Genomic and phenomic analysis of *Hanseniaspora vineae* provides insights for understanding yeast fermentation flavours that contribute to wine quality.

Short title: **Genomic and phenomic analysis of *Hanseniaspora vineae***

Facundo Giorello[1], Maria Jose Valera[2], Valentina Martin[2] , Andres Parada[3], Valentina Salzman[4], Laura Camesasca[5], Laura Fariña[2], Eduardo Boido[2], Karina Medina[2], Eduardo Dellacassa[6], Luisa Berna[4], Pablo S. Aguilar[7], Albert Mas[8], Carina Gaggero[5], Francisco Carrau[2*]

[1]Espacio de Biología Vegetal del Noreste, Centro Universitario de Tacuarembó, Universidad de la República, 45000, Tacuarembó, Uruguay. [2]Area Enología y Biotecnología de Fermentaciones, Facultad de Química, Universidad de la República, 11800, Montevideo, Uruguay. [2] Instituto de Ciencias Ambientales y Evolutivas, Universidad Austral de Chile, Valdivia CP 5090000, Chile. [3]Laboratorio de Biología Celular de Membranas, Institut Pasteur de Montevideo, 11400 Montevideo, Uruguay. [4]Departamento de Biología Molecular, Instituto de Investigaciones Biológicas Clemente Estable (IIBCE), 11600 Montevideo, Uruguay. [5] Laboratorio de Biotecnología de Aromas, Facultad de Quimica, Universidad de la Republica, Montevideo, Uruguay, [6] Laboratorio de Biología Celular de Membranas (LBCM), Instituto de Investigaciones Biotecnológicas "Dr. Rodolfo A. Ugalde" (IIB) Universidad Nacional de San Martin (UNSAM), Buenos Aires, Argentina, [7] Departamento de Bioquímica y Biotecnología, Faculty of Oneology, University Rovira i Virgili, 43007 Tarragona, Spain.

*Contact email: fcarrau@fq.edu.uy

Abstract

Hanseniaspora is the main genus of the apiculate yeast group that represents about 70% of the grape-associated microflora. *Hanseniaspora vineae* is emerging as a promising species for quality wine production compared to other non-*Saccharomyces*. Wines produced by *H. vineae* with *Saccharomyces cerevisiae* consistently exhibit more intense fruity flavours and complexity than wines produced by *S. cerevisiae* alone.

In this work, genome sequencing, assembling and phylogenetic analysis of two strains of *H. vineae* shows that it is a member of the *Saccharomyces* complex and it diverged before the Whole Genome Duplication (WGD) event from this clade. Specific flavour gene duplications and absences were identified in the *H. vineae* genome, as compared to 14 fully sequenced industrial *S. cerevisiae* genomes. The increased formation of 2-phenylethyl acetate and phenylpropanoids such as 2-phenylethyl and benzyl alcohols might be explained due to gene duplications of *H. vineae* aromatic amino acid aminotransferases (*ARO8*, *ARO9*) and phenylpyruvate decarboxylases (*ARO10*). Transcriptome and aroma profiles under fermentation conditions confirmed these genes were highly expressed at the beginning of stationary phase coupled to the production of their related compounds. The extremely high level of acetate esters produced by *H. vineae* compared to *S. cerevisiae* is consistent with the identification of six novel proteins with alcohol acetyltransferase (AATase) domains. The absence of the branched-chain-amino-acid transaminases (*BAT2*) and acyl-CoA/ethanol *O*-acyltransferases (*EEB1*) genes, correlates with *H. vineae* reduced production of branched-chain higher alcohols, fatty acids and ethyl esters, respectively. Our study provides sustenance to understanding and potentially utilizing genes that determine fermentation aromas.

Key words: genome, transcriptome, metabolome, wine aroma, flavor compounds

78 **Importance**
79
80
81 The huge diversity of non-*Saccharomyces* yeasts in grapes is dominated by the apiculate
82 genus *Hanseniaspora*. Two native strains of *H. vineae* applied to winemaking due to their high
83 oenological potential in aroma and fermentation performance, were selected to obtain high quality
84 genomes. Here, we present a phylogenetic analysis, and the complete transcriptome and aroma
85 metabolome of *H. vineae* during three fermentation steps. This species produced significantly richer
86 flavour compound diversity compared to *Saccharomyces*, such as benzenoids, phenylpropanoids,
87 and acetate derived compounds. The identification of six proteins, different from *S. cerevisiae ATF*,
88 with diverse acetyl transferase domains in *H. vineae* offers a relevant source of native genetic
89 variants for this enzymatic activity. The discovery of benzenoid synthesis capacity in *H. vineae*
90 provides a new eukaryotic model to dilucidate an alternative pathway to that catalysed by plants'
91 phenylalanine lyases.
92

93 **Introduction**

94 It is well known that yeast transforms grape sugars to ethanol and CO_2 as the main wine
95 fermentation products; however, cell secondary metabolism generates the highest impact
96 compounds that dramatically affect the final flavour of wine. Flavour traits matter most in
97 fermented beverages and should be considered the key properties when developing yeast selection
98 within food biotechnology industries (1, 2). In wine, non-*Saccharomyces* yeast strains that account
99 for more than 99% of the grape native flora are still poorly explored (2), and their impact on flavour
100 richness will require multidisciplinary studies from genetics to metabolomic analyses of yeast cells.
101 The limited number of commercial yeast strains applied by today's winemakers are not contributing
102 to flavour diversity, decreasing the possibilities to obtain more differentiated wine styles. Besides
103 the grape selection, viticulture and vinification technologies used, that have been traditionally
104 emphasized for quality wine production, yeast aspects should be taken into account. In a highly

105 competitive market with more than one million brands of wines, differentiation and increased
106 flavour diversity will be obtained with application of increased yeast diversity during the
107 fermentation process. Non-*Saccharomyces* species of yeast have been reported as beneficial for
108 winemaking because they contribute to sensory complexity of wines (3, 4). The main non-
109 *Saccharomyces* genus associated with grapes is *Hanseniaspora*. Among the species comprised in
110 this genus, *H. vineae* is one of the most promising with high oenological potential (5). Recently, *H.*
111 *vineae* demonstrated the ability to provide increased levels of acetate esters and benzenoids, and
112 decreased levels of higher alcohols (except benzyl and 2-phenylethyl alcohols) in wines by pure
113 fermentation or by co-fermentation with *S. cerevisiae* (6-10). Aroma sensory analysis of wines,
114 attributed to *H. vineae* winemaking, indicated a significant increase in fruit intensity described as
115 banana, pear, apple, citric fruits and guava (9). These favourable characteristics for the winemaking
116 industry have turned *H. vineae* into a species increasingly regarded as a means to improve aroma
117 quality (5). Flavour diversity, including subtle characteristic differences in fine wines, has been
118 described for other non-*Saccharomyces* species such as *Pichia, Metschnikowia* and *Torulaspora* (2,
119 4, 11). Various genes have been identified as contributors to higher alcohol, acetate ester and ethyl
120 ester biosynthesis in *S. cerevisiae*; however, other species remain uncharacterized in this regard
121 (12).

122 Higher alcohol formation via the Ehrlich pathway is subdivided into three steps: transamination,
123 decarboxylation and reduction (Fig. 1). In transamination, the key enzymes are the branched-chain
124 amino acid transaminases (encoded by *BAT* genes) and the aromatic amino acid aminotransferases
125 (encoded by *ARO8* and *ARO9* genes), which catalyse the transfer of amines between amino acids
126 and their respective α-keto acid. In the second step, the α-keto acids are decarboxylated through
127 pyruvate decarboxylases (encoded mainly by *PDC* and *ARO10* genes) to form the respective
128 aldehydes. Finally, the reduction from aldehydes to alcohols is carried out by alcohol
129 dehydrogenases (encoded by *ADH* genes) and aryl-alcohol dehydrogenases (encoded by *AAD*
130 genes). The formation of the fruity- and flowery-like aroma acetate esters is dependent on acetate

and alcohols, and they are due to two alcohol acetyltransferases (AATases) *ATF* in *S. cerevisiae*. The biosynthesis of ethyl esters is carried out by two acyl-CoA/ethanol *O*-acyltransferases (encoded by the *EEB1* and *EHT1* genes) and involves ethanol and acyl-CoA units (derived from fatty acid synthesis). Ethyl esters as well as acetate esters contribute fruity-like aromas, although their concentration levels in wine are significantly lower than those of the acetate esters (14-18).

In this work, genomic, transcriptomic and metabolomic analyses of the novel and native yeast for winemaking (*H. vineae*) was conducted and the results were compared with those of *S. cerevisiae* to understand the aroma compound differences produced. We identified several changes in the dosage of key genes involved in higher levels of alcohol, fatty acids, acetate esters and ethyl esters biosynthesis in *H. vineae*. We analysed the expression profiles of these genes through transcriptomics and assessing the concentration of several aroma compounds during three different phases of the *H. vineae* fermentation process. A comparative work that analysed the genomic, transcriptomic and metabolomic profiles of a member of the apiculate group of the Saccharomycodaceae yeast family is presented. Understanding the alternative metabolic pathways of *H. vineae* as compared to *S. cerevisiae* will contribute to understanding of apiculate yeast biology, which is the main yeast group associated with fruits (17-18).

Results and Discussion

Yeasts analysed in this work are shown in Table 1 and putative genes and codes related to aroma synthesis by *S. cerevisiae* are described in supplemental materials Dataset S1.

Genome characterization of *H. vineae*

Two strains of *H. vineae* most used at the winemaking level by our group since 2009 were selected for genome sequencing: T02/19AF and T02/05AF. Sequencing of both strains was performed on an Illumina Genome Analyzer IIx platform.

Genome analysis revealed high similarity in both genomes in size and prediction of genes (Table 2; Table S1; Fig. S1a). Therefore, only data obtained from T02/19AF genomes are specified below in

157 detail.

158 The sequencing run generated a mean of 13,302,566 paired-end reads (2 x 100 cycles). After
159 filtering and removing redundant reads, a final set of 9,203,956 reads was used for genome
160 assembly. A total of 87 scaffolds with a median length of 76,832 base pairs were assembled through
161 MaSuRCA software, yielding a genome (haploid) of 11.3 Mb, representing an average coverage of
162 163-fold, with an N50 of 261 Kb and a GC content of 37% (Table 2; Table S1; Fig. S1a). Higher
163 quality data and a more extensive analysis of the genome of *H. vineae* was obtained than in our
164 previous report (19).

165 Genome size and ploidy level were also addressed by flow cytometry (FCM) analysis using linear
166 plots of fluorescence intensity of cell populations stained with propidium iodide (PI). This
167 technique discriminated two cellular subpopulations with different DNA contents, namely R1 and
168 R2 (Fig. S2). All tested samples presented a half-peak coefficient of variation of R1 of less than
169 10% (data not shown), indicating high-resolution DNA measurements. As references for genomic
170 DNA estimation, we used both *S. cerevisiae* haploid (BY4742) and diploid (BY4743) strains,
171 containing genomes of 11.67 and 23.35 Mb, respectively (20). A concurrent FCM analysis of *S.*
172 *cerevisiae* haploid and diploid strains revealed three distinct peaks (Fig. S2), corresponding to 1n,
173 2n and 4n DNA contents, where the mean PI fluorescence intensity of each peak was directly
174 correlated ($r^2 > 0.999$) to the amount of DNA of its corresponding cell subpopulation (Fig. S2). The
175 genome size of each *H. vineae* strain was estimated in accordance with the R1 cell subpopulation
176 (Fig. S2). The analysis by FCM revealed a diploid genome size of 16.71+/- 0.79 Mb (Table S1).
177 Regarding gene copy number, we expected that the *H. vineae* genome would show a certain (but
178 unknown) level of ploidy given its sporulation capacity (21). In any case, diploidy of both strains
179 was confirmed. However, the slight difference in genome size obtained by FCM and our genome
180 assembly-based calculations could be explained by the principles of the technique. *H. vineae*
181 genome size was estimated using *S. cerevisiae* as the control strain. Because cells themselves can
182 act as a lens, changes in cell size or shape can affect the PI fluorescence detected by FCM (22),

resulting in differences in genome size estimations obtained by FCM versus sequencing.

