Genetic and Transcriptomic Evidences Suggest *ARO10* Genes Are Involved in Benzenoid Biosynthesis by Yeast

Maria Jose Valera*[1], Ari Zeida[2], Eduardo Boido[1], Gemma Beltran[3], María Jesús Torija[3], Albert Mas[3], Rafael Radi[2], Eduardo Dellacassa[4] and Francisco Carrau[1]

[1]Facultad de Quimica, Food Science and Technology Department, Enology and Fermentation Biotechnology Area, Universidad de la República, Uruguay

[2]Centro de Investigaciones Biomédicas, Departamento de Bioquímica, Facultad de Medicina; Universidad de la Republica, Uruguay

[3]Department of Biochemistry and Biotechnology, Biotecnología Enológica, Universitat Rovira i Virgili, Spain

[4]Departamento de Química Orgánica, Laboratorio de Biotecnología de Aromas, Universidad de la República, Uruguay

*Correspondance e-mail: mjvalera@fq.edu.uy

Abstract

Benzenoids are compounds associated with floral and fruity flavours in flowers, fruits and leaves and present a role in hormonal signalling in plants. These molecules are produced by the phenyl ammonia lyase pathway. However, some yeasts can also synthesize them from aromatic amino acids using an alternative pathway that remains unknown. *Hanseniaspora vineae* can produce benzenoids at levels up to two orders of magnitude higher than *Saccharomyces* species, so it is a model microorganism for studying benzenoid biosynthesis pathways in yeast. According to their genomes, several enzymes have been proposed to be involved in a mandelate pathway similar to that described for some prokaryotic cells. Among them, the *ARO10* gene product could present benzoylformate decarboxylase activity. This enzyme catalyses the decarboxylation of benzoylformate into benzaldehyde at the end of the mandelate pathway in benzyl alcohol formation.

Two homologous genes of *ARO10* were found in the two sequenced *H. vineae* strains. In this study, nine other *H. vineae* strains were analysed to detect the presence and percent homology of *ARO10* sequences by PCR using specific primers designed for this species. Also, the copy number of the genes was estimated by quantitative PCR. To verify the relation of *ARO10* with the production of benzyl alcohol during fermentation, a deletion mutant in the *ARO10* gene of *S. cerevisiae* was used. The two Hv*ARO10* paralogues were analysed and compared with other α-ketoacid decarboxylases at the sequence and structural level.

Keywords: benzoylformate decarboxylase, wine yeast, 4-hydroxybenzaldehyde, coenzyme Q

Acknowledgements

We wish to thank the following agencies for the financial support yielded to this work: CSIC Group Project #656 and CSIC Productive Sector Project #602 of UdelaR, Uruguay, Agencia Nacional de Investigación e Innovación (ANII) *Hanseniaspora vineae* FMV 6956 project and a postdoctoral fellowship (PD_NAC_2016_1_133945), and a Clarín-COFUND postdoctoral fellowship from Principado de Asturias and European Union.

Introduction

Benzenoids are valuable molecules as aroma compounds from plant metabolism present in different sources such as tea leaves [Wang et al., 2019], mei [Zhao et al., 2017], petunia [Qualley et al., 2012] and grape [Alessandrini et al., 2017]. Phenylalanine and tyrosine metabolism are considered a valuable fount of aromatic molecules [Lapadatescu et al., 2000], although the routes differ depending on the species, particularly in the final steps, and some of them remain uncharted [Zhao et al., 2017]. In plants and some fungi, phenylalanine ammonia lyase (PAL) and tyrosine ammonia lyase (TAL) are the key activities involved in the production of benzenoids from phenylalanine and tyrosine, respectively [Widhalm and Dudareva, 2015]. The production of benzenoids has been investigated thoroughly in plants, and while the biosynthesis of these compounds involves the PAL/TAL pathways, alternative routes have also been explored [Jensen et al., 1994; Lapadatescu et al., 2000, Tsui and Clarke, 2019].

In addition, yeasts can produce benzenoids during fermentation by using phenylalanine and tyrosine as precursors [Martin et al., 2016a; Giorello et al., 2019], by an alternative pathway to the PAL/TAL activities. Due to the importance of this biosynthetic pathway, a description of the metabolites and enzymes implicated in the pathways is necessary for a better understanding of these processes. The selection of model microorganisms is essential for an accurate depiction of the metabolic reactions that take part along this route.

Hanseniaspora vineae is a yeast frequently found in plant environments, being commonly isolated from grapes and grape juices. This species is characterized by its increased production of phenylpropanoids, especially when compared with *Saccharomyces cerevisiae* [Martin et al., 2016a]. In a previous study [Martin et al., 2016a], a benzenoid biosynthesis pathway has been proposed by genomic analysis, searching gene homology in *H. vineae.* The biosynthesis of benzyl alcohol from phenylalanine and 4-hydroxybenzaldehyde from tyrosine in *H. vineae* requires the decarboxylation of benzoylformate and 4-hydroxybenzoylformate, respectively [Martin et al., 2016a]. The *ARO10* gene product has been proposed to catalyse this step. Other authors [Stefely et al., 2016] have hypothesised a relation between the production of 4-hydroxy benzoic acid through 4-hydroxyphenylacetaldehyde from 4-hydroxyphenylpyruvate and although the involvement of *ARO10* gene of *S. cerevisiae* was mentioned, the detailed mechanism remains undefined. In fact, the *ARO10* gene of *S. cerevisiae* is involved in the Ehrlich pathway, the metabolic route for the catabolism of amino acids. This gene encodes phenylpyruvate decarboxylase (PPDC), which catalyses the irreversible decarboxylation of phenylpyruvate into phenylacetaldehyde. This enzyme presents broad substrate specificity [Vuralhan et al., 2005; Kneen et al., 2011]. Similarly to benzoylformate decarboxylase (BFDC) in *Pseudomonas putida,* Aro10p from *S. cerevisiae* decarboxylates benzoylformate into benzaldehyde; however, its efficiency is lower for this substrate [Kneen et al., 2011].

PPDC, BFDC, pyruvate decarboxylase (PDC) and indole-3-pyruvate decarboxylases (IPDC) are thiamine diphosphate (ThDP)-dependent alpha-ketoacid enzymes with diverse substrate specificities [Iding et al., 1998]. The amino acid sequences of the members of this enzymatic group are not conserved among different species [Iding et al., 1998], but the binding-specific sites show coincident residues [Spaepen et al., 2007].

