1   **Mapping global shifts in *Saccharomyces cerevisiae* gene expression across**

2   **asynchronous time trajectories with diffusion maps**

3

4   *Taylor Reiter [1,2,3,#], Rachel Montpetit [2], Ron Runnebaum [2,4], C. Titus Brown [3], and Ben Montpetit*

5   *[1,2,#].*

6

7   [1] Food Science Graduate Group, University of California Davis, Davis, CA, USA

8   [2] Department of Viticulture and Enology, University of California Davis, Davis, CA, USA

9   [3] Department of Population Health and Reproduction, University of California, Davis, CA, USA

10 [4] Department of Chemical Engineering, University of California, Davis, CA 95616, USA

11

12 # Correspondence should be addressed to Taylor Reiter (tereiter@ucdavis.edu) and Ben

13 Montpetit (benmontpetit@ucdavis.edu).

14

Abstract

Grapes grown in a particular geographic region often produce wines with consistent characteristics, suggesting there are site-specific factors driving recurrent fermentation outcomes. However, our understanding of the relationship between site-specific factors, microbial metabolism, and wine fermentation outcomes are not well understood. Here, we used differences in *Saccharomyces cerevisiae* gene expression as a biosensor for differences among Pinot noir fermentations from 15 vineyard sites. We profiled time series gene expression patterns of primary fermentations, but fermentations proceeded at different rates, making analyzes of these data with conventional differential expression tools difficult. This led us to develop a novel approach that combines diffusion mapping with continuous differential expression analysis. Using this method, we identified vineyard specific deviations in gene expression, including changes in gene expression correlated with the activity of the non-*Saccharomyces* yeast *Hanseniaspora uvarum*, as well as with initial nitrogen concentrations in grape musts. These results highlight novel relationships between site-specific variables and *Saccharomyces cerevisiae* gene expression that are linked to repeated wine fermentation outcomes. In addition, we demonstrate that our analysis approach can extract biologically relevant gene expression patterns in other contexts (e.g., hypoxic response of *Saccharomyces cerevisiae*), indicating that this approach offers a general method for investigating asynchronous time series gene expression data**.**

34 **Importance**

35   While it is generally accepted that foods, in particular wine, possess sensory
36 characteristics associated with or derived from their place of origin, we lack knowledge of
37 the biotic and abiotic factors central to this phenomenon. We have used *Saccharomyces*
38 *cerevisiae* gene expression as a biosensor to capture differences in fermentations of Pinot
39 noir grapes from 15 vineyards across two vintages. We find that gene expression by non-
40 *Saccharomyces* yeasts and initial nitrogen content in the grape must correlates with
41 differences in gene expression among fermentations from these vintages. These findings
42 highlight important relationships between site-specific variables and gene expression that
43 can be used to understand, or possibly modify, wine fermentation outcomes. Our work also
44 provides a novel analysis method for investigating asynchronous gene expression data sets
45 that is able to reveal both global shifts and subtle differences in gene expression due to varied
46 cell – environment interactions.

Introduction

47

48   During a wine fermentation, *Saccharomyces cerevisiae* metabolizes sugars and other

49   nutrients to obtain energy and increase biomass, while also dealing with a common set of

50   stresses caused by the must/wine environment. Given these general features of the system,

51   the cellular activities of *S. cerevisiae* throughout a wine fermentation are consistent, which is

52   reflected in a core gene expression program (CGEP) that is present across diverse

53   fermentation conditions (1–4). However, this is not to say that *S. cerevisiae* metabolism is

54   fixed, as *S. cerevisiae* dynamically responds to differences in the fermentation environment

55   (e.g., nutrient levels, temperature, and differences in microbial communities) to maintain

56   cellular metabolism and overall fitness (1, 2, 5, 6). For example, differences in grape must

57   nitrogen concentrations leads to changes in metabolism that are accompanied by altered

58   aroma compound in wine (7). This highlights the fact that metabolic adaptation to varied

59   fermentation environments leads to differences in wine fermentation outcomes, including

60   sensory differences. This relationship is mirrored by many findings showing that genetic

61   changes causing altered expression of select genes or pathways in *S. cerevisiae* leads to

62   quantifiable differences in wine fermentation outcomes (8). These facts support the

63   generally accepted idea that interactions between *S. cerevisiae* and the unique chemical and

64   biological matrix of each grape must are central to defining primary fermentation

65   characteristics. We reason that these differences are the result of 1) the expression of unique

66   genes outside those in the CGEP required for fermentation and/or 2) variation in expression

67   of CGEP genes that changes the activity of various core pathways during fermentation.