A total of 4,708 gene models were predicted using Augustus software, out of which 3,855 had at least one Pfam domain of the Pfam platform databases. We identified 3,861 sequences homologous to *S. cerevisiae* genes and more than 4,141 sequences aligned to the National Center for Biotechnology Information (NCBI) non-redundant protein database (Table 2; Table S1). Due to the presence of different number of homologous genes than those reported for *S. cerevisiae* strain S288c in *H. vineae*, an Augustus prediction number (gXXXX.t1) is provided to clarify the putative gene, which was analysed in each case.

We identified 243 of the 248 core eukaryotic genes (CEGs) and 445 of the 458 CEGs from the Augustus predictions, showing that our genome is ~98% complete. Interestingly, the protein identity between *H. vineae* and *S. cerevisiae* is only 52%, demonstrating a great divergence between these two species. Moreover, a high heterozygosity level was evidenced by single-nucleotide polymorphism (SNP) analysis using different *S. cerevisiae* strains (24). A total of 56,662 SNPs (1 heterozygous SNP per 200 bp) were found, of which 30,740 SNPs (54%) were present in coding sequences (Fig. S1b). According to the high genetic similarity found between T02/19AF and T02/05AF the nucleotide diversity between both *H. vineae* strains was 1 variant per 179 pb (63,021 SNPs), a closely similar rate to those found among different *S. cerevisiae* strains (24, 25).

Genes related to yeast aroma compound synthesis in *H. vineae* were compared with those reported for *S. cerevisiae* (see Tables 3 and S2). Absent homologies and repeated genes were found.

H. vineae diverged before the WGD clade of the _Saccharomyces_ complex

In order to determine the phylogenetic position of *H. vineae*, a phylogenetic tree was inferred by concatenating 227 genes from 29 species. The selection of these proteins was done by orthologous alignment of the predicted proteins from the *H. vineae* genome compared with those from yeast species obtained from databases. The maximum likelihood phylogeny classifies *H. vineae* as part of *Saccharomyces* complex but out of the WGD clade with very high support (Fig. 2). *H. vineae* is recovered as the sister taxa to two lineages, one composed of WGD yeasts: *Kazachstania africana*,

209 *Naumovozyma dairenensis*, *Saccharomyces cerevisiae*, *Candida glabrata*, *Torulaspora delbrueckii*,
210 *Zygosaccharomyces rouxii* and *Tetrapisispora blattae*; and the other composed of species diverged
211 before the WGD, to include *Ashbya gossypii*, *Eremothecium cymbalariae*, *Lachancea*
212 *thermotolerans* and *Kluyveromyces lactis*. Node support for this placement of *H. vineae* was very
213 high (IC = 1.0 and BS = 100).

214 A total of 372 orthologous groups were expanded in *S. cerevisiae* compared to *H. vineae*, which
215 involved 427 genes. These genes have a 2:1 relationship between *S. cerevisiae* and *H. vineae*,
216 supporting the theory that *H. vineae* diverged previously to the WGD and arose out of the fungal
217 CTG clade formed by yeasts that present differences in their genetic code (26). Although this
218 phylogeny presents some differences to that previously reported for *Saccharomyces* complex (27,
219 28), the phylogenetic position of *H. vineae* presents a high node support and is similar to that
220 obtained by Kurtzman and Robnett (27). Phylogenies inferred by these authors were based on
221 divergence in genes of the rDNA repeat (18S, 26S, ITS), single copy nuclear genes (translation
222 elongation factor 1α, actin-1, RNA polymerase II) and mitochondrially encoded genes (small-
223 subunit rDNA, cytochrome oxidase II).

224 **Overview of transcriptome dynamics during fermentation**

225 In order to perform a comprehensive analysis, we obtained transcriptomic profiles of *H. vineae*
226 strain T02/19AF along three different days of the fermentation process (Fig. S3): exponential
227 growth phase (day 1), end of exponential phase (day 4) and end of stationary phase (day 10).
228 Fermentations were carried out in triplicate using Erlenmeyer flasks of 250 mL with 125 mL of
229 chemically defined grape (CDG) medium that presents similar nutrient composition to grape juice.
230 The medium was supplemented with 100 mg N/L yeast available nitrogen (YAN), 200 g/L of an
231 equimolar mixture of glucose and fructose and pH was adjusted to 3.5.
232 These data were analysed to compare the expression of the key genes related to the flavour
233 compounds present in *H. vineae*, and, moreover, to those extra-copies identified exclusively in *H.*
234 *vineae* and not in *S. cerevisiae*. Transcriptome sequencing of nine libraries was done in three

replicates for the three fermentation points, using Illumina. Similar quantities of genes were expressed at the three fermentation points although their expression levels differed considerably. More than 2,500 (~56%) genes were differentially expressed according to the false discovery rate calculation (FDR < 0.05) between each pair of fermentation time points.

Transcriptome assembly allowed the identification of 15 more genes than those obtained by genome analysis, and almost all paralogous genes identified within the genome were confirmed. The transcriptome analyses for days 1, 4 and 10 presented 4,596, 4,558 and 4,468 expressed genes, respectively, of which 4,468 were in common (Figs. 3a, b). The most significant gene ontology (GO) terms associated with the genes shared between the three fermentation points were tRNA processing for biological processes, GTPase regulator activity for molecular function and Golgi apparatus for cellular component.

High number of differentially expressed genes in *H. vineae* during fermentation

For the three fermentation points (1, 4 and 10 days), the differentially expressed genes were analysed using edgeR software. Important changes in gene expression were detected between any pair of the three fermentation points, while the differences between replicates were minimal (Fig. S4).

In *H. vineae*, the large number of differentially expressed genes (DEGs) identified along the fermentation process was remarkable. Of the 4,468 genes shared among the three days, more than 2,500 (~56%) were differentially expressed between each point (FDR < 0.05). However, microarray studies of various *S. cerevisiae* strains reported a smaller number of DEGs, ranging from 1,000 to 1,500 genes (29). The largest number of DEGs was identified between the first and last point; on the other hand, the fewest number were detected between day 4 and day 10. This situation is consistent with the fact that day 4 is the start of a stationary phase and at day 10 the stationary phase is ending. As the fermentation process approaches the stationary phase, fewer genes are expected to be differentially expressed.

Unique and expanded orthologous groups in *H. vineae* compared to *S. cerevisiae*

261 Using OrthoMCL software, 85 expanded orthologous groups were detected in *H. vineae* compared
262 to data from the *S. cerevisiae* S288c sequence.
263 There is consistently higher expression at day 1 and 4 of genes related to growth biochemical
264 cascades (such as amino acid biosynthesis, pentose phosphate pathway, oxidative phosphorylation
265 and TCA) and glycolysis (such as pyruvate metabolism and synthesis of secondary compounds)
266 (Fig. 3c; Table S3). However, at day 10 the protein turnover genes were the most expressed, as at
267 the middle and end of fermentation amino acids are generally exhausted from the medium. The
268 expression of genes related with protein processing at the end of the stationary phase could be
269 related to autophagy processes. Autophagy in yeast is a response to nutrient limitation, and the
270 endoplasmic reticulum and GPI anchor mechanism are activated under this stress situation for the
271 recovery process of proteins (30, 31). Interestingly, methane metabolism genes were mainly
272 expressed at exponential growth to early stationary phase (days 1 to 4), but this could be specific to
273 *Hanseniaspora* yeasts, as they are a methylotrophic group that may be active when oxygen is
274 present at the beginning of fermentation (32).
275 The most complete KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathways were those
276 related to tyrosine and phenylalanine metabolism, both aromatic amino acids that are related to
277 phenylpropanoid synthesis (Table S4a). The main genes that are exclusive to *H. vineae* belong to
278 the following KEGG modules: β-lactam resistance and lysine biosynthesis with five genes, and
279 bacterial proteasome and benzoate degradation, with three genes (Table S4b). Five serine
280 endopeptidases that could be involved in diverse functions were related to the β-lactam resistance
281 module, while for lysine biosynthesis two aldehyde dehydrogenases (g3618.t1–g3619.t1), an
282 unknown (4147.t1), one mlo2 like protein (g2280.t1) and ssm4 (g570.t1) proteins were found.
283 **Genomics and yeast flavours**
284 Several genetic and phenomic characteristics were analysed comparing *H. vineae* and *S. cerevisiae*
285 strains. The comparative genomics analysis included *H. vineae* and up to 14 wine industry strains of
286 *S. cerevisiae* whose genomes were analysed in previous studies (33, 34) (Table S5). Aroma

compound profiles determined by GC-MS (Gas Chromatography- Mass Spectrometry) of *H. vineae* were compared with those of *S. cerevisiae* strain M522 and are shown in Table 5. On the other hand, in Table 6 aroma compounds produced by *H. vineae* at days 4 and 10 are shown and differential expression of genes involved in higher alcohol, acetate ester and ethyl ester metabolism were evaluated (Fig. S5). These results are discussed in the following sections.

Alcohols and 2-phenylethanol

The aroma compound analysis shows that overall higher alcohol production is more than twice as high in *S. cerevisiae* M522 than in *H. vineae* (Fig. 4a; Table 5). In fact, other studies comparing *H. vineae* to the wine yeast *S. cerevisiae* EC1118 have found similar results (6). However, the proportion of 2-phenylethanol in *H. vineae* with respect to *S. cerevisiae* M522 is approximately equivalent (Fig. 4a) if 2-phenylethyl acetate is taken into account as a derived compound of 2-phenylethanol.

The three steps of higher alcohol biosynthesis (transamination, decarboxylation and reduction) (Fig. 4b) were analysed attending to transcriptomics and phenomic results.

Transamination. In *S. cerevisiae*, the most important gene involved in transamination leading to the production of higher alcohols is *BAT2* (34), which encodes the branched-chain amino acid aminotransferase. *BAT2* is absent in the *H. vineae* genome. This could explain the reduced presence of overall branched-chain higher alcohols in *H. vineae* fermentations compared to *S. cerevisiae* M522. In this scenario, the *BAT1* gene in *H. vineae* would perform the two reactions of the reversible transamination step. *BAT1* showed higher expression levels on day 1 and a decay in expression on days 4 and 10, while overall alcohol levels remained constant (Figs. 4d and S5). Therefore, the production of alcohols seems to occur early in fermentation, preceded by expression of this gene.