94 The present study analyses a *S. cerevisiae* mutant strain with a deletion in the *ARO10*
95 gene to demonstrate the relation of this gene with benzyl alcohol and 4-
96 hydroxybenzaldehyde production. The presence of two paralogous *ARO10* in different *H.*
97 *vineae* strains and their role in benzenoid biosynthesis were analysed, comparing
98 predicted amino acid sequences and structural modelling with other α-keto acid
99 decarboxylases.

100 **Materials and methods**

101 *Yeast strains*

102 Eleven strains of *H. vineae* isolated from Uruguayan grape fermentations (Table S1) were
103 used in this study. Among them, *H. vineae* T02/05AF and T02/19AF sequenced by
104 Giorello et al. [2019] were used for genomic comparisons and primer design.

105 *S. cerevisiae* BY4743 wild type (WT) was used as the control strain for mutant
106 experiments. The homozygotic diploid double mutant strain *S. cerevisiae* Δ*aro10* from
107 the Yeast Knockout Collection [Giaever et al., 2002] was supplied by Dharmacon
108 (Lafayette, CO, USA).

109 All the strains were grown in YPD medium (2% glucose, 2% peptone, 1% yeast extract).
110 In the case of the *S. cerevisiae* mutant strain, YPD medium was supplemented with 200
111 µg/mL Gentamycin G418 (Acros-Fischer, Belgium).

112 *DNA extraction*

113 *H. vineae* strains were grown in YPD medium. The cell concentration for each culture
114 was calculated by counting under a microscope in a Neubauer chamber. Aliquots of 10^7
115 cells were collected and centrifuged at 12,000 rpm for 10 minutes. Pellets formed were

116 stored at -20ºC until use. DNA extraction was performed using a DNeasy Plant Mini kit
117 (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

118 *PCR amplification and sequencing*

119 *H. vineae* T02/05AF and *H. vineae* T02/19AF genome sequences [Giorello et al., 2019]
120 were used to design primers for PCR amplification of *ARO10* homologous sequences.
121 *ARO10* were amplified and sequenced using four primers for each homologous gene. The
122 primers used are listed in Table 1. Amplification reactions were performed in a total
123 volume of 50 µL. The PCR mixture contained 100 mM of each forward and reverse
124 primer, 200 mM of each of the four dNTPs (Roche Diagnostics GmBh, Manheim,
125 Germany), 1 × PCR reaction buffer (Ecogen, Spain), 1.5 mM $MgCl_2$, 0.4 U EcoTaq DNA
126 Polymerase (Ecogen, Spain), and 5 µL of DNA template (approximately 50 ng/mL). An
127 initial denaturation step at 94ºC for 5 minutes, followed by 35 cycles of denaturation at
128 94ºC for 1 minute, annealing at 50ºC for 45 seconds and extension at 72ºC for 2 minutes,
129 and a final extension cycle at 72ºC for 10 minutes were performed in a GeneAmp PCR
130 System 2700 (Applied Biosystems). The PCR products were visualised by
131 electrophoresis in 1% (w/v) agarose gels using a DNA XIV 100 bp ladder (Roche
132 Diagnostics, Mannheim, Germany).

133 For complete coverage of the gene sequence, all the positive PCR products were purified
134 and sequenced using additional primers (Table 1) by Macrogen Inc. (Seoul, South Korea)
135 in an ABI3730 XL automatic DNA sequencer. Sequences were manually assembled using
136 ClustalW and translated into amino acid sequences using ExPASy, the SIB bioinformatics
137 resource portal [Artimo et al., 2012]. Alignments were analysed by MEGA version 4
138 [Tamura et al., 2007] and the amino acid sequences were visualised using ESPript
139 software (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi).

140 *Real-time PCR analysis*

141 DNA concentration and purity were determined by using a NanoDrop 1000
142 Spectrophotometer (Thermo Scientific, Whalthan, MA, USA). All the samples were
143 diluted to a concentration 0.1 ng/µL and subsequently to 0.01 ng/µL for copy number
144 analysis. Each *ARO10* homologue was amplified separately by real-time PCR with
145 SYBR-Green fluorescence detection. The calibration curves were constructed by using
146 10-fold dilutions starting with 1 ng/µL of the DNA extracted from the strains *H. vineae*
147 T02/05AF and *H. vineae* T02/19AF, whose genomes were sequenced previously
148 [Giorello et al., 2019], considering one copy for each homologous *ARO10* gene. The
149 primers used for each homologue are listed in Table 1, and the reactions were carried out
150 in a total volume of 25 µL containing 2 µL of DNA solution, 12.5 µL SYBR-Premix Ex
151 Taq II (tli RNase h plus), 600 nM of each primer and 4.5 mL of sterile H_2O. The
152 amplification was performed on a 7300 Real-Time PCR System (Applied Biosystems,
153 Foster City, CA, USA). Reactions took place for 2 minutes at 50 ºC and 10 minutes at 95
154 ºC, followed by 40 cycles of 15 s 95 ºC and 1 min at 60 ºC. The Ci value was determined
155 automatically by the instrument, and NTC reactions were used as negative controls.

156 *Fermentation conditions*

157 *H. vineae* and *S. cerevisiae* fermentations were performed with 100 mg N/L of yeast
158 assimilable nitrogen. Fermentation medium was prepared as described previously [Carrau
159 et al., 2008] with some modifications described below. The final pH of the medium was
160 adjusted to 3.5 with HCl. Equimolar concentrations of glucose and fructose were added
161 to reach 200 g/L and the mixed vitamins and salts were as described previously [Carrau
162 et al., 2008]. Ergosterol was added as the only supplemented lipid at a final concentration
163 of 10 mg/L. Also, in experiments performed with *S. cerevisiae* strains, media were
164 supplemented with 125 mg/L of histidine, 500 mg/L of lysine, 150 mg/L of uracil and

165 500 mg/L of leucine. The inoculum size was 1 x 10^5 cells/mL in the final medium for all
166 strains. Static batch fermentation conditions were conducted at 20 ^oC in triplicate.

167 For treatments with *H. vineae* wine strains, fermentations were carried out in a volume of
168 125 mL in Erlenmeyer flasks cotton plugged to simulate microaerobic conditions. *S.*
169 *cerevisiae* BY4743 WT was used as control strain in the same conditions. The final
170 amount of yeast assimilable nitrogen was reached by the sum of amino acids (50 mg N/L)
171 and diammonium phosphate (50 mg N/L).

