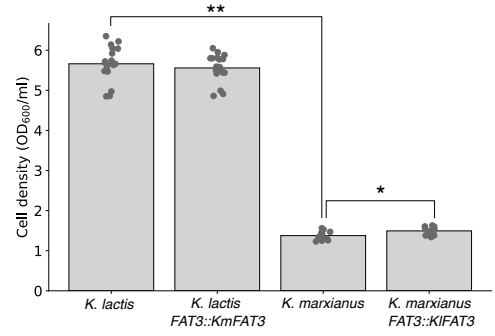
C



D

K. lactis, K. marxianus genes	NI	*Saccharomyces* gene	Standard Name	Function
KLLA0E09329g, KM020H00950	0.19	YOR020W-A	MCO10	Subunit I of the mitochondrial F1F0 ATP synthase
KLLA0E07569g, KM020D04120	0.26	YKL060C	FBA1	Fructose 1,6-bisphosphate aldolase
KLLA0D10780g, KM020E01340	0.32	YJR118C	ILM1	Unknown function, possible mitochondrial involvement
KLLA0C12133g, KM020H02920	0.39	YNL284C	MRPL10	Mitochondrial ribosomal protein
KLLA0A10175g, KM020A03410	0.47	YMR313C	TGL3	Triacylglycerol lipase and LPE acyltransferase
KLLA0F19162g, KM020D01640	0.49	YDR111C	ALT2	Catalytically inactive alanine transaminase
KLLA0B04994g, KM020F01720	0.50	YLR443W	ECM7	Putative integral membrane protein, calcium uptake
KLLA0F00528g, KM020F00240	0.51	YAL061W	BDH2	Putative medium-chain alcohol dehydrogenase
KLLA0B02607g, KM020D05580	0.59	YKL187C	FAT3	Fatty acid uptake
KLLA0D16236g, KM020F04880	0.63	YGR130C	N/A	Component of the eisosome
KLLA0F26048g, KM020E04040	0.68	YMR066W	SOV1	Mitochondrial protein of unknown function

E



F

GO Term	GO Term Function	NI mean
GO:0006096	glycolytic process	0.7180
GO:0015986	proton motive force-driven ATP synthesis	0.7693
GO:0055085	transmembrane transport	1.0418
GO:0005975	carbohydrate metabolic process	0.9794
GO:0006099	tricarboxylic acid cycle	0.8911

797 ## References

798 Al-Qaysi SAS, Abdullah NM, Jaffer MR, Abbas ZA. 2021. Biological Control of Phytopathogenic
799 Fungi by Kluyveromyces marxianus and Torulaspora delbrueckii Isolated from Iraqi Date
800 Vinegar. *J Pure Appl Microbiol.* 15:300–311.

801 Am-In S, Yongmanitchai W, Limtong S. 2008. Kluyveromyces siamensis sp. nov., an
802 ascomycetous yeast isolated from water in a mangrove forest in Ranong Province,
803 Thailand. *FEMS Yeast Research* 8:823–828.

804 Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput
805 sequencing data. *Bioinformatics* 31:166–169.

806 Anderson PJ, McNeil K, Watson K. 1986. High-Efficiency Carbohydrate Fermentation to Ethanol
807 at Temperatures above 40°C by Kluyveromyces marxianus var. marxianus Isolated from
808 Sugar Mills. *Applied and Environmental Microbiology* 51:1314–1320.

809 Araujo FV, Hagler AN. 2011. Kluyveromyces aestuarii, a potential environmental quality
810 indicator yeast for mangroves in the State of Rio de Janeiro, Brazil. *Braz J Microbiol*
811 42:954–958.

812 Avchar R, Lanjekar V, Dhakephalkar PK, Dagar SS, Baghela A. 2022. Compost as an untapped
813 niche for thermotolerant yeasts capable of high-temperature ethanol production. *Letters*
814 *in Applied Microbiology* 74:109–121.

815 Baptista M, Cunha JT, Domingues L. 2021. Establishment of Kluyveromyces marxianus as a
816 Microbial Cell Factory for Lignocellulosic Processes: Production of High Value Furan
817 Derivatives. *Journal of Fungi* 7:1047.

818 Baptista M, Domingues L. 2022. Kluyveromyces marxianus as a microbial cell factory for
819 lignocellulosic biomass valorisation. *Biotechnology Advances* 60:108027.