68   The chemical and biological diversity of grape musts is in large part a reflection of

69   biotic and abiotic pressures encountered by a grapevine during a growing season. For

70   example, wines produced under similar vinification conditions from genetically identical

71   grapes grown in different locations have diverse sensory outcomes (9), many of which are

72   reproducible across multiple vintages (10). After observing diverse sensory outcomes in

73   wines where a consistent variable between fermentations was vineyard location (9), we

74   sought to uncover quantifiable contributions of vineyard site by using *S. cerevisiae* gene

75   expression as a biosensor to detect differences between fermentations. This is motivated by

76   the fact that high throughput gene expression surveys (microarray and RNA sequencing)

77   have revealed the causes of stuck and sluggish fermentations (11), triggers for entry into

78   stationary phase (1, 2), and the impact of inter-species interactions on *S. cerevisiae*
79   metabolism in wine (6, 12, 13). In addition, as an organism commonly used in life science
80   and biotechnology research, the *S. cerevisiae* genome and transcriptome is well understood
81   with published datasets focused on gene expression in diverse environments, including wine
82   (1–3, 14–16). This makes *S. cerevisiae* a powerful tool for understanding the wine
83   fermentation environment and identifying key biotic and abiotic factors underlying
84   fermentation outcomes.

85   Towards this end, we performed time series RNA sequencing on Pinot noir
86   fermentations with the aim of identifying gene expression differences by vineyard site.
87   However, using standard analysis methods (17–22), we only identified the dominant and
88   consistent global shift in gene expression, the CGEP, across fermentations and not gene
89   expression patterns indicating altered *S. cerevisiae* metabolism that would differentiate
90   vineyard site (4). A major issue was that sampled fermentations proceeded at different rates,
91   leading to asynchronous biological progression among sequenced samples with respect to
92   fermentation status (e.g., sugar consumption). This was problematic because samples need
93   to be at the same stage of fermentation to interpret the biological significance of differentially
94   expressed genes (3, 23). This is a common problem in time series experiments with multiple
95   groups and in some experimental systems there are strategies to combat this issue (23). For
96   example, in experiments that study the cell cycle, inhibitors arrest the cell cycle at the same
97   stage across groups thereby enabling comparisons (24). This experimental approach is not
98   applicable to wine fermentations where nutrient availability is more important than stage of
99   the cell cycle (25).

100   To address a similar issue, methods have recently been developed for the analysis of
101   single cell RNA sequencing data from differentiating cells. In these experiments, as cells
102   differentiate, absolute time may not reflect the extent of differentiation in each individual
103   cell (26). Consequently, pseudotime analysis has been used to reorder cells from absolute
104   time to the stage in differentiation relative to other cells undergoing the same process (26).
105   In particular, diffusion maps have been used to reorder asynchronous cell populations
106   because this analysis approach preserves relationships between samples (26). Diffusion
107   mapping is a manifold learning technique that uses information from the k most similar
108   samples to construct non-linear composites of the major sources of variation among samples

109   (27, 28). Diffusion mapping has strengths over other dimensionality reduction algorithms,

110   like principal component analysis (PCA) and t-distributed stochastic neighbor embedding

111   (tSNE) that are applied to sequencing data to identify sources of variation (29, 30). The major

112   advantages are that diffusion mapping is non-linear and robust to the "horseshoe effect"

113   (unlike PCA) (31), preserves distances between samples (unlike tSNE), and is insensitive to

114   sampling density. As a dimensionality reduction algorithm, diffusion maps extract latent

115   variables that are inferred from relationships in the data, which can be used to represent

116   composite sources of variation between samples.