172 For the mutant analysis, fermentation inoculated with *S. cerevisiae* BY4743 WT and *S.*
173 *cerevisiae* Δ*aro10* were carried out in a volume of 250 mL in Erlenmeyer flasks cotton
174 plugged to simulate microaerobic conditions. Fermentations were also performed with
175 100 mg N/L, but diammonium phosphate was not added and double the concentration of
176 each amino acid was used to maximize the production of benzenoids.

177 *Transcriptomic analysis*

178 For transcriptomic study, total RNA obtained from *H. vineae* T02/19AF strain from three
179 different fermentation stages (days 1, 4 and 10) were analysed in independent replicates.
180 The nine samples were paired-end sequenced using Illumina MySeq. Trinity was used to
181 assemble the raw reads from transcriptomic analysis as specified by Giorello et al. [2019].
182 Transcriptomic reads were aligned against the transcriptomic reference implementing
183 RSEM (default settings) [Li et al., 2011]. The obtained expected counts for each gene
184 were then used for the differential gene expression analysis carried out with edgeR
185 [Robinson et al., 2010]. Genes with FDR < 0.05 were considered differentially expressed.

186 *GC-MS analysis*

187 After 12 days of fermentation, the entire volume in the flasks was centrifuged at 18,000
188 rpm for 10 minutes to separate cells and extracellular medium. Extracellular aromatic

189 compounds were separated by liquid-liquid extraction using dichloromethane. The
190 extracellular media were extracted three times with dichloromethane 1:5 v/v. The
191 solutions were dried over anhydrous sodium sulphate, concentrated at 40 ºC on a Vigreux
192 column using a thermostatized water bath and then under a N_2 stream. Sample treatment
193 and GC-MS analysis were performed as described previously [Martin et al., 2016a] in a
194 Shimadzu-QP 2010 ULTRA (Tokyo, Japan) mass spectrometer equipped with a
195 Stabilwax (30 m x 0.25 mm i.d., 0.25-µm film thickness, Restek) capillary column.
196 Volatiles were identified by comparison of their linear retention index, with pure
197 standards for benzyl alcohol and 4-hydroxybenzaldehyde. Comparison of mass spectral
198 fragmentation patterns with those stored in databases was also performed. GC-MS
199 instrumental procedures were applied for quantitative purposes, as described previously
200 [Martin et al., 2016a]. The internal standard was 1-octanol (Sigma, Aldrich, Milwaukee,
201 USA).

202 *Structural analysis*

203 The crystal structure of PPDC from *A. brasilense* in complex with its substrate and
204 cofactors phenylpyruvate, ThDP and Mg^{2+} (PDBid 2Q5O) at 1.5 Å [Versees et al., 2007],
205 was used as a model to analyse and compare the conservation of key active site residues
206 in the amino acid sequences predicted from Hv*ARO10* genes. All molecular drawings
207 were generated with VMD 1.9.1 [Humphrey et al., 1996]

208 *Statistical analysis*

209 ANOVA analyses of benzyl alcohol and 4-hydroxybenzaldehyde detected in
210 fermentation were performed for the *H. vineae* and *S. cerevisiae* strains. All ANOVA
211 analyses were performed with STATISTICA 7.0 software. Differences in mean
212 benzenoid compound concentrations were evaluated by the least significant differences
213 test.

214 **Results and discussion**

215 The route for benzyl alcohol biosynthesis from phenylalanine in yeast (Fig 1) was
216 proposed in *H. vineae* by a genomic analysis (Martin et al. 2016a). In different strains of
217 *H. vineae* the production of benzyl alcohol ranged from 86.51 to 620.27 µg/L (Table S1),
218 while in *S. cerevisiae* at the same fermentation conditions it produces less than 2-3 µg/L,
219 a concentration level that is in the quantification limit of the GC/MS (Table S1). High
220 yeast assimilable nitrogen and also the use of ammonium salts inhibit the synthesis of
221 benzyl alcohol (Martin et al. 2016b). Therefore, in this work the experiments with *S.*
222 *cerevisiae* WT and Δ*aro10* mutant were performed using a low yeast assimilable nitrogen
223 level (100 mg N/L) and composed of amino acids as sole nitrogen source. In this
224 fermentation medium we obtained a production of benzyl alcohol above 20 µg/L that
225 allowed us to make comparisons with the deletion mutant.

226 *Benzyl alcohol production in S. cerevisiae and the Δaro10 mutant*

227 *S. cerevisiae* presents one *ARO10* gene that encodes phenylpyruvate decarboxylase
228 activity. This enzyme is involved in the Erlich pathway as part of amino acid metabolism,
229 but it presents a broad substrate specificity [Vuralhan et al., 2005; Romagnoli et al., 2012].
230 One of the substrates of Aro10p is benzoylformate, which is decarboxylated to
231 benzaldehyde. The affinity for this substrate is reduced in *S. cerevisiae*, especially
232 compared with benzoylformate decarboxylase activity in the bacteria *P. putida,* which is
233 part of the mandelate pathway [Vuralhan et al., 2005].

234 In order to probe the role of *ARO10* in the production of benzyl alcohol the *S. cerevisiae*
235 deletion mutant Δ*aro10* was used. The production of benzyl alcohol during fermentation
236 with this mutant was practically null as was that of 4-hydroxybenzaldehyde, the product

237 of the parallel route from tyrosine (Fig 2A). Furthermore, similar results were found using
238 the inhibitor molecule methyl-benzoylphosphonate (MBP) added to the fermentation
239 medium in *S. cerevisiae* WT [Valera et al., 2020]. MBP is an analogue of benzoylformate
240 that forms a covalent MBP-ThDP adduct, blocking the decarboxylation of
241 benzoylformate. These results suggest the involvement of *ARO10* in the benzenoids
242 synthesis pathway of *S. cerevisiae* as predicted at two conversion steps (Fig 1).
243

244 *Homology of HvARO10 genes with α-keto acid decarboxylases*

245 The study of the whole genomes of *H. vineae* T02/05AF and T02/19AF strains revealed
246 in both cases the presence of two orthologous genes similar to *S. cerevisiae ARO10.* The
247 two homologous sequences were designated as Hv*ARO10*A and Hv*ARO10*B. The length
248 of the predicted sequences are 699 and 620 amino acids, and they present 52% similarity
249 between Hv*ARO10*A and Hv*ARO10*B. The presence of two isoforms of *ARO10* in
250 *Brettanomyces bruxellensis* (syn. *Dekkera bruxellensis*) has been reported, and the
251 authors hypothesize that these two Aro10p activities could present different biological
252 functions regarding their expression profiles [Liberal et al., 2012]. *ARO10*A and *ARO10*B
253 in *H. vineae* also present different expression profiles during fermentation (Fig S2) as
254 described by Giorello et al. [2019]; however, no functional consequences have yet been
255 analysed.
256 Regarding the amino acids of the whole protein predicted from the *ARO10* DNA
257 sequences of *H. vineae* T02/05AF, *ARO10*B presents higher similarity to all model α-
258 keto acid decarboxylases structurally and functionally described from other
259 microorganisms (Table S2). The predicted amino acid sequences of Hv*ARO10*A and
260 Hv*ARO10*B present the highest homology with *Sc*PPDC. This protein was purified and

261 characterized biochemically [Kneen et al., 2011], but the crystallographic structure was
262 not obtained.