820 Burton WG. 1982. Post-harvest Physiology of Food Crops. Longman

821 Cernak P, Estrela R, Poddar S, Skerker JM, Cheng Y-F, Carlson AK, Chen B, Glynn VM, Furlan
822 M, Ryan OW, et al. 2018. Engineering Kluyveromyces marxianus as a Robust Synthetic
823 Biology Platform Host. *mBio* 9:10.1128/mbio.01410-18.

824 Chacón-Vargas K, Torres J, Giles-Gómez M, Escalante A, Gibbons JG. 2020. Genomic profiling
825 of bacterial and fungal communities and their predictive functionality during pulque
826 fermentation by whole-genome shotgun sequencing. *Sci Rep* 10:15115.

827 Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff
828 F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational
829 molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423.

830 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T,
831 McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools.
832 *GigaScience* 10:giab008.

833 Dekker WJC, Ortiz-Merino RA, Kaljouw A, Battjes J, Wiering FW, Mooiman C, Torre P de la,
834 Pronk JT. 2021. Engineering the thermotolerant industrial yeast Kluyveromyces
835 marxianus for anaerobic growth. *Metabolic Engineering* 67:347–364.

836 Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, de Montigny J, Marck
837 C, Neuvéglise C, Talla E, et al. 2004. Genome evolution in yeasts. *Nature* 430:35–44.

838 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
839 throughput. *Nucleic Acids Research* 32:1792–1797.

840 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative
841 genomics. *Genome Biology* 20:238.

842 Falcão Moreira R, Fernandes PA, Moradas-Ferreira P. 1998. Kluyveromyces marxianus
843 flocculence and growth at high temperature is dependent on the presence of the protein
844 p37. *Microbiology* 144:681–688.

845 Ferenci T. 2016. Trade-off Mechanisms Shaping the Diversity of Bacteria. *Trends Microbiol*
846 24:209–223.

847 Flores-Cosío G, García-Béjar JA, Sandoval-Nuñez D, Amaya-Delgado L. 2024. Stress
848 response and adaptation mechanisms in *Kluyveromyces marxianus*. In: Gadd GM,
849 Sariaslani S, editors. Advances in Applied Microbiology. Vol. 126. Advances in Applied
850 Microbiology. Academic Press. p. 27–62. Available from:
851 https://www.sciencedirect.com/science/article/pii/S0065216424000030

852 Friedrich A, Gounot J-S, Tsouris A, Bleykasten C, Freel K, Caradec C, Schacherer J. 2023.
853 Contrasting Genomic Evolution Between Domesticated and Wild Kluyveromyces lactis
854 Yeast Populations. *Genome Biology and Evolution* 15:evad004.

855 Fu X, Li P, Zhang L, Li S. 2019. Understanding the stress responses of Kluyveromyces
856 marxianus after an arrest during high-temperature ethanol fermentation based on
857 integration of RNA-Seq and metabolite data. *Appl Microbiol Biotechnol* 103:2715–2729.

858 Gethins L, Rea MC, Stanton C, Ross RP, Kilcawley K, O'Sullivan M, Crotty S, Morrissey JP.
859 2016. Acquisition of the yeast Kluyveromyces marxianus from unpasteurised milk by a
860 kefir grain enhances kefir quality. *FEMS Microbiology Letters* 363:fnw165.

861 Goodspeed D, Liu JD, Chehab EW, Sheng Z, Francisco M, Kliebenstein DJ, Braam J. 2013.
862 Postharvest Circadian Entrainment Enhances Crop Pest Resistance and Phytochemical
863 Cycling. *Current Biology* 23:1235–1241.

864 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
865 Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-
866 Seq data without a reference genome. *Nat Biotechnol* 29:644–652.

867 Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith Jr RK, Hannick LI, Maiti R, Ronning CM,
868 Rusch DB, Town CD, et al. 2003. Improving the Arabidopsis genome annotation using
869 maximal transcript alignment assemblies. *Nucleic Acids Research* 31:5654–5666.

870 Henderson CM, Block DE. 2014. Examining the Role of Membrane Lipid Composition in
871 Determining the Ethanol Tolerance of Saccharomyces cerevisiae. *Applied and*
872 *Environmental Microbiology* 80:2966–2972.