117   Here, we use diffusion mapping to analyze time-series RNA-sequencing data from *S.*

118   *cerevisiae* during wine fermentation. We use the resulting diffusion maps to synchronize

119   gene expression across treatment groups and to extract latent variables, termed diffusion

120   components (DCs), which represent the dominant sources of structure in the data. Diffusion

121   maps *per se* provide no suggestion of the underlying genes that lead to separation of samples

122   along diffusion components; therefore, we apply continuous differential expression analysis

123   using each diffusion component to determine what genes vary among samples across a given

124   diffusion component. Notably, this method was not exclusively useful for the analysis of RNA

125   sequencing data from wine fermentations. We additionally apply this method to time series

126   data collected from *S. cerevisiae* during hypoxia, which showed that diffusion mapping

127   captures the dominant global shift in gene expression that occurs when yeast transitions

128   from aerobic to anaerobic metabolism. These findings suggest that our method enables

129   analysis of diverse asynchronous time series gene expression data, revealing both global

130   shifts and subtle differences in gene expression among groups. In the context of wine, we

131   find that diffusion mapping extracted the CGEP across Pinot noir fermentations, in addition

132   to distinguishing subtler differences between fermentations that reflect differences in the

133   grape musts. These findings offer important insights into variable wine fermentation and

134   sensory outcomes driven by vineyard specific cell-environment interactions.

Results and Discussion

Diffusion maps reorder asynchronous cell populations while preserving relationships between samples (26), which provides latent variables that represent non-linear composites of genes that vary between samples (**Figure 1**). We refer to these latent variables as *diffusion components*, the number of which is constrained by the number of samples in the data. Within each diffusion component (DC), a sample is represented by a single value and samples that have similar gene expression profiles among the genes captured in that component will have similar values. Moreover, samples at the origin of a DC (i.e., near 0) have gene expression profiles that do not vary along that component, while samples with positive or negative values diverge. Each DC captures diminishing structure among samples with the first diffusion component (DC1) accounting for the largest variation among all samples. Differential expression analysis can then be used to identify genes that vary among the samples along a given diffusion component. We expect the information extracted from each DC in this way will provide important insight into gene expression patterns caused by the time in fermentation (e.g., based here on Brix level) and vineyard site.

Diffusion mapping captures the global shift of gene expression during primary fermentation

We had previously performed inoculated primary fermentations of genetically similar Pinot noir grapes grown in 15 vineyard sites in California and Oregon over five vintages at the UC Davis Teaching and Research Winery (**Figure 2A**) (4, 9, 32, 33). We profiled the 2017 and 2019 vintages using time-course RNA sequencing data, with the aim of using *S. cerevisiae* gene expression as an indicator of similarities and differences across fermentations. In the 2017 vintage, we took samples approximately corresponding to early growth phase (16 hours), late growth phase (40 hours), stationary phase (64, 88 hours), and end of fermentation (112 hours) (**Figure 2B**). In the 2019 vintage, we shifted sampling to capture cellular adaptation after inoculation (2 and 6 hours) and reduced sampling later in fermentation (16, 64, and 112 hours). The initial grape musts varied in parameters like initial nitrogen, pH, malic acid, tartaric acid, non-*Saccharomyces* microbial profile, and elemental profile, while the final wines differed in volatile profiles and sensory characteristics (9, 32,

165   33). From these data we determined that *S. cerevisiae* had a consistent core gene expression

166   program (CGEP) across fermentations from different vineyards and vintages (4).

167   Given the variable inputs and sensory differences described for wines from these

168   vineyards (9), we expected that there would be differences in *S. cerevisiae* gene expression

169   that may include genes known to impact the sensory outcome of wine (34). Yet, in our

170   previous analysis (4), site-specific differences were hard to quantify because fermentations

171   progressed at different rates, even with rigorous control of temperature at a 200 L scale,

172   leading to asynchronous biological progression among samples with respect to sampling

173   time (**Figure 2C**). To address this issue and gain insight into vineyard-specific factors

174   altering fermentation outcomes, here we have applied diffusion mapping to extract DCs and

175   performed differential gene expression analysis using DC values as a continuous variable to

176   identify key gene expression patterns differentiating these fermentations.

177   In DC1, which accounts for the largest variation among all samples, we captured a

178   clear transition during fermentation that is observed by the ordering of samples across DC1

179   in both the 2017 and 2019 vintages based on Brix (**Figure 3A–B**). In a previous analysis, we

180   used Brix to perform continuous differential expression and identified the CGEP of *S.*

181   *cerevisiae* in these Pinot noir fermentations (4). To test whether diffusion mapping and DC1

182   captured the CGEP during fermentation, we performed differential expression over DC1 and

183   compared this to values calculated previously across the Brix variable (4). Log_2 fold change

184   values were strongly correlated between both methods of differential expression (**Figure**

185   **3C, Figure S1**), indicating that DC1 captured the dominant global shift in gene expression

186   during both 2017 and 2019 fermentations.