On the other hand, the amount of 2-phenylethanol/2-phenylethyl acetate remains constant between days 4 and 10, while the expression of the *ARO*8 and *ARO9* genes reaches a peak by day 4 (Fig. 4c,d; Fig. S5). *S. cerevisiae* industrial strains present only one copy of these *ARO* genes (Table S5);

313 however, *H. vineae* presents three copies of *ARO8* and four of *ARO9* that are all very similarly
314 expressed during fermentation (Fig. 4d; Fig. S5). *ARO8* and *ARO9* encode aromatic amino acid
315 transaminases, which can act as broad-substrate-specificity amino acid transaminases in the Ehrlich
316 pathway (16), and they are involved in the anabolism and catabolism of the aromatic amino acids
317 phenylalanine and tyrosine. These data are in agreement with the KEGG pathways over-represented
318 in *H. vineae* as shown in Table S4a. Therefore, the overexpression of these two expanded genes
319 could explain the larger proportion of 2-phenylethanol in two ways: first, for their incremented
320 specificity for aromatic amino acids present in the medium; and second, for an increased synthesis
321 of phenylalanine that is known as 2-phenylethanol precursor (36).
322 **Decarboxylation.** Five genes are involved in the decarboxylation step in *S. cerevisiae* (*PDC1*,
323 *PDC5*, *PDC6*, *ARO10* and *THI3*) (16), of which *H. vineae* has two copies of *ARO10* and two of
324 *PDC1*. The most highly expressed paralogous copy of *PDC1* had an expression pattern similar to
325 that of *BAT1*, on day 1, prior to alcohol production (Fig. 4c, d; Fig. S5a).
326 It is possible that *ARO10* duplication (Table 3) enables an efficient decarboxylation of aromatic α-
327 keto acids derived from the enhanced transamination step. In fact, this is supported by the
328 expression profile (Fig. 4d; Fig. S5a) of both *ARO10* genes that are very similar to the expression
329 profiles found for *ARO8* and *ARO9* copies. It should be noted that the co-fermentation of *H. vineae*
330 with *S. cerevisiae* resulted in an increased intensity of citrusy aromas of which 2-phenylethanol
331 (and therefore the *ARO* gene duplications) could be responsible (9). All 14 *S. cerevisiae* industrial
332 strains only showed one copy of *ARO10* (Table S5). Further, *ARO10* has been shown to be related
333 to production of benzyl alcohol in a putative metabolic pathway of mandelate (7). Therefore, this
334 decarboxylase activity could be involved in the enhanced synthesis of more than two orders of
335 magnitude of benzylic alcohol in *H. vineae* compared to *S. cerevisiae* (Table 5).
336 **Reduction to higher alcohol.** Surprisingly, *H. vineae* did not contain homologous sequences or any
337 transcriptional evidence of the seven aryl alcohol dehydrogenases (*AAD* genes) present in the *S.*
338 *cerevisiae* S288c sequenced genome (Table 3). This activity catalyses the chemical reaction

between aromatic aldehydes and alcohols. Given the overproduction of benzyl and 2-phenylethyl alcohol (precursor of the 2-phenylethyl acetate) in *H. vineae*, compared to *S. cerevisiae* M522 (Table 5), at least one aryl-alcohol dehydrogenase protein would be expected. However, it should be noted that the final step of the Ehrlich pathway (higher alcohol formation) can be catalysed by any one of the ethanol dehydrogenases (Adh1, Adh2, Adh3, Adh4, and Adh5) or by Sfa1 (a formaldehyde dehydrogenase) in *S. cerevisiae* (37). In *S. cerevisiae*, the alcohol dehydrogenases are present in two multigenic families, with four genes in each one (according to Ensembl): *ADH6*, *ADH7*, *YAL061W* and *YAL060W* in one family, and *ADH1*, *ADH2*, *ADH3*, *ADH5* in the other. *H. vineae* presents two copies of *ADH1* and *ADH3*, and four of *ADH6*, totalling eight genes, as in *S. cerevisiae* (Table 3). The *ADH4* gene does not belong to either of these multigenic families and is absent in *H. vineae*. In *H. vineae* not all paralogous copies of *ADH* genes showed a significant transcriptional activity (many paralogous copies were assembled but they were filtered out before differential expression analysis). Interestingly, two of the four paralogous copies of *ADH6* of the four gene copies found in *H. vineae* were not expressed in these conditions (Fig. S5a).

In regard to the expression levels, four of the other five alcohol dehydrogenase genes, as well as *SFA1*, are significantly more expressed on days 1 and 4, while just one copy of *ADH1* and *ADH3* is more expressed on days 4 and 10 (Fig. 4d; Fig. S5a). One of the *ADH6* copies showed a significant decline in expression levels between days 1 and 4, which is consistent with that previously reported for *S. cerevisiae* (29). In contrast, one *ADH3* gene copy showed a two-fold increase in expression by day 4 relative to day 1, similar to that of the *AAD10* and *AAD14* genes during *S. cerevisiae* wine fermentation (29). As a result, we suggest that the *ADH* genes that may be replacing the *AAD* genes might be those that show the same expression profile found in *S. cerevisiae*. Further biochemical studies will be necessary to confirm this suggestion.

Acetate esters

H. vineae and *S. cerevisiae* M522 also showed notable differences in overall acetate production, whereby *H. vineae* produced concentrations one order of magnitude higher than *S. cerevisiae* (Fig

4a). As mentioned, *H. vineae* also showed a larger turnover from 2-phenylethanol to 2-phenylethyl acetate than did *S. cerevisiae*. For example, 2-phenylethyl acetate only constituted a small fraction in *S. cerevisiae* of the total 2-phenylethanol produced, compared to *H. vineae* (Fig. 4a; Table 5).

With regard to the genes involved in acetate ester formation, the *H. vineae* genome presented a highly divergent putative ortholog of the *S. cerevisiae ATF2* gene, and it did not present any sequences homologous to *ATF1*. However, there were also five predictions containing the AATase Pfam domain. The four *SLI1* N-acetyltransferase homologues are repeated in tandem in the *H. vineae* genome (one of them is out of transcriptomic analysis according to threshold evaluation; Fig. 5d). Three of these genes that are highly expressed in *H. vineae* (Fig. 5d) have weak similarity (22–24% at the amino acid level) with *S. cerevisiae SLI1*, which is a unique copy gene with N-acetyltransferase activity. It is known that *SLI1* has wide specificity for aromatic amines similar to the *ATF* genes (38). The other *H. vineae* AATase predicted (g4599.t1) has no homology with any *S. cerevisiae* gene previously reported; however, is the most highly expressed gene at the end of stationary phase (Fig. 5d). The *ATF2* gene and the most highly expressed *SLI1* gene copy were both highly expressed on day 4, which can explain the notable two-fold increment of acetate esters between days 4 and 10 (Fig. 5c; Fig. S5b). Curiously, only *S. cerevisiae* strain M522 did not present the *ATF1* gene, while none of the 14 industrial *S. cerevisiae* strains showed more than one gene similar to *SLI1* (Table S5).

Therefore, the presence of six sequences with AATase domains (one *ATF2*, four *SLI1* and g4599.t1) could explain why *H. vineae* produces significantly more acetate esters than *S. cerevisiae*. The higher turnover of 2-phenylethanol to its corresponding acetate esters in *H. vineae* compared to *S. cerevisiae* clearly suggests that some of the *H. vineae* AATases (e.g., *SLI1* paralogs) might be specific for this aromatic alcohol. The increased level of acetate esters in *H. vineae* can explain the more intense fruity aroma resulting from the fermentation of *H. vineae* in Chardonnay (9) and Macabeo (8) wines. In fact, other apiculate yeasts from the *Hanseniaspora* genus are higher acetate esters producers compared to *S. cerevisiae* (39). However, 2-phenylacetate high production is a

particular characteristic of *H. vineae* compared to other species of this genus (40). Other *Hanseniaspora* species commonly produce increased levels of ethyl acetate. It is noteworthy that with regard to information about sequenced genomes of other *Hanseniapora* species available in databases (40, 41), most do not present *SLI1* homologous sequences. The exception is *H. osmophila* with two putative *SLI1*. The detection of six AATases in *H. vineae* provides a relevant higher number of proteins for acetate esters biosynthesis compared with the three copies of *S. cerevisiae*. These variations might contribute to improved functional designs for 2-phenylethanol acetylation and other phenylpropanoid aroma compound synthesis, which are scarce pathways in *S. cerevisiae* strains.

Ethyl esters

EEB1 and *EHT1* code for ethanol *O*-acyltransferases responsible for medium-chain fatty acid ethyl ester biosynthesis in *S. cerevisiae* (44). A decrease in the production of ethyl esters was observed in *H. vineae* compared to *S. cerevisiae* M522 (Fig. 6; Table 5). Furthermore, the absence of one of the main genes involved in ethyl ester production (*EEB1)* in *H. vineae* is consistent with this result (Fig. 6b). Only three strains of *S. cerevisiae* did not present this gene (Table S5). *EHT1* is present in the *H. vineae* genome and it is highly and significantly expressed on days 1 and 4 relative to day 10 (Fig. 6c; Fig. S4). This might be consistent with the fact that esterified fatty acids were quantified on day 10 but were not detected on day 4 (Table 6).

Even so, an important inter-strain expression of acyltransferases was found during the fermentation process in *S. cerevisiae*. In general, the expression of *EHT1* in *S. cerevisiae* increased somewhat as fermentation progressed (30), which differs from our findings in *H. vineae*. Regarding our data, ethyl ester compounds were detectable on day 10 of the fermentation process. Here, it should be noted that our results are consistent with those obtained with this species in wines of Chardonnay (9) and Macabeo (8) fermentations, in which they exhibited decreased levels of ethyl esters compared to acetate esters.

Conclusion

417 The use of non-*Saccharomyces* yeasts in winemaking is limited due to the insufficient
418 characterization of many species that naturally participate in these processes. *H. vineae* has proved
419 to contribute with flavour diversity in winemaking conditions. Here we present a deep genomic,
420 transcriptomic and metabolomic analyses, and their comparison with *Saccharomyces* strains data.
421 Based on our results in a synthetic simil-grape juice medium, this work represents a relevant
422 contribution to understand the biology and phylogenic relationship of the main yeast genus
423 associated with grapes. The larger production of acetate esters, the increased ratio of 2-phenylethyl
424 acetate to 2-phenylethanol and the reduced amount of ethyl esters found in *H. vineae* may be due to
425 the high presence of putative alcohol acetyltransferase proteins and the absence of *EEB1*. These
426 results are in agreement with previous reports studied in real winemaking conditions. As it was
427 shown, *H. vineae* produced a large amount of phenylpropanoids compared to *S. cerevisiae* and other
428 yeasts that might be explained by gene duplications and highly expressed *ARO* genes. This work
429 established that *H. vineae* may be a potential model eukaryotic species to study benzenoids
430 synthesis pathways, an alternative to the phenylalanine ammonia lyase (PAL) pathway commonly
431 found in plants and Basidiomycetes. These phenolic volatile compounds have several known key
432 functions in plants, such as cell-cell communication, antimicrobial activity or phytohormone
433 production that make them highly attractive to the yeast biotechnology industry.
434
435 **Materials and Methods**
436 **Yeasts**
437 Table 1 shows all the yeast strains utilized in this work.
438 **Genomic characterization of *H. vineae***
439 DNA and RNA isolation from *H. vineae* strains: *H. vineae* T02/19AF and T02/05AF strains were
440 isolated from the Uruguayan Tannat vineyards. These strains were identified as *H. vineae* by
441 sequencing the ribosomal D1/D2 region, and the strains were differentiated using the tandem
442 repeats t-RNA PCR technique (44). Genomic DNA was obtained from *H. vineae* cultures grown in

443 yeast extract, peptone (YP) media (1% yeast extract and 2% peptone, supplemented with 2%
444 glucose) at 30 °C, using the Wizard Genomic DNA Purification Kit (Promega, NY, USA),
445 according to the manufacturer's instructions. Total RNA was obtained from *H. vineae* T02/19AF
446 strain grown in static batch fermentation conditions using the Ribopure RNA Purification Kit Yeast
447 (Life Technologies Grand Island, USA). The poly-A mRNA fraction was then isolated using the
448 Oligotex mRNA Mini Kit (Qiagen Hilden, Germany) and converted to indexed RNAseq libraries
449 with the ScriptSeq™ v2 RNA-Seq Library Preparation Kit (Epicentre Biotechnologies, Madison,
450 WI, USA).