873 Hittinger CT, Rokas A, Bai F-Y, Boekhout T, Gonçalves P, Jeffries TW, Kominek J, Lachance
874 M-A, Libkind D, Rosa CA, et al. 2015. Genomics and the making of yeast biodiversity.
875 *Current Opinion in Genetics & Development* 35:100–109.

876 Hoshida H, Kagawa S, Ogami K, Akada R. 2020. Anoxia-induced mitophagy in the yeast
877 Kluyveromyces marxianus. *FEMS Yeast Research* 20:foaa057.

878 Hsu P-C, Lu T-C, Hung P-H, Jhou Y-T, Amine AAA, Liao C-W, Leu J-Y. 2021. Plastic Rewiring
879 of Sef1 Transcriptional Networks and the Potential of Nonfunctional Transcription Factor
880 Binding in Facilitating Adaptive Evolution. *Molecular Biology and Evolution* 38:4732–
881 4747.

882 Illarionov A, Lahtvee P-J, Kumar R. 2021. Potassium and Sodium Salt Stress Characterization
883 in the Yeasts Saccharomyces cerevisiae, Kluyveromyces marxianus, and Rhodotorula
884 toruloides. *Applied and Environmental Microbiology* 87:e03100-20.

885 Ishtar Snoek IS, Yde Steensma H. 2006. Why does Kluyveromyces lactis not grow under
886 anaerobic conditions? Comparison of essential anaerobic genes of Saccharomyces
887 cerevisiae with the Kluyveromyces lactis genome. *FEMS Yeast Research* 6:393–403.

888 Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A,
889 Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification.
890 *Bioinformatics* 30:1236–1240.

891 Jung S, Smith JJ, von Haller PD, Dilworth DJ, Sitko KA, Miller LR, Saleem RA, Goodlett DR,
892 Aitchison JD. 2013. Global Analysis of Condition-specific Subcellular Protein Distribution
893 and Abundance*. *Molecular & Cellular Proteomics* 12:1421–1435.

894 Kiers J, Zeeman A-M, Luttik M, Thiele C, Castrillo JI, Steensma HY, Van Dijken JP, Pronk JT.
895 1998. Regulation of alcoholic fermentation in batch and chemostat cultures of
896 Kluyveromyces lactis CBS 2359. *Yeast* 14:459–469.

897 Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate
898 alignment of transcriptomes in the presence of insertions, deletions and gene fusions.
899 *Genome Biology* 14:R36.

900 Kosaka T, Tsuzuno T, Nishida S, Pattanakittivorakul S, Murata M, Miyakawa I, Lertwattanasakul
901 N, Limtong S, Yamada M. 2022. Distinct Metabolic Flow in Response to Temperature in
902 Thermotolerant Kluyveromyces marxianus. *Applied and Environmental Microbiology*
903 88:e02006-21.

904 Kurtzman C, Fell JW, Boekhout T. 2011. The Yeasts: A Taxonomic Study. Elsevier

905 Lairón-Peris M, Routledge SJ, Linney JA, Alonso-del-Real J, Spickett CM, Pitt AR, Guillamón
906 JM, Barrio E, Goddard AD, Querol A. 2021. Lipid Composition Analysis Reveals
907 Mechanisms of Ethanol Tolerance in the Model Yeast Saccharomyces cerevisiae.
908 *Applied and Environmental Microbiology* 87:e00440-21.

909 Lane MM, Burke N, Karreman R, Wolfe KH, O'Byrne CP, Morrissey JP. 2011. Physiological and
910 metabolic diversity in the yeast Kluyveromyces marxianus. *Antonie van Leeuwenhoek*
911 100:507–519.

912 Lane MM, Morrissey JP. 2010. Kluyveromyces marxianus: A yeast emerging from its sister's
913 shadow. *Fungal Biology Reviews* 24:17–26.

914 Lappe-Oliveras P, Avitia M, Sánchez-Robledo SD, Castillo-Plata AK, Pedraza L, Baquerizo G,
915 Le Borgne S. 2023. Genotypic and Phenotypic Diversity of Kluyveromyces marxianus
916 Isolates Obtained from the Elaboration Process of Two Traditional Mexican Alcoholic
917 Beverages Derived from Agave: Pulque and Henequen (Agave fourcroydes) Mezcal.
918 *Journal of Fungi* 9:795.