187   We next explored subsequent diffusion components (e.g., DC2 through DC8) to

188   determine other less dominant drivers of structure among samples. In general, within the

189   2019 data set we observed that lower diffusion components (DC2–DC4) captured structure

190   across samples relating to the time of fermentation (**Figure 4**). For example, we observed

191   separations along DCs 2-4 based on Brix levels and not vineyard site (compare **Figure 4A**

192   **and 4B**). These differences were driven by cellular remodeling in early fermentation (DC2,

193   DC3) and starvation during late fermentation (DC4), based on the differentially expressed

194   genes associated with each diffusion component (**Table S1**). In contrast, higher diffusion

195   components (DC5–DC8) showed patterns that suggest these DCs varied more based on

196   vineyard site than on time in fermentation (**Figure 5 and Table S1**). In addition, we saw the

197   total number of differentially expressed genes diminished as the diffusion component

198   number increased, indicating more subtle differences between samples (**Table 1**).

199

200   ***Table 1***: *Number of significantly differentially expressed genes for the top diffusion com-*

201   *ponents for the 2017 and 2019 vintages.*

Diffusion component	2017		2019	
	induced	repressed	induced	repressed
DC1	335	435	470	457
DC2	320	331	169	16
DC3	122	127	32	293
DC4	200	74	57	74
DC5	80	481	13	0
DC6	36	26	92	46
DC7	82	56	50	7
DC8	36	8	0	24

202

203   Notably, while we detected the global shift in gene expression during primary

204   fermentation from data collected in 2017 along DC1 (**Figure 3A**), we did not observe clear

205   delineations among higher DCs (**Figure S2**). All gene expression data are inherently noisy

206   (35), which is continuously improved upon by technical and methodological advancements

207   in sequencing. Applying these advances, we sequenced the 2019 vintage with unique

208   molecular identifier (UMI) barcodes and were able to demultiplex sequencing lanes and

209   remove PCR duplicates (36). With this improvement in data quality in the 2019 vintage over

210   2017, we expect that signals associated with higher diffusion components were better

211   captured in the 2019 sequencing data. These improvements are also likely reflected in the

212   number of differentially expressed genes across each DC in the 2017 vs. 2019 vintages

213   (**Table 1**). Consequently, our are focused on the 2019 data.

214

215   **Diffusion mapping identified global shifts in gene expression during hypoxia**

216   While diffusion maps successfully distinguished major fermentation transitions, we

217   wanted to confirm that this method could consistently report on other RNA-sequencing data

218   where global shifts in gene expression occur. To address this, we identified a publicly

219   available gene expression dataset during hypoxia in *S. cerevisiae* (GSE85595 and
220   GSE115171). Hypoxia occurs when a cell becomes oxygen limited, which is accompanied by
221   large scale reprograming of gene expression for continued growth (37). We applied diffusion
222   mapping to this dataset and observed an ordered time-dependent transition to a hypoxic
223   phenotype along DC1 (**Figure 6A**). Sample positions along DC1 show a rapid transition
224   within 5 minutes of nitrogen exposure, indicating a fast metabolic transition to hypoxia that
225   matured over the remainder of the time course. As part of this genetic reprogramming, a
226   transient shift in gene expression has been previously identified at ~30 minutes of the
227   hypoxic response and shown to partially overlap with the environmental stress response
228   (37). Within the diffusion mapping data, DC6 differentiated this transient state at 30 minutes
229   of hypoxia (**Figure 6B**).

230   We next investigated whether genes differentially expressed along DC1 matched
231   oxygen-regulated genes identified by previous studies of hypoxia. Across seven microarray
232   studies, 11 genes (3 aerobic, 8 hypoxic) were consistently identified as being involved in
233   aerobiosis or anaerobiosis in at least six studies (compiled by Bendjilali et al. in (37)). We
234   identified all 11 of these genes as differentially expressed ($p < 0.05$) along DC1. We further
235   compared our results to time-series RNA-sequencing profiles of wild type *S. cerevisiae* during
236   a hypoxic response (37). Along DC1, 239 of 291 (82.1%) aerobic genes were significantly
237   expressed prior to exposure to nitrogen, while 422 of 519 (81.3%) hypoxic genes were
238   significantly induced after prolonged exposure to nitrogen (**Table S2**). Genes expressed at
239   time zero were enriched in pathways like ribosome biogenesis, oxidative phosphorylation,
240   and sterol metabolic process, while genes induced after prolonged exposure to nitrogen
241   were enriched in pathways like oxidation reduction process, cell wall, glycogen metabolic
242   process, and glycolysis/gluconeogenesis (**Figure S3**). These findings align well with our
243   knowledge of the hypoxic transition in yeast (37, 38). Moreover, these results indicate that
244   diffusion mapping accurately captured global changes in gene expression during the hypoxic
245   shift, including transient gene expression states.