263 Most enzymes with α-keto acid decarboxylase activity are ThDP-dependent. They create
264 a ThDP-bound carbanionic intermediate upon cleavage of the C-C bond [Iding et al.,
265 1998]. Generally, ThDP-dependent decarboxylases are tetrameric; however, *Sc*PPDC
266 exists primarily in solution as a dimer with a small proportion of tetramer [Kneen et al.,
267 2011]. Regarding *Ab*PPDC from *Azospirillum brasilense* whose crystallographic
268 structure was previously resolved [Versées et al., 2007], the residues involved in the
269 active site are necessarily positioned in two different subunits [Spaepen et al., 2007].
270 There are two histidine residues conserved in the active site of ThDP-dependent
271 decarboxylases [Kneen et al., 2011]. Furthermore, the residues in the catalytic triad of
272 *Ab*PPDC D^{45}-H^{132}-D^{303} are partially conserved as D^{49}-H^{144}-E^{334} in *Sc*PPDC. Also, both
273 Hv*ARO10*A and Hv*ARO10*A sequences present these amino acids conserved as
274 represented in Fig 3A and 3B. These residues are proposed to play a role in protonation
275 of the enamine/carbanion intermediate in the decarboxylation in *Ab*PPDC [Versées et al.,
276 2007]. Another conserved motif is found in α-keto acid decarboxylase corresponding to
277 a ThDP-binding domain (Fig 3C) which starts with GDG and ends with NN [Spaepen et
278 al., 2007].

279 *Ab*PPDC cannot decarboxylate benzoylformate [Spaepen et al., 2007], in contrast to other
280 α-keto acid decarboxylases such as IPDC from *Enterobacter cloacae* (*Ec*IPDC), which
281 can also decarboxylate pyruvate but not phenylpyruvate [Schütz et al., 2003].
282 Phylogenetic analysis suggests these activities are derived from at least two different
283 ancestors [Spaepen et al., 2007].
284 The amino acids proposed as part of the active site in BFDC of *P. putida* are partially
285 conserved in other α-keto acid decarboxylases (Table 2). The two histidine residues

286 present, H^{70} and H^{281} in *Pp*BBFD, fulfil the functions of His^{132} and His^{133} in *Ab*PPDC.
287 These residues originate in very different places, as shown by the sequence alignments;
288 however, the functional groups of the imidazole rings are in nearly the same positions in
289 the active sites, as was reported in *Saccharomyces* [Hasson et al., 1998]. In the active site
290 of *Pp*BFDC, N^{27} presents a similar position to N^{51} in *Sc*PPDC and N^{45} Hv*ARO10*A but
291 not in Hv*ARO10*B (Table 2). Moreover, S^{26} in *Pp*BFDC and D^{49} in *Sc*PPDC are in the
292 same position (Table 2) in the protein sequence and fold probably contributing to different
293 roles in the active site, as it was observed in *Sc*PDC [Hasson et al., 1998]. The sequence
294 of the active site of *Pp*BFDC is not highly conserved; however, the binding site for the
295 cofactor is similar in *Pp*BFDC and other ThDP-dependent enzymes (Fig 3C). In fact,
296 residues E^{47} and Y^{458} are conserved in *Sc*PPDC, Hv*ARO10*A and Hv*ARO10*B, forming
297 part of the ThDP-binding region (Table 2).
298
299 *Presence, homology and copy number of ARO10 genes in different strains of H. vineae*

300 In order to study the presence of both isoforms Hv*ARO10*A and Hv*ARO10*B in different
301 strains of *H. vineae,* specific primers were designed on the basis of the sequences of
302 T02/05AF and T02/19AF. The sequences of these two genes presented high homology in
303 both strains (98% Hv*ARO10*A and 99% Hv*ARO10*B).
304 Nine other strains of *H. vineae* were analysed by PCR. Amplicons of approximately
305 1600−2000 bp were obtained for *ARO10*A and *ARO10*B. The two homologues for all the
306 strains were sequenced and assembled. The corresponding amino acid sequence of each
307 one was predicted and compared by ClustalW alignment. The similarity between strains
308 in these amino acid sequences in homologous *ARO10*A ranged from 59% to 99% and
309 *ARO*10B was highly conserved, presenting 99%−100% homology. All these sequences
310 are separated in two different clusters, one for each homologue (Fig 4) and present less

311 similarity with other *ARO10* from different yeast species regarding the sequences
312 obtained in public databases.
313 *H. vineae* M12/196AF was the highest producer of benzyl alcohol (Table S1) as
314 previously described by Martin et al. [2016a]. The concentration of benzyl alcohol
315 detected in the fermentation medium after 12 days was three times higher than that
316 quantified in *H. vineae* T02/19AF (Figure 2B). Curiously, both homologous Hv*ARO10*A
317 and Hv*ARO10*B genes in *H. vineae* M12/196AF presented three copies of the target
318 sequence used in real-time PCR detection, while most of the strains analysed showed just
319 one for each (Table S1). Regarding the homology of these sequences, in M12/196F
320 Hv*ARO10*A was 96% similar in amino acids to those from T02/05AF and T02/19AF but
321 the coverage of the sequence assembled was just 375 amino acids out of the 699 amino
322 acid present in the two genomes wholly sequenced. The amplification of the gene was
323 correctly performed, yielding the amplicon size expected, but sequencing results revealed
324 overlapped peaks probably corresponding to different alleles for this gene (data not
325 shown). Conversely, the Hv*ARO10*B gene from M12/196F is highly similar in sequence
326 (99.8%) compared with T02/05AF and T02/19AF. In these three strains, *ARO10*B shows
327 a predicted sequence of 620 amino acids that is remarkably conserved.