919 Lertwattanasakul N, Kosaka T, Hosoyama A, Suzuki Y, Rodrussamee N, Matsutani M, Murata
920 M, Fujimoto N, Suprayogi, Tsuchikane K, et al. 2015. Genetic basis of the highly efficient

921 yeast Kluyveromyces marxianus: complete genome sequence and transcriptome
922 analyses. *Biotechnol Biofuels* 8:47.

923 Lertwattanasakul N, Suprayogi, Murata M, Rodrussamee N, Limtong S, Kosaka T, Yamada M.
924 2013. Essentiality of respiratory activity for pentose utilization in thermotolerant yeast
925 Kluyveromyces marxianus DMKU 3-1042. *Antonie van Leeuwenhoek* 103:933–945.

926 Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping
927 and population genetical parameter estimation from sequencing data. *Bioinformatics*
928 27:2987–2993.

929 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–
930 3100.

931 Linder T. 2018. Evaluation of the chitin-binding dye Congo red as a selection agent for the
932 isolation, classification, and enumeration of ascomycete yeasts. *Arch Microbiol* 200:671–
933 675.

934 Liti G, Louis EJ. 2012. Advances in Quantitative Trait Analysis in Yeast. *PLOS Genetics*
935 8:e1002912.

936 Madeira-Jr JV, Gombert AK. 2018. Towards high-temperature fuel ethanol production using
937 Kluyveromyces marxianus: On the search for plug-in strains for the Brazilian sugarcane-
938 based biorefinery. *Biomass and Bioenergy* 119:217–228.

939 Masly JP, Azom MG. 2022. Molecular divergence with major morphological consequences:
940 development and evolution of organ size and shape. *Essays in Biochemistry* 66:707–
941 716.

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

Mejía-Barajas JA, Martínez-Mora JA, Salgado-Garciglia R, Noriega-Cisneros R, Ortiz-Avila O, Cortés-Rojo C, Saavedra-Molina A. 2017. Electron transport chain in a thermotolerant yeast. *J Bioenerg Biomembr* 49:195–203.

Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2021. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36:5516–5518.

Merico A, Sulo P, Piškur J, Compagno C. 2007. Fermentative lifestyle in yeasts belonging to the Saccharomyces complex. *The FEBS Journal* 274:976–989.

Montini N, Doughty TW, Domenzain I, Fenton DA, Baranov PV, Harrington R, Nielsen J, Siewers V, Morrissey JP. 2022. Identification of a novel gene required for competitive growth at high temperature in the thermotolerant yeast Kluyveromyces marxianus. *Microbiology* 168:001148.

Nagahama T, Hamamoto M, Nakase T, Horikoshi K. 1999. Kluyveromyces nonfermentans sp. nov., a new yeast species isolated from the deep sea. *International Journal of Systematic and Evolutionary Microbiology* 49:1899–1905.

Nagy LG, Tóth R, Kiss E, Slot J, Gácser A, Kovács GM. 2017. Six Key Traits of Fungi: Their Evolutionary Origins and Genetic Bases. *Microbiology Spectrum* 5:10.1128/microbiolspec.funk-0036–2016.

961 Nespolo RF, Solano-Iguaran JJ, Paleo-López R, Quintero-Galvis JF, Cubillos FA, Bozinovic F.
962 2020. Performance, genomic rearrangements, and signatures of adaptive evolution:
963 Lessons from fermentative yeasts. *Ecology and Evolution* 10:5240–5250.

964 Ortiz-Merino RA, Varela JA, Coughlan AY, Hoshida H, da Silveira WB, Wilde C, Kuijpers NGA,
965 Geertman J-M, Wolfe KH, Morrissey JP. 2018. Ploidy Variation in Kluyveromyces
966 marxianus Separates Dairy and Non-dairy Isolates. *Frontiers in Genetics* [Internet] 9.
967 Available from: https://www.frontiersin.org/articles/10.3389/fgene.2018.00094

968 Papouskova K, Akinola J, Ruiz-Castilla FJ, Morrissey JP, Ramos J, Sychrova H. 2024. The
969 superior growth of Kluyveromyces marxianus at very low potassium concentrations is
970 enabled by the high-affinity potassium transporter Hak1. *FEMS Yeast Research*
971 24:foae031.