451 Genome length and ploidy estimation by flow cytometry (FCM): *H. vineae* strains were grown in
452 YP media supplemented with 2% glucose. 1 X 10^7 cells were pelleted at 3,000 x g for 3 min and
453 washed with ice-cold phosphate-buffered saline (PBS) with 138mM NaCl, 3mM KCl, 8.1mM
454 Na_2HPO_4 and 1.5 mM KH_2PO_4. To fix cells 1 ml of 70% cold ethanol was slowly added and
455 samples were stored at 4°C overnight. After removing ethanol by centrifugation the cell pellet was
456 washed with PBS and resuspended in 700 µl of this buffer. Each sample was sequentially treated
457 with 250 µl of 1 mg/ml Ribonuclease A (Applichem, USA) (1 h at 50°C), 50 µl of Proteinase K
458 (Sigma-Aldrich, USA) 20 mg/mL (1 h at 50°C) and incubated overnight at 4°C in the dark with 50
459 µl of propidium iodide (PI) 1mg/ml (Life Technologies, USA). Analysis of DNA content by FCM
460 requires staining yeasts with PI, a fluorochrome that binds to DNA.

461 FCM analyses were performed using a CyAn™ ADP LX, 7 colour (Beckman Coulter, USA) flow
462 cytometer. The blue laser (488 nm) was selected to excite the PI fluorophore. Fluorescence area
463 signal was detected with a 575/25 nm (FL2) emission filter and plotted on a linear scale. Data
464 acquisition and analysis were achieved using Summit v4.3 software (Dako Cytomation, UK), and
465 10,000 events per sample were collected. The gating strategy comprised a forward scatter (FSC) vs.
466 side scatter (SSC) cell region that excluded cellular debris and irrelevant small particles. This region
467 was applied to a PI histogram so that only gated events were displayed. *S. cerevisiae* strains
468 BY4742 and BY4743 (Table 1) were used as controls. The mean fluorescence intensity of stained

cells as measured by FCM was taken as indicative of total DNA content, and a direct correlation between fluorescence intensity measurements and the amount of DNA in each control strain was established. All cultures generated bimodal fluorescence profiles composed of two peaks: one corresponding to a population of a majority of cells in G phase (lower intensity peak) and the other (higher intensity peak) attributed to cells in S-phase undergoing DNA synthesis. The genome size of each *H. vineae* strain was estimated in accordance with the mean fluorescence of the peak subpopulation that showed lower intensity values. Three independent biological experiments were performed, and samples were analysed in triplicate for each experiment.

Genome assembly and gene annotation: Genomes were sequenced using an Illumina Genome Analyzer IIx platform in paired end mode. A shotgun genomic library was generated based on standard methods.

Reads were filtered and trimmed with the QC Toolkit (45). The first 15 bases at the 5′ end and the last bases of the 3′ end with a phred quality smaller than 30 were trimmed. The reads with an average phred quality smaller than 20 were filtered.

Digital normalization to the paired reads was applied to systematize the coverage, from uneven x200 to x30, across the genome, to gain computation efficiency and to eliminate most of the erroneous Kmer (46, 47). The *de novo* genome assembly was performed using MaSuRCA (48) (insert length = 900). To reduce heterozygosis redundancy and find any potential gene tandem repeats, HaploMerger (49) was applied using default parameters.

Gene prediction was carried out using Augustus (50) trained with *S. cerevisiae* gene models. Peptide predictions were then annotated through BLASTp (cutoff for e-value 1e-10) against *S. cerevisiae* proteins, obtained from the *Saccharomyces* Genome Database (20). The Pfam protein families database (51) was used to predict possible protein domains. To evaluate genome completeness, Core Eukaryotic Genes (CEGs) (52) were sought with BLASTp (cutoff for e-value 1e-10). Gene ontology analysis was carried out using topGO (53).

SNP identification: Genomic short reads sequences were mapped to the assembled genome of

T02/19AF using Bowtie2 in paired-end mode with default conditions (54) and processed using samtools (55) and Picard (http://broadinstitute.github.io/picard/). Through the GATK pipeline (56, 57), SNPs were identified using Unified Genotyper applying hard-filter (QD < 2.0 || FS > 60.0 || MQ < 40.0 || HaplotypeScore > 13.0 || MappingQualityRankSum < -12.5 || ReadPosRankSum < -8.0). Base pair coverage was calculated using bedtools (58). The reads of *H. vineae* T02/05AF were aligned to those of T02/19AF to estimate the nucleotide divergence between these two strains.

Analysis of 14 *S. cerevisiae* industrial wine strains

For several genes with known functions in the biosynthesis of acetate esters, ethyl esters and higher alcohols we determined which ones were present, duplicated or absent in the *H. vineae* genome, as compared to *S. cerevisiae* S288c and an additional 14 *S. cerevisiae* wine strains. These strains were selectively chosen because they are used in wine fermentation and commercial winemaking studies (Table 1).

Ortholog cluster analysis

The proteome of 31 fungal species from OrthoDB (59) was downloaded. This web service has the orthologous relationship between a broad group of predefined species. For orthologous identification, we first used pairwise BLASTp against *H. vineae* and select the reciprocal best hit. Then, we compared our orthologous group with those present on the OrthoDB database and if they contain at least one gene not belonging to the corresponding OrthoDB group they were filtered out. The protein alignment was done using MUSCLE v3.8.31(60). We used PAL2NAL (61) for aligning the nucleotides based on the protein alignment and Gblocks v0.91b (62) for eliminate poorly aligned positions. We finally obtained 227 proteins for 29 species (two species had to be discarded because we could not find the correspondence between their protein and nucleotides sequences) to recover the phylogenetic position of *H. vineae*.

To establish orthologous clusters between *S. cerevisiae* S288c and *H. vineae* T02/19AF, the predicted proteins were analysed through the OrthoMCL web server (63). Orthologous clusters

were classified as expanded in *H. vineae* if the number of *H. vineae* genes in one OrthoMCL group was larger than the number of *S. cerevisiae* genes present in that group. To identify the pathways involved in each group, *S. cerevisiae* genes were used as input on the DAVID functional annotation pipeline (64). Those orthologous cluster groups exclusive to *H. vineae* (not containing any *S. cerevisiae* sequences) were analysed using the EC enzymes and KEGG modules of the corresponding orthologous group (65) using custom Python scripts.

Phylogenetic analysis

A supermatrix tree was constructed using a set of 227 genes from 29 species, including *H. vineae*. First, FASconCAT (62) was used to concatenate the supermatrix of 214,302 bases. The problematic aligned regions were previously removed with Gblocks v0.91b (66). For this supermatrix, the best partition scheme was chosen through PartitionFinder (68). Phylogentic inference under Maximun Likelihood was done with RaxML employing a GTRCAT substitution model for each of the 32 partitions suggested by PartitionFinder and using 200 starting trees. Node support was summarized in RAxML. Bootstrap support (BS) was calculated using extended majority-rule consensus for the bootstrapped-trees set. Support is also shown as internode certainty (IC) values, a recently developed metric that considers the frequency of the bipartition defined by the internode in a given set of trees jointly with that of the most prevalent conflicting bipartition in the same tree set (68).

Transcriptome analyses

Nine transcriptomes, three replicates from three different fermentation stages, at days 1, 4 and 10, were analysed. Paired-end transcriptome sequencing was performed using Illumina MySeq. High-quality raw sequencing reads were directly assembled using Trinity (47). They yielded a total of 7.8 Gb of data and 52 million 75-bp paired reads. The transcriptomic reference constructed resulted in 4,725 contigs with an average and median length of 1,982 and 1,683 bp, respectively (Table 4).

A transcriptomic reference was constructed using the transcriptome of each sample, and an assembly was constructed by joining all of the reads for the subsequent gene expression analysis. For the construction of the transcriptomic reference, we selected the best reciprocal hit between the

contigs, among the 10 assembled transcriptomes and the subject sequences (19). The subject sequences were constructed using *H. vineae* T02/19AF protein predictions and *S. cerevisiae* proteins from the OMA Browser (69). The alignments were carried out using reciprocal BLASTx (e-value cutoff 1e-10).

Reads were aligned against the transcriptomic reference implementing RSEM (default settings) (70). The obtained expected counts for each gene were then used for the differential gene expression analysis carried out with edgeR (71). Genes with cpm < 5 in 2 samples or more per each fermentation point, were removed from the differential expression analysis. Genes with FDR < 0.05 were considered differentially expressed.

Aroma compounds analysis in a synthetic medium.

Fermentation conditions: Chemically defined grape (CDG) fermentation medium (simulating the nutrient components of grape juice but devoid of grape precursors) was prepared with the same composition to study the *de novo* formation of aroma compounds and for the transcriptome analysis by a previously described process (72) with some variations. Modifications were as follows: the total nitrogen content was adjusted to a basic amount of 50 mg of nitrogen (N)/L with each amino acid and ammonium component added in the same proportions as indicated previously (72). The final CDG medium used for inoculum preparation and fermentations was made by increasing the basic concentration by supplementation with diammonium phosphate (DAP) up to Yeast Available Nitrogen (YAN) concentration of 100 mg N/L. This YAN amount was not in limiting concentration for complete fermentation of sugars by the yeast strains used. The final pH of the medium was adjusted to 3.5 with HCl. Equimolar concentrations of glucose and fructose were added to reach a total of 200 g/L and the mixed vitamins and salts as described previously (73). Tween 80 was excluded from the medium because it was not found to be necessary for complete fermentation and it had a negative impact on the sensory characteristics of the resultant wines. Ergosterol was added as the only supplemented lipid at a final concentration of 10 mg/L.

Inocula were prepared in 10 mL of the same CDG medium by incubation for 12 h in a rotary shaker

573 at 150 rpm and 25 °C. Fermentations were carried out in 125 mL of medium contained in 250-mL
574 Erlenmeyer flasks closed with cotton plugs to simulate microaerobic conditions (74). Inoculum size
575 was 1 x 10^5 cells/mL in the final medium for all strains. Static batch fermentations were conducted
576 at 20 °C in triplicate, simulating winemaking conditions. Wine samples for GC analysis were taken
577 at days 4 and 10 during fermentation and at the end of the process. Samples were filtered through
578 0.45-µm pore membranes; SO_2 was added as 50 mg/L of sodium metabisulfite.

579 **Aroma volatile compounds**: Extraction of aroma compounds was performed using adsorption and
580 separate elution from an Isolute (IST Ltd, Mid Glamorgan, UK) ENV1 cartridge packed with 1 g of
581 a highly cross-linked styrene-divinyl benzene (SDVB) polymer. Treatment of samples and GC-MS
582 analysis were performed as described previously (4) in a Shimadzu-QP 2010 ULTRA (Tokyo,
583 Japan) mass spectrometer equipped with a Stabilwax (30 m x 0.25 mm i.d., 0.25-µm film thickness,
584 Restek) capillary column.