328 **Conclusion**

329 *ARO10* genes play a role in benzenoid formation in *S. cerevisiae,* and our genetic and
330 transcriptomic results suggest a similar role of this decarboxylase enzyme in *H. vineae*.
331 The conversion of phenylalanine and tyrosine into benzyl alcohol and 4-
332 hydroxybenzaldehyde in yeast present a step catalysed by an *ARO10* product, as
333 demonstrated by the results obtained from fermentation with the Δ*aro10* deletion mutant
334 strain. Moreover, information from sequence alignments and structures suggests that the
335 Hv*ARO10*A and Hv*ARO10*B genes have benzoylformate decarboxylase activities. These

336 protein models and the higher copy number of *ARO10* in *H. vineae* strains, might
337 explained the highly increased formation of benzyl alcohol during fermentation compared
338 to *S. cerevisiae*. Functional analysis of *ARO10* genes in *H. vineae* is necessary to confirm
339 this activity, which is putatively involved in yeast aroma production.

340 **References**

341 Alessandrini M, Gaiotti F, Belfiore N, Matarese F, D'Onofrio C, Tomasi D. 2017.
342 Influence of vineyard altitude on Glera grape ripening (*Vitis vinifera* L.): effects on aroma
343 evolution and wine sensory profile. *J. Sci. Food Agric*. *97*(9): 2695-2705.

344 Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, Duvaud S,
345 Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D,
346 Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H
347 (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res. 40*(W1): W597-
348 W603.

349 Carrau FM, Medina K, Farina L, Boido E, Henschke PA, Dellacassa E. 2008. Production
350 of fermentation aroma compounds by *Saccharomyces cerevisiae* wine yeasts: effects of
351 yeast assimilable nitrogen on two model strains. *FEMS Yeast Res. 8*(7): 1196-1207.

352 Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A.,
353 Anderson K, André B, Arkin AP, Astromoff A, El Bakkoury M, Bangham R, Benito R,
354 Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P,
355 Foury F, Garfinkel DJ, Gerstein M, Gotte D, Güldener U, Hegemann JH, Hempel S,
356 Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kötter P, LaBonte D, Lamb DC, Lan N,
357 Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL,
358 Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B,
359 Shoemaker DD, Sookhai-Mahadeo S, K. Storms R, Strathern JN, Valle G, Voet M,
360 Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman
361 E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M. 2002.
362 Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*. *418*(6896): 387.

363 Giorello F, Valera MJ, Martin V, Parada A, Salzman V, Camesasca L, Fariña L, Boido
364 E, Medina K, Dellacassa E, Berna L, Aguilar PS, Mas A, Gaggero C, Carrau F. 2019.
365 Genomic and transcriptomic basis of *Hanseniaspora vineae*'s impact on flavor diversity
366 and wine quality. *Appl. Environ. Microbiol.*, *85*(1): e01959-18.

367 Hasson MS, Muscate A, McLeish MJ, Polovnikova LS, Gerlt JA, Kenyon GL, Petsko
368 GA, Ringe, D. 1998. The crystal structure of benzoylformate decarboxylase at 1.6 Å
369 resolution: diversity of catalytic residues in thiamin diphosphate-dependent enzymes.
370 *Biochem. 37*(28): 9918-9930.

371 Humphrey W, Dalke A, Schulten K. 1996. VMD: Visual Molecular Dynamics. J. Mol.
372 Graph. 14: 33–38.

Iding H, Siegert P, Mesch K, Pohl M. 1998. Application of α-keto acid decarboxylases in biotransformations. *Biochimica et Biophysica Acta (BBA)-Prot. Struct. Mol. Enzymology*. *1385*(2): 307-322.

Jensen KA, Evans KM, Kirk TK, Hammel KE. 1994. Biosynthetic pathway for veratryl alcohol in the ligninolytic fungus *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol. 60*(2): 709-714.

Kneen MM, Stan R, Yep A, Tyler RP, Saehuan C, McLeish MJ. 2011. Characterization of a thiamin diphosphate-dependent phenylpyruvate decarboxylase from *Saccharomyces cerevisiae*. *FEBS J. 278*(11): 1842-1853.

Lapadatescu C, Giniès C, Le Quéré JL, Bonnarme P. 2000. Novel Scheme for Biosynthesis of Aryl Metabolites from l-Phenylalanine in the Fungus *Bjerkandera adusta*. *Appl. Environ. Microbiol. 66*(4): 1517-1522.

Li, B., Dewey, C. N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, *12*(1): 323.

Liberal ATS, Carazzolle MF, Pereira GA, Simões DA, de Morais MA. 2012. The yeast *Dekkera bruxellensis* genome contains two orthologs of the *ARO10* gene encoding for phenylpyruvate decarboxylase. *World J. Microbiol. and Biotech. 28*(7): 2473-2478.

Martin V, Giorello F, Fariña L, Minteguiaga M, Salzman V, Boido E, Aguilar P, Gaggero C, Dellacassa E, Mas A, Carrau F. 2016. De novo synthesis of benzenoid compounds by the yeast *Hanseniaspora vineae* increases the flavor diversity of wines. *J Agric Food Chem*. *64*(22): 4574-4583.

Martin V, Boido E, Giorello F, Mas A, Dellacassa E, Carrau F. 2016. Effect of yeast assimilable nitrogen on the synthesis of phenolic aroma compounds by *Hanseniaspora vineae* strains. *Yeast*. *33*(7): 323-328.

Qualley AV, Widhalm JR, Adebesin F, Kish CM, Dudareva N. 2012. Completion of the core β-oxidative pathway of benzoic acid biosynthesis in plants. *Proc. Nat. Ac. Sci. USA*. *109*(40): 16383-16388.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1): 139-140.

Romagnoli G, Luttik MA, Kötter P, Pronk JT, Daran JM. 2012. Substrate specificity of thiamine pyrophosphate-dependent 2-oxo-acid decarboxylases in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol. 78*(21): 7538-7548.

Schütz A, Sandalova T, Ricagno S, Hübner G, König S, Schneider G. 2003. Crystal structure of thiamindiphosphate-dependent indolepyruvate decarboxylase from *Enterobacter cloacae*, an enzyme involved in the biosynthesis of the plant hormone indole-3-acetic acid. *European J. Biochem. 270*(10): 2312-2321.