972 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer
973 P, Weiss R, Dubourg V, et al. 2011. Scikit-learn: Machine Learning in Python. *Journal of*
974 *Machine Learning Research* 12:2825–2830.

975 Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies.
976 *Bioinformatics* 21:676–679.

977 Prado MR, Blandón LM, Vandenberghe LPS, Rodrigues C, Castro GR, Thomaz-Soccol V,
978 Soccol CR. 2015. Milk kefir: composition, microbial cultures, biological activities, and
979 related products. *Frontiers in Microbiology* [Internet] 6. Available from:
980 https://www.frontiersin.org/articles/10.3389/fmicb.2015.01177

981 Proux-Wéra E, Armisén D, Byrne KP, Wolfe KH. 2012. A pipeline for automated annotation of
982 yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics*
983 13:237.

984 Rainey TJ, Covey G. 2016. Pulp and paper production from sugarcane bagasse. In: Sugarcane-
985 Based Biofuels and Bioproducts. John Wiley & Sons, Ltd. p. 259–280. Available from:
986 https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118719862.ch10

987 Rajkumar AS, Morrissey JP. 2022. Protocols for marker-free gene knock-out and knock-down in
988 Kluyveromyces marxianus using CRISPR/Cas9. *FEMS Yeast Research* 22:foab067.

989 Ren L, Zha H, Zhang Q, Xie Y, Li J, Hu Z, Tao X, Xu D, Li F, Zhang B. 2024. Altered sterol
990 composition mediates multiple tolerance of Kluyveromyces marxianus for xylitol
991 production. *Microb Cell Fact* 23:271.

992 Ribeiro RA, Bourbon-Melo N, Sá-Correia I. 2022. The cell wall and the response and tolerance
993 to stresses of biotechnological relevance in yeasts. *Front. Microbiol.* [Internet] 13.
994 Available from:
995 https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.953479/full

996 Rij NJWK. 2013. The Yeasts: A Taxonomic Study. Elsevier

997 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential
998 expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

999 Saini P, Beniwal A, Vij S. 2017. Physiological response of *Kluyveromyces marxianus* during
1000 oxidative and osmotic stress. *Process Biochemistry* 56:21–29.

1001 Seabold S, Perktold J. 2010. Statsmodels: Econometric and Statistical Modeling with Python.
1002 In: Austin, Texas. p. 92–96. Available from:
1003 https://conference.scipy.org/proceedings/scipy2010/seabold.html

1004 Shen X-X, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, Haase MAB, Wisecaver
1005 JH, Wang M, Doering DT, et al. 2018. Tempo and Mode of Genome Evolution in the
1006 Budding Yeast Subphylum. *Cell* 175:1533-1545.e20.

1007 Shimada TL, Takano Y, Hara-Nishimura I. 2015. Oil body-mediated defense against fungi: From
1008 tissues to ecology. *Plant Signaling & Behavior* 10:e989036.

1009 Smith SD, Pennell MW, Dunn CW, Edwards SV. 2020. Phylogenetics is the New Genetics (for
1010 Most of Biodiversity). *Trends in Ecology & Evolution* 35:415–425.

1011 Sokolov SS, Popova MM, Pohl P, Horner A, Akimov SA, Kireeva NA, Knorre DA, Batishchev
1012 OV, Severin FF. 2022. Structural Role of Plasma Membrane Sterols in Osmotic Stress
1013 Tolerance of Yeast Saccharomyces cerevisiae. *Membranes* 12:1278.

1014 Sorrells TR, Booth LN, Tuch BB, Johnson AD. 2015. Intersecting transcription networks
1015 constrain gene regulatory evolution. *Nature* 523:361–365.

1016 Sorrells TR, Johnson AN, Howard CJ, Britton CS, Fowler KR, Feigerle JT, Weil PA, Johnson
1017 AD. 2018. Intrinsic cooperativity potentiates parallel cis-regulatory evolution. *eLife*
1018 7:e37563.

1019 Spurley WJ, Fisher KJ, Langdon QK, Buh KV, Jarzyna M, Haase MAB, Sylvester K, Moriarty
1020 RV, Rodriguez D, Sheddan A, et al. 2022. Substrate, temperature, and geographical
1021 patterns among nearly 2000 natural yeast isolates. *Yeast* 39:55–68.