585 **Identification and quantification**: The components of wine aromas were identified by comparing
586 their linear retention indices with pure standards (Aldrich, Milwaukee, WI). A comparison of mass
587 spectral fragmentation patterns with those stored in databases was also performed. GC-FID and GC-
588 MS instrumental procedures using an internal standard (1-heptanol) were applied for quantitative
589 purposes, as described previously (4). All fermentations and chemical analysis were performed in
590 triplicates. ANOVA analyses were conducted to determine differences in aroma compound
591 concentrations among the strains with Statistica 7.0 (StatSoft Inc., Tulsa Oklahoma, USA).

592 **Accession Codes.**

593 This whole genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the
594 accession numbers LSNF01000001:LSNF01000741 and JFAV02000001:JFAV02000305.

595

596 **ACKNOWLEDGMENTS**

597 We wish to thank Comisión Sectorial de Investigación Científica (CSIC) CSIC Group Project No.
598 656 and CSIC Productive Sector Project No. 602 of UdelaR, Uruguay (Grant Nos ANII

599 Postgraduate POS_NAC_2012_1_9099); Agencia Nacional de Investigación e Innovación (ANII)
600 *Hanseniaspora vineae* FMV 6956 project for financial support and postdoctoral fellowship
601 PD_NAC_2016_1_133945; postdoctoral fellowship Clarín-COFUND from Principado de Asturias
602 and European Union.

603

604 **References**

605 1. Fleet GH. 2003. Yeast interactions and wine flavour. Int J Food Microbiol 86:11–22.
606 2. Carrau F, Gaggero C, Aguilar PS. 2015. Yeast diversity and native vigor for flavor
607 phenotypes. Trends Biotechnol 33:148–154.
608 3. Steensels J, Snoek T, Meersman E, Nicolino MP, Voordeckers K, Verstrepen KJ. 2014.
609 Improving industrial yeast strains: exploiting natural and artificial diversity. FEMS Microbiol
610 Rev 38:947–995.
611 4. Jolly NP, Varela C, Pretorius IS. 2014. Not your ordinary yeast: Non-*Saccharomyces* yeasts
612 in wine production uncovered. FEMS Yeast Res 14: 215–237.
613 5. Martin V, Valera MJ, Medina K, Boido E, Carrau F. 2018. Oenological impact of the
614 *Hanseniaspora/Kloeckera* yeast genus on wines - A review. Fermentation 4: 76
615 6. Martin V, Boido E, Giorello F, Mas A, Dellacassa E, Carrau F. 2016. Effect of yeast
616 assimilable nitrogen on the synthesis of phenolic aroma compounds by *Hanseniaspora*
617 *vineae* strains. Yeast 33:323–328.
618 7. Martin V, Giorello F, Fariña L, Minteguiaga M, Salzman V, Boido E, Aguilar PS, Gaggero
619 C, Dellacassa E, Mas A, Carrau F. 2016. De Novo Synthesis of Benzenoid Compounds by
620 the Yeast *Hanseniaspora vineae* Increases the Flavor Diversity of Wines. J Agric Food
621 Chem. 64: 4574-4583.
622 8. Lleixà J, Martín V, Portillo M del C, Carrau F, Beltran G, Mas A. 2016. Comparison of
623 fermentation and wines produced by inoculation of *Hanseniaspora vineae* and
624 *Saccharomyces cerevisiae*. Front Microbiol 7:338.

9. Medina K, Boido E, Fariña L, Gioia O, Gomez ME, Barquet M, Gaggero C, Dellacassa E, Carrau F. 2013. Increased flavour diversity of Chardonnay wines by spontaneous fermentation and co-fermentation with *Hanseniaspora vineae*. Food Chem 141:2513–2521.
10. Viana F, Belloch C, Vallés S, Manzanares P. 2011. Monitoring a mixed starter of *Hanseniaspora vineae-Saccharomyces cerevisiae* in natural must: Impact on 2-phenylethyl acetate production. Int J Food Microbiol 151:235–240.
11. Ciani M, Comitini F, Mannazzu I, Domizio P. 2009. Controlled mixed culture fermentation: a new perspective on the use of non *Saccharomyces* yeasts in winemaking. FEMS Yeast Res 10:123–133.
12. Curtin CD, Pretorius IS. 2014. Genomic insights into the evolution of industrial yeast species *Brettanomyces bruxellensis*. FEMS Yeast Res 14:997–1005.
13. Bisson LF, Karpel JE. 2010. Genetics of Yeast Impacting Wine Quality. Annu Rev Food Sci Technol 1:139–162.
14. Cordente AG, Curtin CD, Varela C, Pretorius IS. 2012. Flavour-active wine yeasts. Appl Microbiol Biotechnol 96:601–618.
15. Hazelwood LA, Daran J-M, van Maris AJA, Pronk JT, Dickinson JR. 2008. The Ehrlich Pathway for Fusel Alcohol Production: a Century of Research on *Saccharomyces cerevisiae* Metabolism. Appl Environ Microbiol 74:2259–2266.
16. Pires EJ, Teixeira JA, Brányik T, Vicente AA. 2014. Yeast: The soul of beer's aroma - A review of flavour-active esters and higher alcohols produced by the brewing yeast. Appl Microbiol Biotechnol 98:1937–1949.
17. Rosini G, Federici F, Martini A. 1982. Yeast flora of grape berries during ripening. Microb Ecol 8:83–89.
18. Loureiro V, Ferreira MM, Monteiro S, Ferreira RB. 2012. The microbial community of grape berry, p. 241–268. *In* The Biochemistry of the Grape Berry.
19. Giorello FM, Berná L, Greif G, Camesasca L, Salzman V, Medina K, Robello C, Gaggero C,

Aguilar PS, Carrau F. 2014. Genome Sequence of the Native Apiculate Wine Yeast *Hanseniaspora vineae* T02/19AF. Genome Announc 2:514–530.

20. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED. 2012. *Saccharomyces* Genome Database: The genomics resource of budding yeast. Nucleic Acids Res 40:700–705.

21. Goffeau a, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. 1996. Life with 6000 Genes. Science (80) 274:546–567.

22. Haase SB, Lew DJ. 1997. Flow cytometric analysis of DNA content in budding yeast. Methods Enzymol 283:322–332.

23. Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ. 2011. Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genet 7: e1001287.

24. Borneman AR, Forgan AH, Pretorius IS, Chambers PJ. 2008. Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain, p. 1185–1195. *In* FEMS Yeast Research.

25. Wei W, McCusker JH, Hyman RW, Jones T, Ning Y, Cao Z, Gu Z, Bruno D, Miranda M, Nguyen M, Wilhelmy J, Komp C, Tamse R, Wang X, Jia P, Luedi P, Oefner PJ, David L, Dietrich FS, Li Y, Davis RW, Steinmetz LM. 2007. Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. Proc Natl Acad Sci U S A 104:12825–12830.

26. Santos MAS, Gomes AC, Santos MC, Carreto LC, Moura GR. 2011. The genetic code of the fungal CTG clade. Comptes Rendus - Biol 33:607–611.

27. Kurtzman CP, Robnett CJ. 2003. Phylogenetic relationships among yeasts of the

"*Saccharomyces* complex" determined from multigene sequence analyses. FEMS Yeast Res 3:417–432.

28. Suh S-O, Blackwell M, Kurtzman CP, Lachance M-A. 2006. Phylogenetics of Saccharomycetales, the ascomycete yeasts. Mycologia 98:1006–1017.

29. Rossouw D, Næs T, Bauer FF. 2008. Linking gene regulation and the exo-metabolome: A comparative transcriptomics approach to identify genes that impact on the production of volatile aroma compounds in yeast. BMC Genomics 9:530.

30. Yorimitsu T, Nair U, Yang Z, Klionsky DJ. 2006. Endoplasmic reticulum stress triggers autophagy. J Biol Chem 281:30299–30304.

31. Paulick MG, Bertozzi CR. 2008. The glycosylphosphatidylinositol anchor: A complex membrane-anchoring structure for proteins. Biochemistry 47:6991–7000.

32. Negruţa O, Csutak O, Stoica I, Rusu E, Vassu T. 2010. Methylotrophic yeasts: Diversity and methanol metabolism. Rom Biotechnol Lett 15:5369–5375.

33. Borneman AR, Forgan AH, Kolouchova R, Fraser JA, Schmidt SA. 2016. Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of *Saccharomyces cerevisiae*. Genes, Genomes, Genet 6:957–971.

34. Hauser NC, Fellenberg K, Gil R, Bastuck S, Hoheisel JD, Pérez-Ortín JE. 2001. Whole genome analysis of a wine yeast strain. Comp Funct Genomics 2:69–79.

35. Yoshimoto H, Fukushige T, Yonezawa T, Sone H. 2002. Genetic and physiological analysis of branched-chain alcohols and isoamyl acetate production in *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 59:501–508.

36. Trinh TTT, Woon WY, Yu B, Curran P, Liu SQ. 2010. Effect of L-isoleucine and L-phenylalanine addition on aroma compound formation during longan juice fermentation by a co-culture of *Saccharomyces cerevisiae* and *Williopsis saturnus*. South African J Enol Vitic 31:116–124.

37. Dickinson JR, Salgado LEJ, Hewlins MJE. 2003. The catabolism of amino acids to long

chain and complex alcohols in *Saccharomyces cerevisiae*. J Biol Chem 278:8028–8034.

38. Momoi M, Tanoue D, Sun Y, Takematsu H, Suzuki Y, Suzuki M, Suzuki A, Fujita T, Kozutsumi Y. 2004. SLI1 (YGR212W) is a major gene conferring resistance to the sphingolipid biosynthesis inhibitor ISP-1, and encodes an ISP-1 N-acetyltransferase in yeast. Biochem J 381:321–328.

39. Moreira N, Mendes F, Hogg T, Vasconcelos I. 2005. Alcohols, esters and heavy sulphur compounds production by pure and mixed cultures of apiculate wine yeasts. Int J Food Microbiol 103:285–294.

40. Medina K. 2014. Biodiversidad de levaduras no-*Saccharomyces*: efecto del metabolismo secundario en el color y el aroma de vinos de calidad. Universidad de la República, Uruguay.

41. Riley R, Haridas S, Wolfe KH, Lopes MR, Hittinger CT, Göker M, Salamov AA, Wisecaver JH, Long TM, Calvey CH, Aerts AL, Barry KW, Choi C, Clum A, Coughlan AY, Deshpande S, Douglass AP, Hanson SJ, Klenk H-P, LaButti KM, Lapidus A, Lindquist EA, Lipzen AM, Meier-Kolthoff JP, Ohm RA, Otillar RP, Pangilinan JL, Peng Y, Rokas A, Rosa CA, Scheuner C, Sibirny AA, Slot JC, Stielow JB, Sun H, Kurtzman CP, Blackwell M, Grigoriev I V, Jeffries TW. 2016. Comparative genomics of biotechnologically important yeasts. Proc Natl Acad Sci 113:9882–9887.

42. Sternes PR, Lee D, Kutyna DR, Borneman AR. 2016. Genome Sequences of three species of *Hanseniaspora* isolated from spontaneous wine fermentations. Genome Announc 4:e01287-16.

43. Verstrepen KJ, Van Laere SDM, Vanderhaegen BMP, Derdelinckx G, Dufour JP, Pretorius IS, Winderickx J, Thevelein JM, Delvaux FR. 2003. Expression levels of the yeast alcohol acetyltransferase genes *ATF1, Lg-ATF1*,and *ATF2* control the formation of a broad range of volatile esters. Appl Environ Microbiol 69:5228–5237.