Spaepen S, Versées W, Gocke D, Pohl M, Steyaert J, Vanderleyden J. 2007. Characterization of phenylpyruvate decarboxylase, involved in auxin production of *Azospirillum brasilense*. *J. Bac. 189*(21): 7626-7633.

413 Stefely JA, Pagliarini DJ. 2017. Biochemistry of mitochondrial coenzyme Q biosynthesis.
414 *Trends Biochem. Sci. 42*(10): 824-843.

415 Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics
416 analysis (MEGA) software version 4.0. *Mol Biol Evol. 24*(8): 1596-1599.

417 Tsui HS, Clarke CF. 2019. Ubiquinone Biosynthetic Complexes in Prokaryotes and
418 Eukaryotes. *Cell Chem. Biol.* 26(4):465-467

419 Valera MJ, Boido E, Ramos JC, Manta E, Radi R, Dellacassa E, Carrau F. 2020.
420 Mandelate pathway, an alternative to the PAL pathway for the synthesis of benzenoids in
421 Ascomycete yeasts. *Appl. Environ. Microbiol.86(17).* DOI: 10.1128/AEM.00701-20

422 Versees W, Spaepen S, Vanderleyden J, Steyaert J. 2007. The crystal structure of
423 phenylpyruvate decarboxylase from *Azospirillum brasilense* at 1.5 Å resolution:
424 Implications for its catalytic and regulatory mechanism. *FEBS J. 274*(9): 2363-2375.

425 Vuralhan Z, Luttik MA, Tai SL, Boer VM, Morais MA, Schipper D, Almering MJH,
426 Kötter P, Dickinson JR, Daran JM, Pronk, J. T. 2005. Physiological characterization of
427 the *ARO10*-dependent, broad-substrate-specificity 2-oxo acid decarboxylase activity of
428 *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol. 71*(6): 3276-3284.

429 Wang X, Zeng L, Liao Y, Zhou Y, Xu X, Dong F, Yang Z. 2019. An alternative pathway
430 for the formation of aromatic aroma compounds derived from L-phenylalanine via
431 phenylpyruvic acid in tea (*Camellia sinensis* (L.) O. Kuntze) leaves. *Food Chem. 270*:
432 17-24.

433 Widhalm JR, Dudareva N. 2015. A familiar ring to it: biosynthesis of plant benzoic acids.
434 *Mol. Plant*. *8*(1): 83-97.

435 Zhao K, Yang W, Zhou Y, Zhang J, Li Y, Ahmad S, Zhang Q. 2017. Comparative
436 transcriptome reveals benzenoid biosynthesis regulation as inducer of floral scent in the
437 woody plant *Prunus mume*. *Front. Plant Sci. 8*: 319.

438

439

440 **Tables**

441 Table 1. Primers used in this study

Gene	Primer name	Oligonucleotide sequence (5' → 3')	Analytical use
Hv*ARO10*A	A1Fw	GGCTCTAAGGTCTCTTTGT	PCR and sequencing
	A2Fw	ATCGTAAGAGTCGCTCCATT	Sequencing
	A3Rv	ATCCTTGCACCATTTACTAC	Sequencing
	A4Rv	GATATGTTTAATAAAGTTGATGGTGGTAG	PCR and sequencing
	AqFw	CAATGCCATGCATACACCCA	RT-PCR
	AqRv	TCAGGATCCCAGAATTGAGCA	RT-PCR
Hv*ARO10*B	B1Fw	GAAGCTCAAGGATATATCAACAAGCT	PCR and sequencing
	B2Fw	GTGAATAATCTAGAGTTCTTTCCATATCCAA	Sequencing
	B3Rv	GGACATGTTGTTGATGAAAGCTTGC	Sequencing
	B4Rv	CACAACCATTGTACCATGACCGAAGA	PCR and sequencing
	BqFw	CAATGGCGTACCCAGAGTTG	RT-PCR
	BqRv	TCTGCCCAGATGACATTCCA	RT-PCR

442

443

444

445

446

447

448

449

450

451

452

453

454 Table 2. Conserved and non-conserved residues in the active site and cofactor-binding region
455 from benzoylformate decarboxylase of *P. putida* compared with phenylpyruvate decarboxylases
456 of *S. cerevisiae* and *A. brasilense* and the two *ARO10* homologous of *H. vineae*.

	***Pp*BFDC**	**Hv*ARO10*A**	**Hv*ARO10*B**	***Sc*PPDC**	***Ab*PPDC**
Conserved residues	N^{27}	N^{45}	T^{36}	N^{51}	F^{46}
	L^{29}	L^{46}	L^{37}	L^{52}	L^{48}
	E^{47}	E^{100}	E^{59}	E^{76}	E^{68}
	Y^{458}	Y^{607}	Y^{527}	Y^{542}	W^{479}
Non conserved residues	S^{26}	D^{43}	D^{34}	D^{49}	D^{45}
	H^{70}	T^{123}	T^{82}	T^{99}	T^{91}
	H^{281}	N^{388}	T^{319}	I^{335}	N^{304}
	F^{464}	I^{613}	I^{533}	I^{548}	F^{485}
	Y^{433}	M^{582}	M^{502}	Q^{516}	M^{454}

457

458

459

460 **Figure captions**

461 Fig 1. Proposed biosynthesis route of benzyl alcohol and 4-hydroxybenzaldehyde in
462 yeast. Sc*ARO10* codifies phenylpyruvate decarboxylase and ScAro10p presents weak
463 benzoylformate decarboxylase activity in *S. cerevisiae*. Modified from Martin et al.,
464 2016a

465 Fig 2. Benzenoids production A) by three strains of *H. vineae*; B) by *S. cerevisiae*
466 BY4743 Wild type (WT) and *S. cerevisiae* Δ*aro10* mutant. Samples were taken after 12
467 days of fermentation and quantified by GC-MS. Results are expressed per litre of
468 fermentation medium. Fermentations were performed in triplicate, bars indicate standard
469 deviation. Letters represent significant different values p-value 0.01.