246

247   **Diffusion mapping detected metabolic remodeling throughout wine fermentation**

248   With the knowledge that diffusion mapping effectively identified gene expression
249   changes across two different data sets, we set out to investigate the gene expression patterns

250   captured within each DC in our wine fermentation datasets. In the 2019 data, along DC2
251   there were clear separations among the 2, 6, and 16 hour samples, while the 64 and 112 hour
252   samples fell on the origin (**Figure 4B**). Within the genes captured along this component,
253   arginine biosynthetic process was enriched in genes that were more highly expressed in the
254   2 hour samples (*ARG1*, *ARG3*, *ARG5,6*, *ARG8*) (**Figure S4, Figure S5**). Arginine is likely the
255   most abundant amino acid in Pinot noir grape must (39) and genes that encode proteins
256   involved in arginine biosynthesis are suppressed by the presence of arginine (40).
257   Expression of these biosynthetic genes in early fermentation likely reflects that *S. cerevisiae*
258   is yet adapted to the wine environment by 2 hours after inoculation. By 6 hours of
259   fermentation, expression of these genes decreases, potentially signaling completion of
260   cellular adaptation to the grape must environment. Interestingly, four of the 16 genes
261   (YMR244W, YPR078C, YGL117W, YER085C) expressed in the 2 hour samples have no known
262   function. Given that very few genes were differentially expressed at 2 hours and they were
263   enriched for arginine biosynthesis, one speculation is that these genes may have functions
264   related to nitrogen and arginine biosynthetic processes. Alternatively, expression of these
265   genes may be associated with other cellular processes associated with early adaptation to
266   the must environment.

267   The 6 hours samples segregated to the opposite extrema of DC2 and were the most
268   differentiated from the 2 hours samples along this component (**Figure 4B**). Glycolysis was
269   enriched among genes induced in these samples (**Figure S5**), which was also accompanied
270   by gene expression changes that support a transition to anaerobic metabolism. For example,
271   we detected induction of anaerobic translation elongation factor encoded by *ANB1* (**Figure
272   S4**), which is optimally expressed below 0.5 µmol/L O_2 (41), likely indicating must oxygen
273   levels at this timepoint. Genes important for cell wall processes were also induced at 6 hours,
274   with *TIR1-4* being four of the top five genes induced (**Figure S4**). These genes encode cell
275   wall mannoproteins that are required for anaerobic growth (42). These genes are also
276   important in DC3 for separating the 6 and 16 hours samples, along with many genes induced
277   by anaerobiosis that included *DAN1* and *PAU* genes (*PAU2*-PAU*5*, *PAU7*, *PAU8*, *PAU10-PAU12*,
278   *PAU15-PAU17*, *PAU19*, *PAU20*, *PAU23*, and *PAU24*) in the 16 hour samples (43, 44). Together,
279   the induction of these genes regulated in response to oxygen across DC2 and DC3 likely

280   signal the transition to anaerobiosis. In DC3, there were also many other biological

281   processes, cellular compartments, and molecular functions enriched among the 293 genes

282   that were induced in the 16 hour samples (**Figure S6**), consistent with a transition to an

283   active growth phase at this stage of fermentation. As diffusion components are ordered with

284   the most variation among samples occurring first, DC2 and DC3 demonstrated that early

285   metabolic remodeling was second only to gene expression changes that occur as Brix

286   decrease (e.g., captured in DC1) during fermentation.

287   Along DC4, we observed separation of the 64 hour samples from the 112 hour

288   samples. In the 64 hour samples, transmembrane transport, including amino acid and

289   polyamine transport, were enriched categories among the genes that were induced (**Figure**

290   **S7**). Many induced genes (*DUR3*, *DAL5*, *DAL7*) are involved in allantoin metabolism, a non-

291   preferred nitrogen source (**Figure S4**). Induction of these genes at 64 hours likely indicates

292   relief of nitrogen catabolite repression consistent with decreasing nitrogen concentrations

293   and nutrient availability. Genes repressed by the presence of amino acids were also induced

294   in the 112 hours samples (*GAT2, ARG3*). Interestingly, *HXT13* and *MAN2* were among the top

295   induced genes, along with *HXT17*, in the 112 hours samples (**Figure S4**). These two *HXT*

296   genes encode mannitol transporters (45) and *MAN2* encodes mannitol dehydrogenase.