44. Barquet M, Martín V, Medina K, Pérez G, Carrau F, Gaggero C. 2012. Tandem repeat-tRNA (TRtRNA) PCR method for the molecular typing of non-*Saccharomyces* subspecies. Appl

729 Microbiol Biotechnol 93:807–814.
730 45. Patel RK, Jain M. 2012. NGS QC Toolkit: A Toolkit for Quality Control of Next Generation
731 Sequencing Data. PLoS One 7: e30619.
732 46. Brown CT, Howe A, Zang Q, Pyrksz AB, Brom THA. 2012. Reference-free algorith for
733 comutational normalization of shotgun sequencing data. ArXiv 1203.4802
734 47. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB,
735 Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N,
736 Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. 2013.
737 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for
738 reference generation and analysis. Nat Protoc 8:1494–1512.
739 48. Zimin A V, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA
740 genome assembler. Bioinformatics 29:2669–2677.
741 49. Huang S, Chen Z, Huang G, Yu T, Yang P, Li J, Fu Y, Yuan S, Chen S, Xu A. 2012.
742 HaploMerger: Reconstructing allelic relationships for polymorphic diploid genome
743 assemblies. Genome Res 22:1581–1588.
744 50. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab
745 initio prediction of alternative transcripts. Nucleic Acids Res 34:W435--W439.
746 51. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A,
747 Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the
748 protein families database. Nucleic Acids Res 42:D222--30.
749 52. Parra G, Bradnam K, Korf I. 2007. CEGMA: A pipeline to accurately annotate core genes in
750 eukaryotic genomes. Bioinformatics 23:1061–1067.
751 53. Alexa A, Rahnenführer J. 2007. Gene set enrichment analysis with topGO. Bioconductor
752 Improv 27.
753 54. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods.
754 55. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin

R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

56. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498.

57. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303.

58. Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

59. Waterhouse RM, Zdobnov EM, Kriventseva E V. 2011. Correlating traits of gene retention, sequence divergence, duplicability and essentiality in vertebrates, arthropods, and fungi. Genome Biol Evol 3:75–86.

60. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

61. Suyama M, Torrents D, Bork P. 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34.

62. Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552.

63. Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189.

64. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane H, Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 4:P3.

65. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: Kyoto

encyclopedia of genes and genomes. Nucleic Acids Res 27:29–34.

66. Kück P, Meusemann K. 2010. FASconCAT: Convenient handling of data matrices. Mol Phylogenet Evol 56:1115–1118.

67. Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29:1695–1701.

68. Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497:327–331.

69. Altenhoff AM, Schneider A, Gonnet GH, Dessimoz C. 2011. OMA 2011: orthology inference among 1000 complete genomes. Nucleic Acids Res 39:D289--D294.

70. Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

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

72. Henschke PA, Jiranek V. 1993. Yeast: Metabolism of nitrogen compounds. In: Wine Microbiology and Biotechnology, p. 77–164. *In* Graham H. Fleet (ed.), Wine Microbiology and Biotechnology. Harwood Academic Publishers.

73. Fariña L, Medina K, Urruty M, Boido E, Dellacassa E, Carrau F. 2012. Redox effect on volatile compound formation in wine during fermentation by *Saccharomyces cerevisiae*. Food Chem 134:933–939.

802

803 **Table 1.** Yeast strains analysed in this work

Species	Strain	Ploidy	Source	BioSample ID NCBI* database	Use in this Work
H. vineae	T02/19AF	Haploid	Uruguayan Tannat grape vines	SAMN02644989	Genomic transcriptomic and phenomic study
H.vineae	T02/05AF	Haploid	Uruguayan Tannat grape vines	SAMN04487210	Genomic study
S. cerevisiae	BY4742	Haploid	Laboratory strain, derived from S288c	SAMN03020230	FCM analysis
S. cerevisiae	BY4743	Diploid	Laboratory strain, derived from S288c	SAMN01822968	FCM analysis
S. cerevisiae	Montrachet 522	Diploid	Fortified wines, California	SAMN03325349	Flavor compounds analysis
S. cerevisiae	S288c	Haploid	Laboratory strain, California	SAMD00065885	Genomic comparison
S. cerevisiae	AWRI1631	Haploid	Australian derivative of South African commercial wine strain N96	SAMN02953734	Genomic comparison
S. cerevisiae	AWRI796	Diploid	South African wine strain	SAMN04286136	Genomic comparison
S. cerevisiae	BC187	Diploid	Derivative of California wine barrel isolate	SAMEA687137	Genomic comparison
S. cerevisiae	DBVPG6044	Diploid	West African isolate	SAMEA687132	Genomic comparison
S. cerevisiae	EC1118	Diploid	Commercial wine strain	SAMEA2272624	Genomic comparison
S. cerevisiae	L1528	Diploid	Chilean wine strain	SAMN03020223	Genomic comparison
S. cerevisiae	LalvinQA23	Diploid	Portuguese Vinho Verde white wine strain	SAMN02981266	Genomic comparison
S. cerevisiae	M22	Diploid	Italian vineyard isolate	SAMN00189351	Genomic comparison
S. cerevisiae	PW5	Diploid	Nigerian Raphia palm wine isolate	SAMN00199004	Genomic comparison
S. cerevisiae	RM11-1A	Haploid	Natural isolate collected from a vineyard, California	SAMN02953602	Genomic comparison
S. cerevisiae	T73	Near-diploid	Spanish red wine strain	SAMN00198997	Genomic comparison
S. cerevisiae	Vin13	Diploid	South African white wine strain	SAMN02981268	Genomic comparison
S. cerevisiae	VL3	Diploid	French white wine strain	SAMN02981289	Genomic comparison
S. cerevisiae	YJM269	Diploid	Austrian wine from Blauer Portugieser grapes isolate	SAMN02981310	Genomic comparison

804

805 *National Center of Biotechnology Information

Table 2. Genome assembly report of the two strains of *H. vineae*.

Strain	Genome size assembly (Mb)	Total number of contigs	Number of ORFs	Predicted proteins homologous to *S.cerevisiae*
H.vineae T02/05AF	11.37	741	4741	3862
H.vineae T02/19AF	11.33	305	4708	3861

Table 3. Comparison of genes involved in biosynthesis routes for key flavor compounds production in *S. cerevisiae* and *H. vineae*. Crossed letters represent absent homologous genes in *H. vineae* and repeated genes are indicated as "number of copies x gene abbreviation". Predicted amino acidsequences from the genome of *H. vineae* were compared with protein homologous found in *S. cerevisiae.*

Biosynthesis route	Enzymatic activity	Genes identified (% amino acid identity with *S. cerevisiae* homologous protein)
Higher alcohols	Aromatic amino acid transferases	3x*ARO8* (45.51, 59.84, 56.06); 4x*ARO9* (42.70, 35.27, 36.08; 36.91)
	Branched chain amino acid transferases	*BAT1* (78.84); ~~*BAT2*~~
	Decarboxylase	2x*ARO10* (34.10, 30.99); 2x*PDC1* (80.46; 50.66); ~~*PDC5*~~; ~~*PDC6*~~; ~~*THI3*~~
	Alcohol dehydrogenase	2x*ADH1* (77.71, 78.74), ~~*ADH2*~~; 2x*ADH3* (74.78, 74.80); ~~*ADH4*~~; ~~*ADH5*~~; 4x*ADH6* (44.74, 44.47, 44.74, 44.06); ~~*ADH7*~~; *SFA1* (68.16); 4x*GRE2* (44.74, 50.73, 47.51, 43.02); ~~*YPR1*~~; ~~*PAD1*~~; ~~*SPE1*~~; 3x*OYE2* (55.10, 58.06, 57.25); *HOM2* (78.24)
	Aryl alcohol dehydrogenase	~~*AAD3*~~; ~~*AAD4*~~; ~~*AAD6*~~; ~~*AAD10*~~; ~~*AAD14*~~; ~~*AAD15*~~; ~~*AAD16*~~
	Regulation	*ARO80* (34.80); ~~*GAT2*~~; ~~*GLN3*~~; ~~*GZF3*~~; ~~*DAL80*~~
Acetate esters	Alcohol acetyl transferases	~~*ATF1*~~; *ATF2* (26.58); 4xSLI1 (22-24%); g4599.t1
Ethyl esters	Ethanol *O*-acyltransferase and esterase	~~*EEB1*~~; *EHT1* (51.35); *MGL2* (30.06); ~~*AAD*~~; *IAH1* (54.67)
Volatile organic acids	Aldehyde dehydrogenase	2x*ALD2* (40.55, 44.01); ~~*ALD3*~~; ~~*ALD4*~~; *ALD5* (53.45); *ALD6* (55.07)
Aromatic amino acid synthesis	Synthesis of chorismate, phenylalanine, tryptophan and tyrosine	*ARO1* (66.79); *ARO2* (80.59); *ARO3* (77.03); *ARO4* (83.51); *TRP2* (70.84); *TRP3* (69.14); *ARO7* (67.97); *PHA2* (41.99); *TYR1* (62.37)
Benzyl alcohol/benzaldehyde synthesis	Mandelate pathway	2x*ARO10* (34.10, 30.99); 2x*PDC1* (80.46; 50.66); SCS7 (66.50); *ALD6* (55.07); 2x*ALD2* (40.55, 44.01); *DLD1* (53.00); *DLD2* (70.00); ~~*DLD3*~~

Table 4. Transcriptomic assembly reference metrics for *H. vineae* T02/19AF.

Parameter	Transcriptomic reference
Total length (bp)	9,362,444
Total contig number	4,725
Max. contig length (bp)	17,336
Min. contig length (bp)	226
Mean contig length (bp)	1,982
Median contig length (bp)	1,683
Number of genes annotated	4,725

865 **Table 5**. Exometabolome of *H. vineae* and *S. cerevisiae*. Flavor compounds at the end of the fermentation; results are the average of triplicate fermentations at 20 ºC
866 in the chemical defined grape (CDG) synthetic medium.

		H. vineae						*S. cerevisiae*		
		T02/05AF			T02/19AF			M522		
Compounds	**L.R.I.***	**Average content (g/L)****		**Standard deviation**	**Average content (g/L)**		**Standard deviation**	**Average content (g/L)**		**Standard deviation**
Alcohols										
2-methyl-2-butanol	975	168	±	78	159	±	1	n.d.		
1-propanol	996	n.d.			2	±	2	42	±	1
2-methyl-1-propanol	1067	631	±	490	750	±	22	3,488	±	4
1-butanol	1128	31	±	14	33	±	5	91	±	2
3-methyl-1-butanol	1187	25,028	±	3,699	28,326	±	954	54,953	±	41
2,3-butanediol	1526	422	±	68	1,076	±	65	n.d.		
3-ethoxy-1-propanol	1389	75	±	17	135	±	6	175	±	1
2-ethyl-1-hexanol	1453	29	±	4	26	±	6	312	±	1
methionol	1716	2,032	±	230	2,601	±	170	4,980	±	6
benzyl alcohol	1822	141	±	25	179	±	8	n.d.	±	
2-phenylethanol	1906	8,029	±	2,067	9,879	±	120	18,387	±	2
tyrosol	3012	814	±	188	1,006	±	11	7,683	±	4
Esters										
3-methylbutyl acetate	1126	33	±	19	20	±	4	54	±	1
ethyl lactate	1341	66	±	5	81	±	17	116	±	1
benzyl acetate	1690	6	±	0	4	±	0	n.d.		
2-phenylethyl acetate	1813	10,054	±	929	9,205	±	1,435	1,185	±	6
Fatty acids										
2-methylpropanoic acid	1588	301	±	21	668	±	52	168	±	1
butanoic acid	1625	59	±	6	55	±	6	133	±	1
3-methylbutanoic acid	1650	67	±	10	146	±	3	448	±	1
hexanoic acid	1843	82	±	19	67	±	4	461	±	1
octanoic acid	2070	127	±	37	89	±	14	875	±	2
decanoic acid	2243	170	±	111	81	±	26	96	±	2

Other compounds											
3-hydroxy-2-butanone	1270	4,328	±	1,858	5,165	±	742	303	±	20
γ-butyrolactone	1620	90	±	22	153	±	14	338	±	2

* Linear retention index based on a series of n-hydrocarbons reported according to their elution order on Carbowax 20M.
*Mean of three repetitions and standard deviation for three fermentation

867

868

869 **Table 6**. Exometabolome of *H. vineae*. Flavor compounds produced at day 4 and 10; results are the average
870 of triplicate fermentations at 20 ºC in chemical defined grape (CDG) synthetic medium.