470 Fig 3. Conserved regions in active site and cofactor binding region in Hv*ARO10*
471 homologous. A) Detailed view of *Ab*PPDC active site (PDBid 2Q5O). All residues
472 involved in substrate (PhPy) binding are showed, but only the ones conserved in
473 Hv*ARO*10A are highlighted and numbered in black (subscripts). The numbers in red
474 (superscripts) correspond to *Ab*PPDC gene and primes indicate residues belonging to the
475 adjacent subunit in the tetrameric protein array. Thiamine di-phosphate (ThDP) cofactor
476 is also showed in darker colours. B) Detailed view of *Ab*PPDC active site (PDBid 2Q5O)
477 showing conserved residues in Hv*ARO10*B. C) ThDP binding region conserved in both
478 Hv*ARO10* homologous and different α-keto acid carboxylases: *Sc*PPDC phenylpyruvate
479 decarboxylase of *Saccharomyces cerevisiae, Ec*IPDC indol-pyruvate decarboxylase of
480 *Enterobacter cloacae, Ab*PPDC phenylpyruvate decarboxylase of *Azospirillum*
481 *brasilense Pp*BFDC benzoyl formate decarboxylase of *Pseudomonas putida*. Conserved
482 amino acids are coloured depending on their chemical classification and degree of
483 conservation among the sequences analysed.

484 Fig 4. Dendrogram comparing different predicted amino acid sequences of Hv*ARO10*A and
485 Hv*ARO10*B homologous with other *ARO10* from different yeast clustered by Neighbour Joining
486 method. BFDC of *Pseudomonas putida* was used as external group.

487

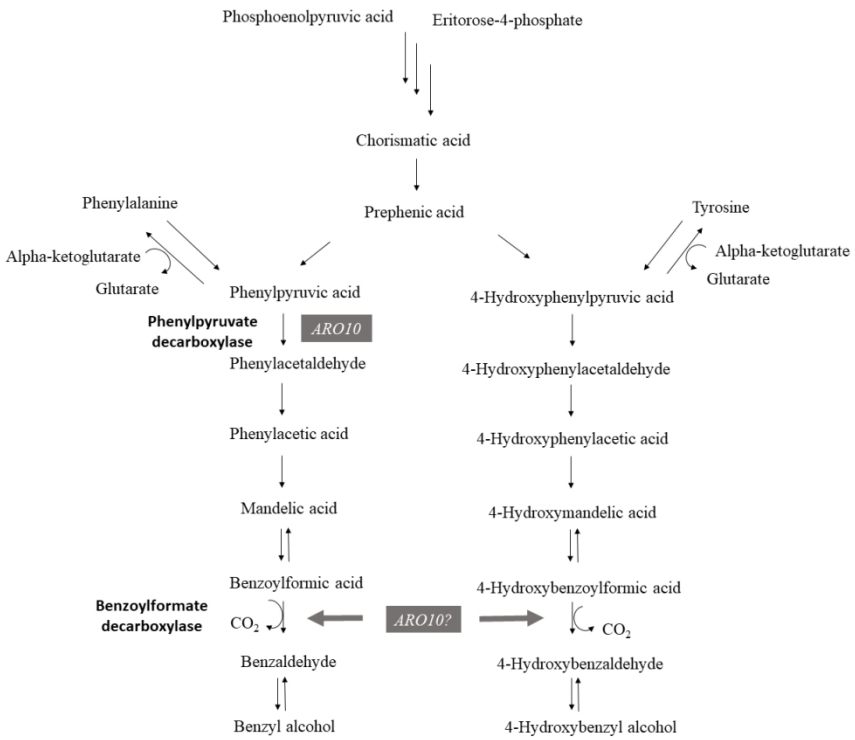

Fig 1. Proposed biosynthesis route of benzyl alcohol and 4-hydroxybenzaldehyde in yeast. ScARO10 codifies phenylpyruvate decarboxylase and ScAro10p presents weak benzoylformate decarboxylase activity in S. cerevisiae. Modified from Martin et al., 2016a

338x451mm (96 x 96 DPI)

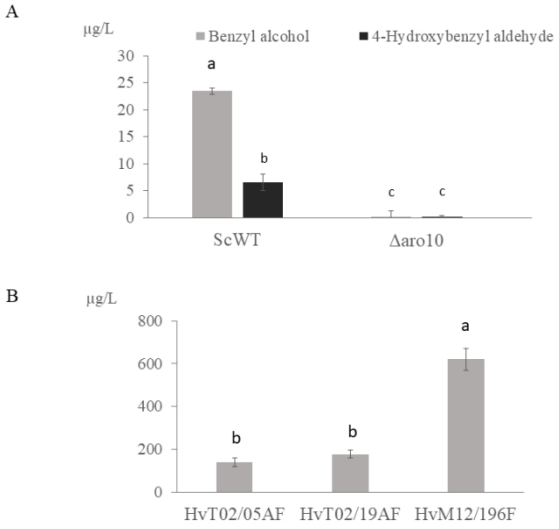

Fig 2. Benzenoids production A) by three strains of H. vineae; B) by S. cerevisiae BY4743 Wild type (WT) and S. cerevisiae Δaro10 mutant. Samples were taken after 12 days of fermentation and quantified by GC-MS. Results are expressed per litre of fermentation medium. Fermentations were performed in triplicate, bars indicate standard deviation. Letters represent significant different values p-value 0.01.

338x190mm (96 x 96 DPI)