1022 Stemmer M, Thumberger T, Keyer M del S, Wittbrodt J, Mateo JL. 2015. CCTop: An Intuitive,
1023 Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLOS ONE* 10:e0124633.

1024 Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence
1025 alignments into the corresponding codon alignments. *Nucleic Acids Research* 34:W609–
1026 W612.

1027 Suzuki T, Hoshino T, Matsushika A. 2014. Draft Genome Sequence of Kluyveromyces
1028 marxianus Strain DMB1, Isolated from Sugarcane Bagasse Hydrolysate. *Genome*
1029 *Announcements* 2:10.1128/genomea.00733-14.

1030 Taylor JW, Branco S, Gao C, Hann-Soden C, Montoya L, Sylvain I, Gladieux P. 2017. Sources
1031 of Fungal Genetic Variation and Associating It with Phenotypic Diversity. *Microbiology*
1032 *Spectrum* 5:10.1128/microbiolspec.funk-0057–2016.

1033 Varela JA, Puricelli M, Montini N, Morrissey JP. 2019. Expansion and Diversification of MFS
1034 Transporters in Kluyveromyces marxianus. *Front. Microbiol.* [Internet] 9. Available from:
1035 https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2018.03330/full

1036 Varela JA, Puricelli M, Ortiz-Merino RA, Giacomobono R, Braun-Galleani S, Wolfe KH,
1037 Morrissey JP. 2019. Origin of Lactose Fermentation in Kluyveromyces lactis by
1038 Interspecies Transfer of a Neo-functionalized Gene Cluster during Domestication.
1039 *Current Biology* 29:4284-4290.e2.

1040 Verma S, Awasthi MK, Liu T, Awasthi SK, Syed A, Bahkali AH, Verma M, Zhang Z. 2023.
1041 Influence of biochar on succession of fungal communities during food waste composting.
1042 *Bioresource Technology* 385:129437.

1043 Wang JJ-T, Steenwyk JL, Brem RB. 2024. Natural trait variation across Saccharomycotina
1044 species. *FEMS Yeast Res* 24:foae002.

1045 Weiss CV, Brem RB. 2019. Dissecting Trait Variation across Species Barriers. *Trends in*
1046 *Ecology & Evolution* 34:1131–1136.

1047 Williams DL, Schückel J, Vivier MA, Buffetto F, Zietsman AJJ. 2019. Grape pomace
1048 fermentation and cell wall degradation by Kluyveromyces marxianus Y885. *Biochemical*
1049 *Engineering Journal* 150:107282.

1050 Wu D, Wang D, Hong J. 2020. Effect of a Novel Alpha/Beta Hydrolase Domain Protein on
1051 Tolerance of K. marxianus to Lignocellulosic Biomass Derived Inhibitors. *Frontiers in*
1052 *Bioengineering and Biotechnology* [Internet] 8. Available from:
1053 https://www.frontiersin.org/articles/10.3389/fbioe.2020.00844

1054 Xie S, Shen B, Zhang C, Huang X, Zhang Y. 2014. sgRNAcas9: A Software Package for
1055 Designing CRISPR sgRNA and Evaluating Potential Off-Target Cleavage Sites. *PLOS*
1056 *ONE* 9:e100448.

1057 Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and*
1058 *Evolution* 24:1586–1591.

1059 Yang Z, Wong WSW, Nielsen R. 2005. Bayes Empirical Bayes Inference of Amino Acid Sites
1060 Under Positive Selection. *Molecular Biology and Evolution* 22:1107–1118.

1061 Yazdi M, Pahlevanlo A, Jamab MS. 2022. Diversity of Yeasts Microbiota in Iranian Doogh and
1062 Influence of Kluyveromyces Marxianus on its Protein profiles. *Applied Food*
1063 *Biotechnology* 9:41–52.

1064 Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an Improved Branch-Site Likelihood Method
1065 for Detecting Positive Selection at the Molecular Level. *Molecular Biology and Evolution*
1066 22:2472–2479.

1067 Zhang L, Zhan H, Xu W, Yan S, Ng SC. 2021. The role of gut mycobiome in health and
1068 diseases. *Therap Adv Gastroenterol* 14:17562848211047130.