871

Compounds	L.R.I.*	DAY 4 Average content (g/L)**		DAY 4 Standard deviation	DAY 10 Average content (g/L)		DAY 10 Standard deviation
Alcohols							
2-methyl-2-butanol	975	66	±	5	42	±	2
1-propanol	996	116	±	5	40	±	5
2-methyl-1-propanol	1067	3,620	±	268	2,990	±	290
1-butanol	1128	149	±	51	122	±	10
3-methyl-1-butanol	1187	42,525	±	1,288	36,859	±	1,693
2,3-butanediol	1526	1,310	±	74	1,450	±	252
3-ethoxy-1-propanol	1389	13	±	13	177	±	8
2-ethyl-1-hexanol	1453	n.d.			39	±	2
methionol	1716	1,605	±	60	1,925	±	60
3-acethoxy-1-propanol	1756	1,335	±	109	1,520	±	50
benzyl alcohol	1822	280	±	9	407	±	33
2-phenylethanol	1906	6,657	±	317	7,587	±	361
tyrosol	3012	33	±	33	2,213	±	638
Esters							
3-methylbutyl acetate	1126	91	±	21	112	±	33
ethyl lactate	1341	n.d			62	±	3
ethyl 2-hydroxyhexanoate	1650	n.d.			20	±	10
benzyl acetate	1690	n.d.			10	±	1
2-phenylethyl acetate	1813	5,862	±	627	10,260	±	995
ethyl 4-hydroxy-butoanoate	1819	n.d.			1,344	±	47
diethyl 2 hydroxy glutarate	2202	n.d.			10	±	2
Fatty acids							
2-methylpropanoic acid	1588	2,366	±	158	3,024	±	138
butanoic acid	1625	57	±	12	97	±	6
3-methylbutanoic acid	1650	71	±	11	128	±	5
hexanoic acid	1843	50	±	4	110	±	4
octanoic acid	2070	44	±	12	164	±	17
decanoic acid	2243	15	±	15	308	±	67
Other compounds							
2,3-butanedione	935	407	±	53	58	±	9
2,3-pentanedione	1046	76	±	25	15	±	3
3-hydroxy-2-butanone	1270	12,691	±	348	9,669	±	275
3-hydroxy-2-pentanone	1330	1,353	±	45	1,121	±	184
γ-butyrolactone	1620	64	±	32	116	±	7
n-formyl tyramine	2890	727	±	145	8,788	±	451

* Linear retention index to their based on a series of n-hydrocarbons reported according elution order on Carbowax 20M
*Mean of three repetitions and standard deviation for three fermentations.

872

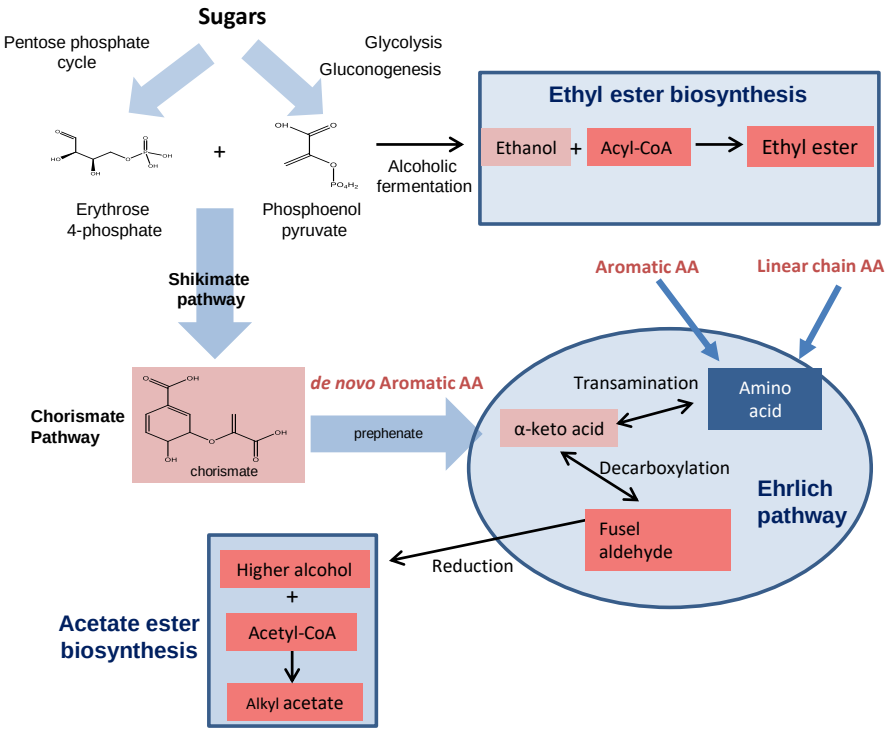
873

874

875

876

877



878

879

880 **FIG 1**. Metabolic pathways studied in this work involved in wine aroma formation. Ehrlich
881 pathway for higher alcohols production, acetate ester biosynthesis and ethyl ester biosynthesis from
882 amino acids (AA) and sugars.
883

884

885

886

887

888

889

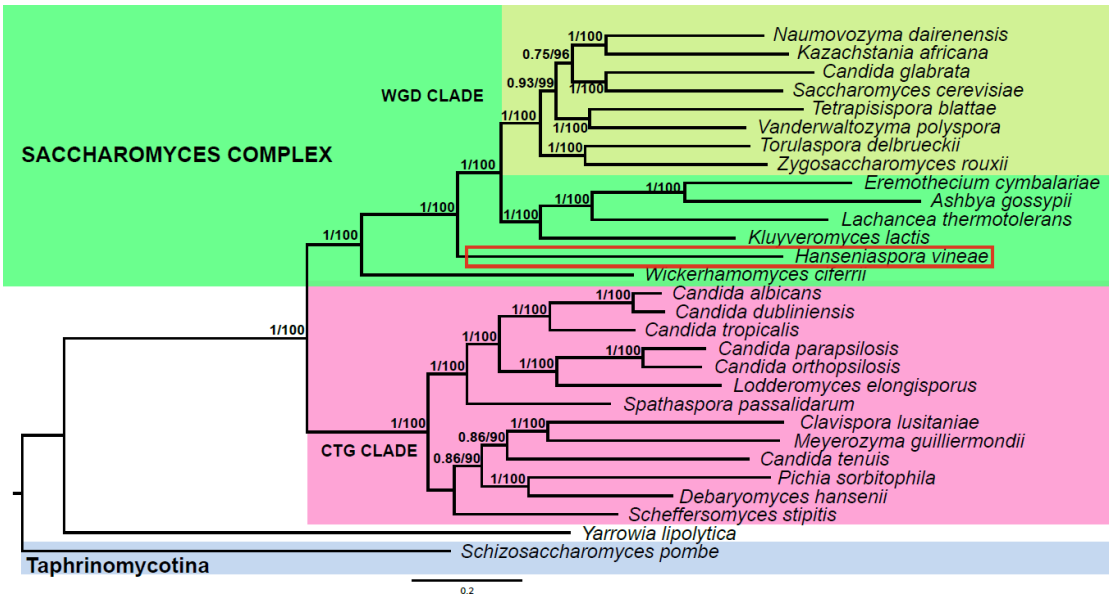


FIG 2. Maximum likelihood phylogeny of *Saccharomyces* complex species from concatenation of 227 genes. *H. vineae* is framed in red inside the *Saccharomyces* complex and outside of the Whole Genome Duplication (WGD) clade. The clade CTG groups yeasts with alternative genetic code. Numbers close to the node match Bootstrap Support (BS) for those values above 70 and Internode Certainty (IC), respectively. The scale bar represents units of amino acid substitutions per site. The tree has a mid-point root for easier visualization.

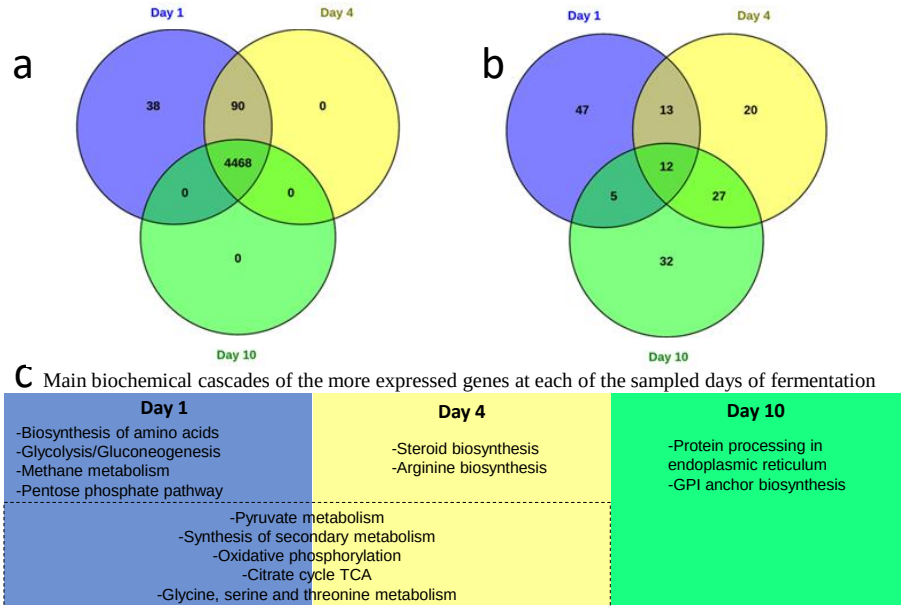
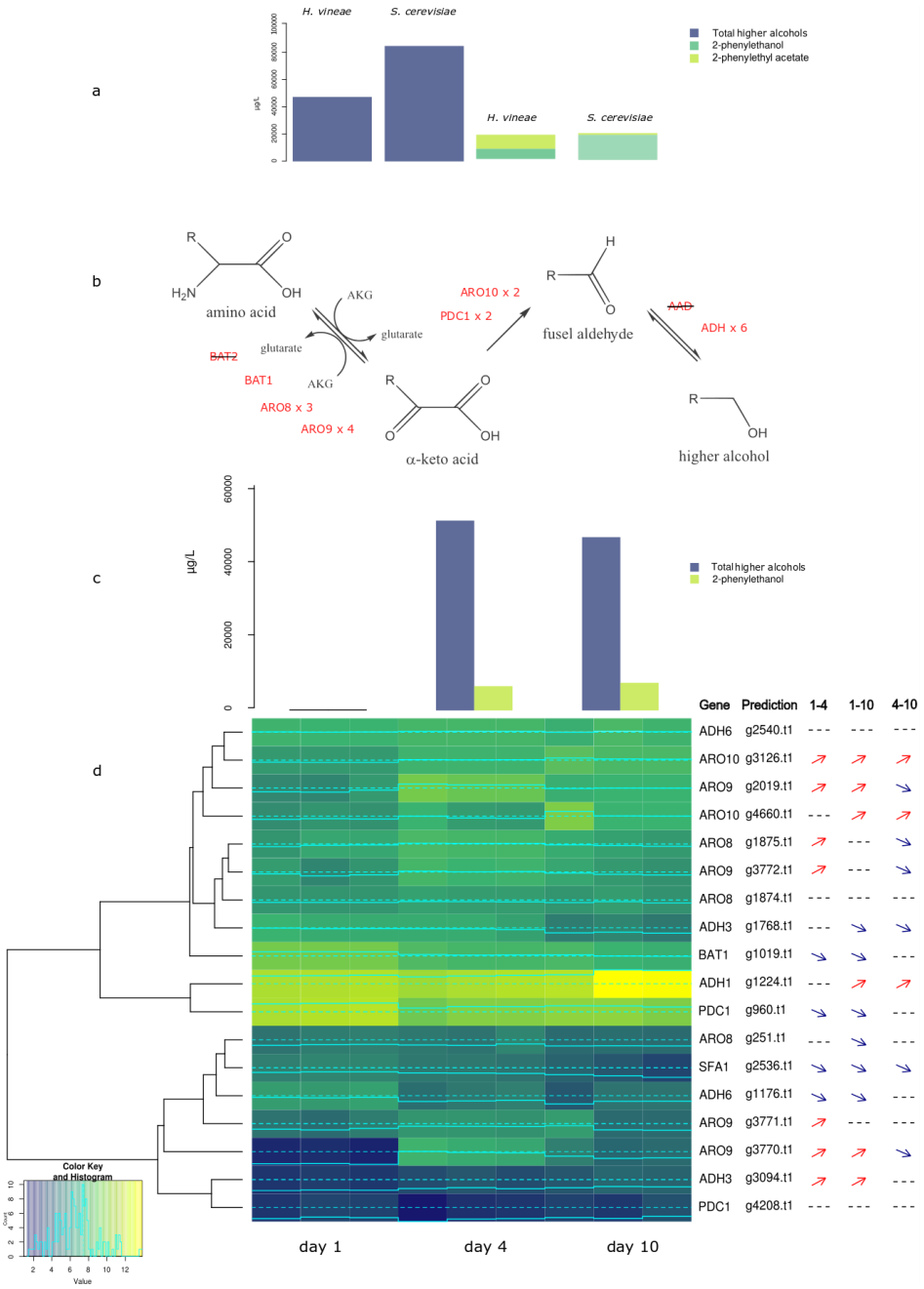