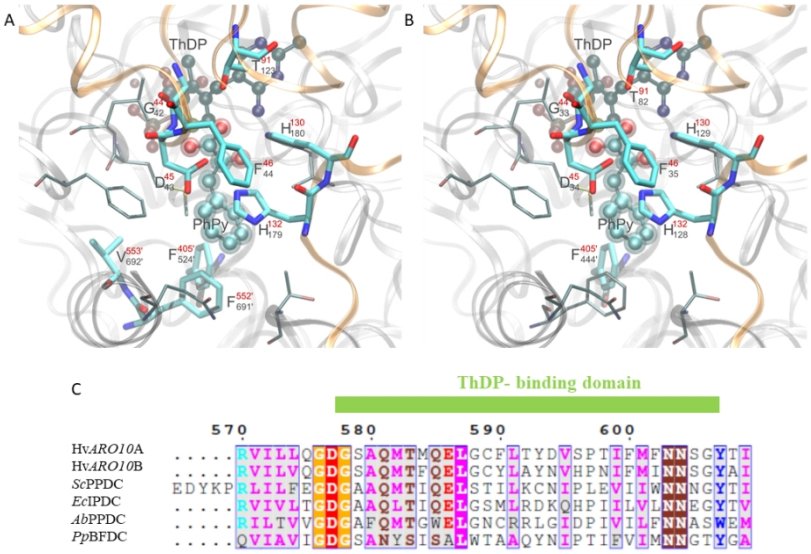

Fig 3. Conserved regions in active site and cofactor binding region in HvARO10 homologous. A) Detailed view of AbPPDC active site (PDBid 2Q5O). All residues involved in substrate (PhPy) binding are showed, but only the ones conserved in HvARO10A are highlighted and numbered in black (subscripts). The numbers in red (superscripts) correspond to AbPPDC gene and primes indicate residues belonging to the adjacent subunit in the tetrameric protein array. Thiamine di-phosphate (ThDP) cofactor is also showed in darker colours. B) Detailed view of AbPPDC active site (PDBid 2Q5O) showing conserved residues in HvARO10B. C) ThDP binding region conserved in both HvARO10 homologous and different α-keto acid carboxylases: ScPPDC phenylpyruvate decarboxylase of Saccharomyces cerevisiae, EcIPDC indol-pyruvate decarboxylase of Enterobacter cloacae, AbPPDC phenylpyruvate decarboxylase of Azospirillum brasilense PpBFDC benzoyl formate decarboxylase of Pseudomonas putida. Conserved amino acids are coloured depending on their chemical classification and degree of conservation among the sequences analysed.

338x190mm (96 x 96 DPI)

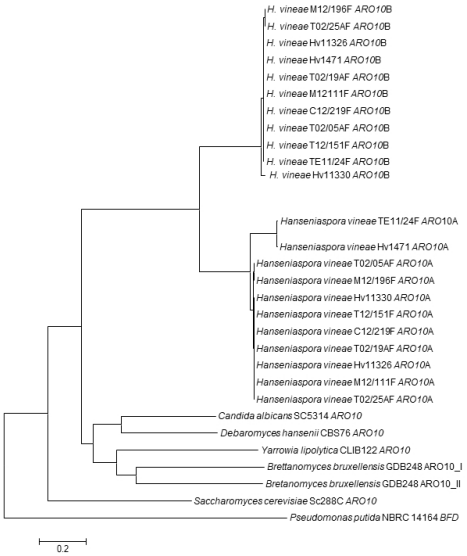

Fig 4. Dendrogram comparing different predicted amino acid sequences of HvARO10A and HvARO10B homologous with other ARO10 from different yeast clustered by Neighbour Joining method. BFDC of Pseudomonas putida was used as external group.

338x190mm (96 x 96 DPI)

1 **Supplementary material**

2

3 Table S1. Production of benzyl alcohol by eleven strains of *H. vineae* after 12 days of
4 fermentation and copy number approximation regarding to RT-PCR results. *S.cerevisiae*
5 was used as fermentation control. Benzyl alcohol values are expressed as mean of

Species	Strain	Source	Benzyl alcohol (µg/L)	Predicted copy number	
				***ARO10*A**	***ARO10*B**
S.cerevisiae	BY4743	Laboratory strain	ND[a]	-	-
H.vineae	T02/05F	Grape wine	141.18±24.61[bc]	1	1
H.vineae	T02/19F	Grape wine	179.45±7.63[bcd]	1	1
H.vineae	T02/25F	Grape wine	409.46±68.97[e]	2	1
H.vineae	TE11/24F	Grape wine	191.68±59.53[bcd]	1	1
H.vineae	Hv11326	Grape wine	140.31±12.88[bc]	1	1
H.vineae	M12/111F	Grape wine	255.30±45.41[cd]	1	1
H.vineae	T12/151F	Grape wine	164.82±31.77[bcd]	1	1
H.vineae	Hv1471	Grape wine	192.82±89.14[bcd]	1	1
H.vineae	M12/196F	Grape wine	620.27±81.86[f]	3	3
H.vineae	Hv11330	Grape wine	86.51±24.03[ab]	1	1
H.vineae	C12/219F	Grape wine	287.47±103.81[de]	2	1

6 independent fermentation triplicates ± standard deviation. Letter represents significant
7 different values p-value 0.01.

8

9

10

11 Table S2. Percent homology in amino acid sequences of Hv*ARO10* homologues and
12 different microbial α-keto acid decarboxylases: *Pp*BFDC benzoylformate decarboxylase
13 of *Pseudomonas putida*, *Ec*IPDC indole-pyruvate decarboxylase of *Enterobacter*
14 *cloacae*, *Ab*PPDC phenylpyruvate decarboxylase of *Azospirillum brasilense*, *Sc*PPDC
15 phenylpyruvate decarboxylase of *Saccharomyces cerevisiae* and *Kl*PDC pyruvate
16 decarboxylase of *Kluyveromyces lactis*.

	Hv*ARO10*A	**Hv*ARO*10B**	***Pp*BFDC**	***Ec*IPDC**	***Ab*PPDC**	***Sc*PPDC**
Hv*ARO10*A	100					
Hv*ARO10*B	51.77	100				
***Pp*BFDC**	10.04	14.58	100			
***Ec*IPDC**	19.02	22.28	17.23	100		
***Ab*PPDC**	10.80	13.81	17.80	21.38	100	
***Sc*PPDC**	21.57	25.97	12.50	23.37	13.45	100
***Kl*PDC**	20.25	23.80	13.07	35.33	20.78	29.13

17
18
19
20
21
22
23
24
25
26
27
28

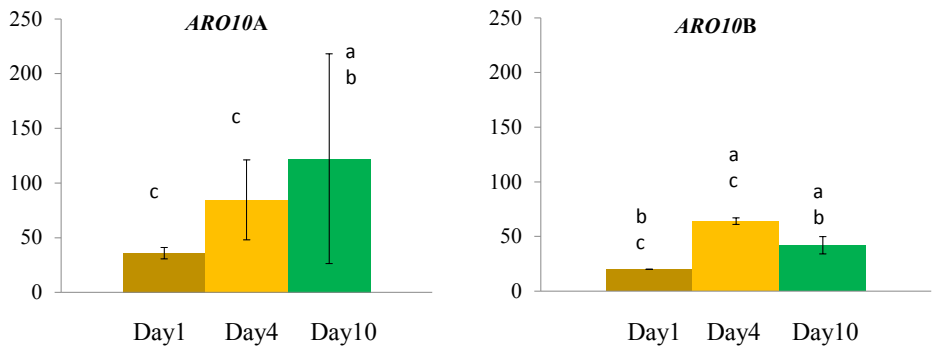

Fig S1. Relative expression *ARO10* genes from *H. vineae* at 1, 4 and 10 days of fermentation. Expression values for genes are expressed in TPM (transcripts per million) units. Bars represent standard deviation. Significative differences were calculated using FDR < 0.05

a significative differences with day 1
b significative differences with day 4
c significative differences with day 10