297   Expression of these genes would enable *S. cerevisiae* to metabolize mannitol as a non-

298   preferred carbon source (45–47). Mannitol is produced by non-*Saccharomyces* organisms,

299   including lactic acid bacteria (48) and other non-*Saccharomyces* yeast (49). Expression of

300   these genes late in fermentation likely signals an increasingly generalized metabolic

301   program that is utilizing non-preferred carbon sources as preferred sugars were exhausted.

302   Overall, we interpret the patterns of separation along DC1-4 to be reflective of gene

303   expression changes occurring as *S. cerevisiae* proceeds through fermentation, adapts to the

304   increasingly nutrient limited environment, and deals with associated stresses. While these

305   changes appear common to the fermentations conducted here, future work will be required

306   to address if individual processes captured in DC2-4 occur in the context of other wine

307   strains and grape varieties or are unique to the wine yeast RC212 and Pinot noir

308   fermentations. Nonetheless, these observations indicate that this analysis approach is a

309   robust means for dealing with asynchronous gene expression data across fermentations.

310   Moreover, it raises many questions about the genes important for defining separation along

311 these DCs, including gene products involved in arginine, mannitol, and anaerobic

312 metabolism. Of notable interest are the large family of *PAU* genes, the vast majority of which

313 have no known function in *S. cerevisiae,* but have been previously noted for being induced

314 during fermentation and in response to stress (50).

315

316 **Lower diffusion components captured varied progression through fermentation**

317 Along lower diffusion components, patterns of separation were based on the time of

318 fermentation (discussed above), but outliers from select vineyard sites were also noted. We

319 expect this indicates site-specific differences that influence *S. cerevisiae* activities during

320 fermentation. For example, in the 2017 vintage, samples taken at 40 hours from SMV sites

321 and OR1 were shifted toward the 16 hour samples along DC1 indicating a different metabolic

322 trajectory (**Figure 3**). Interestingly, we detected *Lactobacillus kunkeei* transcripts in primary

323 fermentations from SMV sites (4, 32). *L. kunkeei* is a fructophilic lactic acid bacteria that

324 produces mannitol and lactic and acetic acids during wine fermentation (48). The

325 observation of an altered metabolic program in early fermentation with the unique presence

326 of *L. kunkeei* may indicate inter-species interactions that modulated *S. cerevisiae* gene

327 expression. In support of this, expression of genes involved in mannitol metabolism at 112

328 hours were the highest for the SMV sites (*HXT13, HXT17, MAN2*). One possibility is that *L.*

329 *kunkeei* may have induced a [GAR+] prion state, which is known to alter the metabolic

330 strategy employed by *S. cerevisiae* (51). Following these observations in 2017, we attempted

331 to detect the [GAR+] prion state in yeast isolated from the 2019 fermentations, but failed to

332 detect [GAR+] yeast.

333 Other outliers included SRH1, which in the 2019 vintage at 6 hours, and to a lesser

334 extent at 2 hours, did not cluster with other samples from these same time points along DC2

335 and DC3 (**Figure 4**). The SRH1 6 hour samples were instead shifted toward other 16 hour

336 samples, potentially indicating faster cellular adaptation to the wine environment. In

337 support of this, fermentations from SRH sites decreased in Brix faster than other sites in the

338 2017 vintage (**Figure 2C**). Another example involved 64 hour samples from OR1 and OR2

339 sites being shifted along DC4 toward the 112 hour cluster (**Figure 4**), which may relate to

340 nutrient conditions specific to OR sites (see further discussion below). Similarly, 112 hour

341 samples from SNC1 and AS2 were shifted toward the 64 hour cluster (**Figure 4**). We expect

342   these outliers in the lower DC components related to the time in fermentation reflect

343   differences between the musts (e.g., nutrient levels or presence of specific non-

344   *Saccharomyces* organisms) that impact *S. cerevisiae* metabolism and the timing of gene

345   expression transitions as fermentations progress.