FIG 3. Overview of transcriptomic analysis.
a) Venn diagram showing the differentially expressed genes shared between each fermentation point; b) Venn diagram showing the genes shared between each fermentation point for the top 100 most highly expressed genes; c) main biochemical cascades of the most expressed genes at each sampled day of fermentation. GPI: glycosylphosphatidylinositol.

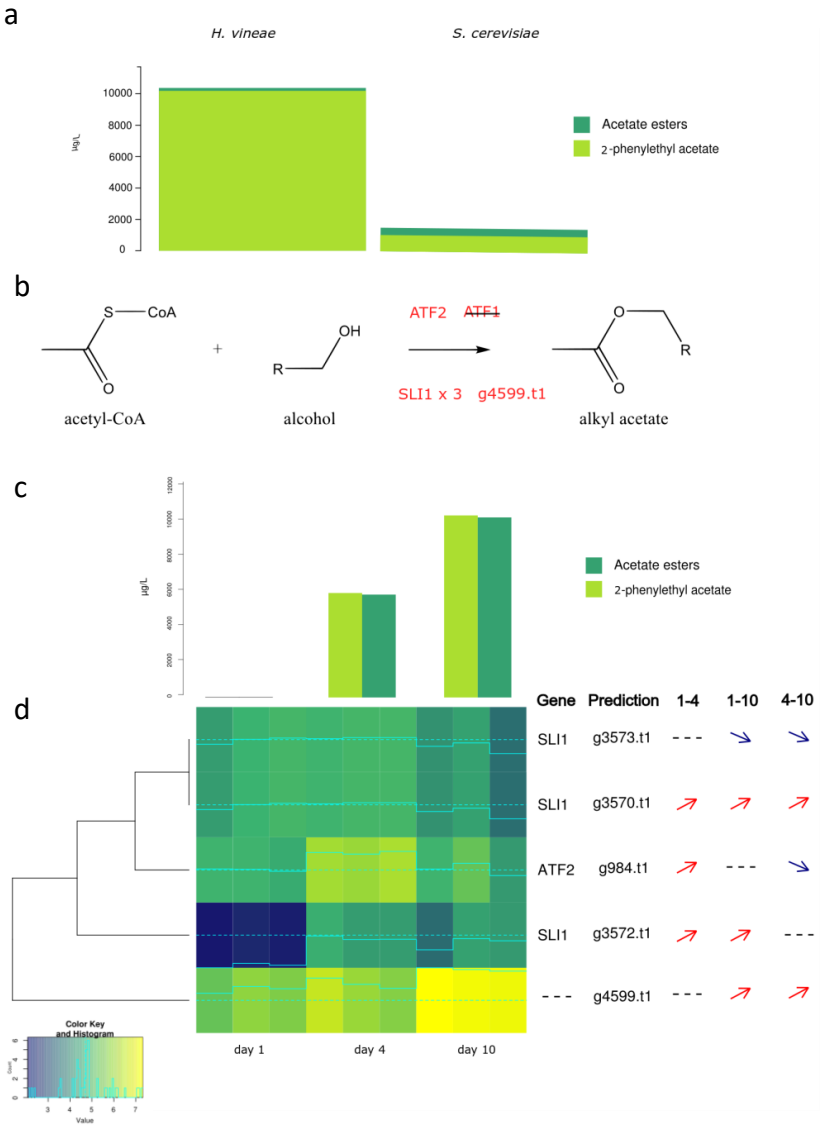


911

912 **FIG 4.** Higher alcohols and 2-phenylethanol production and putatively related genes.
913 a) Comparison of total higher alcohols, 2-phenyelthanol and 2-phenylethyl acetate produced in *H.*
914 *vineae* and *S. cerevisiae* at day 10 of fermentation; b) the three steps of metabolic pathway of higher
915 alcohols biosynthesis with putative enzymes involved in *H. vineae*; c) production of total higher
916 alcohols and 2-phenylacetate by *H. vineae* at 1, 4 and 10 days of fermentation; d) Expression
917 heatmap of genes putatively involved in higher alcohols and 2-phenylethanol production from *H.*
918 *vineae* at 1, 4 and 10 days of fermentation. Lighter colors indicate higher expression values and data

919 are shown for triplicates. Significant changes in expression of each gene are indicated with arrows
920 to the right of the heatmap as analyzed using the package edgeR (FDR < 0.05): (1–4) indicates
921 differential expression between days 1 and 4; (1–10) indicates differential expression between days
922 1 and 10; and (4–10) indicates differential expression between days 4 and 10.
923

924



925

926 **FIG 5.** Acetate esters production and putatively related genes.
927 a) Comparison of total acetate esters and 2-phenylethyl acetate produced in *H. vineae* and *S.*
928 *cerevisiae* at day 10 of fermentation; b) metabolic pathway of acetate esters biosynthesis with
929 putative enzymes involved in *H. vineae*; c) production of total acetate esters and 2-phenylethyl

acetate by *H. vineae* at 1, 4 and 10 days of fermentation; c) Expression heatmap of genes putatively involved in total acetate esters and 2-phenylethyl acetate production from *H. vineae* at 1, 4 and 10 days of fermentation. Lighter colors indicate higher expression values and data are shown for triplicates. Significant changes in expression of each gene are indicated with arrows to the right of the heatmap as analyzed using the package edgeR (FDR < 0.05): (1–4) indicates differential expression between days 1 and 4; (1–10) indicates differential expression between days 1 and 10; and (4–10) indicates differential expression between days 4 and 10.

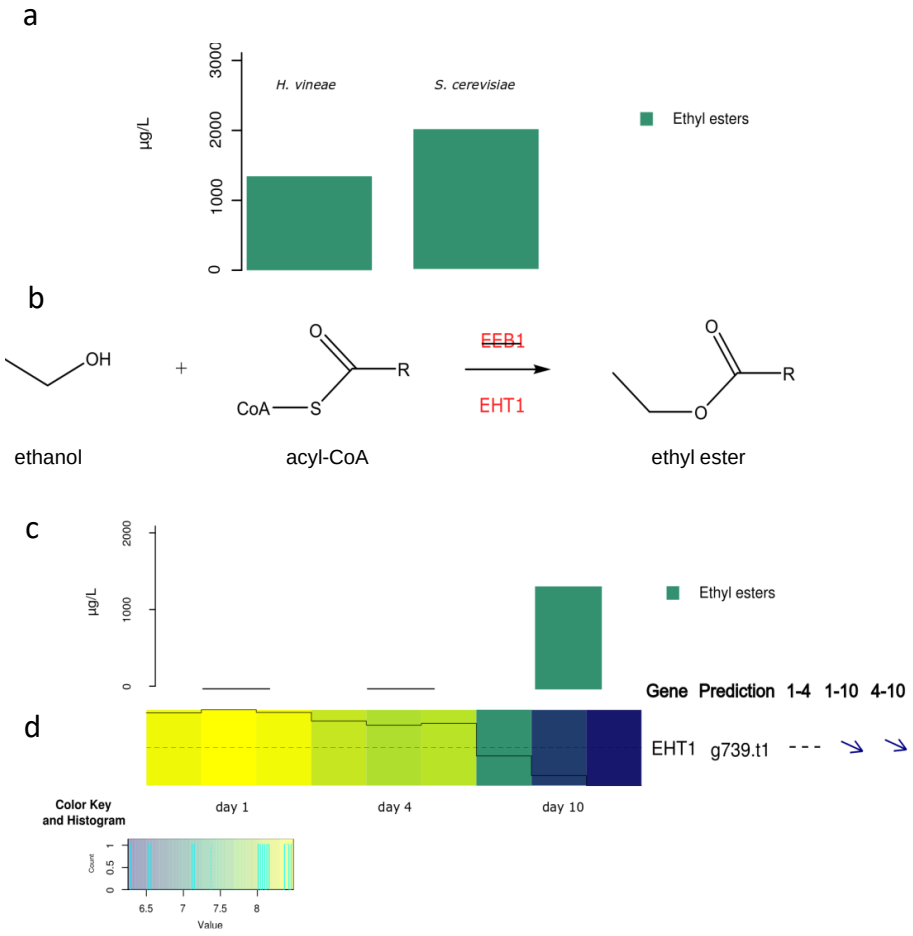


FIG 6. Ethyl esters production and putatively related genes.
a) Comparison of ethyl esters produced in *H. vineae* and *S. cerevisiae* at day 10 of fermentation; b) metabolic pathway of acetate esters biosynthesis with putative enzymes involved in *H. vineae*; c) production of ethyl esters by *H. vineae* at 1, 4 and 10 days of fermentation; d) Expression heatmap of genes putatively involved in ethyl esters production from *H. vineae* at 1, 4 and 10 days of fermentation. Lighter colors indicate higher expression values and data shown are of triplicates. Significant changes in expression of each gene are indicated with arrows to the right of the heatmap as analyzed using the package edgeR (FDR < 0.05): (1–4) indicates differential expression between days 1 and 4; (1–10) indicates differential expression between days 1 and 10; and (4–10) indicates differential expression between days 4 and 10.