346

347   **Higher diffusion components identified vineyard specific gene expression patterns**

348   The common patterns and existence of outliers across lower diffusion components,

349   which aligned with fermentation progression and presence of non-*Saccharomyces* species,

350   indicated that information about specific sites were captured by these analyses. Given that

351   some higher diffusion components clearly separated samples within a stage of fermentation

352   **(Figure 5)**, we used gene expression differences across the higher DCs to investigate

353   vineyard-specific patterns (**Figure S8** and **Table S1**). In this way, we aimed to identify *S.*

354   *cerevisiae* activities, inferred by the genes involved, specific to a vineyard site(s). We

355   specifically concentrated on the 2019 samples that separated to the extremes of each DC, as

356   this separation indicates that these samples were the most differentiated at the

357   transcriptome level. For example, at 2 hours SMV1, SMV2, SRH1, AV2, and RRV3 were most

358   separate from RRV2, CRN1, SNC1, and AS2 along DC5 (**Figure 5**). When comparing these

359   sites, a standout difference was the induction of genes in samples from SMV1, SMV2, SRH1,

360   AV2, and RRV3 that were enriched for vitamin metabolic and cell wall processes (**Figure S8A**

361   **and S9**). Given that co-culture experiments have demonstrated that *S. cerevisiae* induces

362   genes involved in cell wall remodeling and vitamin biosynthesis in response to the presence

363   of non-*Saccharomyces* yeasts (6, 52, 53), we tested whether detected gene expression by

364   non-*Saccharomyces* yeasts in 2 hour samples correlated with DC5. Indeed, DC5 values

365   correlated with total gene expression of *Hanseniaspora uvarum* (R^2 = 0.49, p < 0.001), but

366   not with total gene expression of other organisms (**Table 2**), suggesting that *H. uvarum*

367   activity leading up to these early fermentation samples may have impacted *S. cerevisiae*

368   metabolism. This is consistent with a previous study which reported that *S. cerevisiae*

369   remodels its cell wall in the presence of *H. uvarum* at three hours post-inoculation in a wine

370   fermentation (6). Interestingly, *PDC5* was among genes induced in fermentations with higher

371   gene expression by *H. uvarum* along DC5 (**Table S1**). *PDC5* encodes one of three isoforms of

372   pyruvate decarboxylase, an enzyme involved in the formation of flavor–active higher

373   *Table 2: Correlation between total gene expression by non-*Saccharomyces *organisms and*

374   *DC5.*

Organism	R^2	p value
Aureobasidium pullulans	-0.03555	0.947
Botrytis cinerea	-0.03261	0.774
Cladosporium sp SL 16	-0.03341	0.804
Hanseniaspora opuntiae	-0.03524	0.911
Hanseniaspora uvarum	0.490605	< 0.001
Lachancea thermotolerans	-0.02741	0.638
Metschnikowia fructicola	0.069637	0.086
Pichia kudriavzevii	-0.02819	0.654
Rhizopus stolonifer	-0.02439	0.582

375

376

377   alcohols in wine via the Ehrlich pathway (54). This suggests that the presence of *H. uvarum*

378   may lead to gene expression changes that directly impact wine sensory outcomes. The

379   activity of this non-*Saccharomyces* yeast may also relate to the altered metabolic program

380   detected in SMV1/2 and SRH1 fermentations within the lower diffusion components (as

381   discussed above). Notably, in a previous study we found that all fermentations had

382   detectable *H. uvarum* DNA, but only a subset had detectable *H. uvarum* gene expression (32).

383   Given the potential for *H. uvarum* to impact *S. cerevisiae* gene expression and metabolism, in

384   the future it will be important to determine the factors that lead to *H. uvarum* (in)activity in

385   select fermentations.

386   SMV and SRH are neighboring American Viticultural Areas in southern California

387   (**Figure 2A**). While samples from the SMV sites and SRH1 group together at 2 hours, we

388   detected separation of these fermentations at 16 hours along DC7 (**Figure 5**). This suggests

389   that while these sites were initially similar, they differed later in fermentation. While few

390   genes were significantly induced in SMV vs. SRH samples along DC7, *ADH4* was the top

391   induced gene (**Figure S8C**). *ADH* genes encode alcohol dehydrogenases that play an

392   important role in fermentation by facilitating transitions between acetaldehyde and ethanol

393   involving the redox cofactor NAD+. *ADH1* encodes the primary alcohol dehydrogenase

394   isoform responsible for this reaction during wine fermentation (55). Alcohol

395   dehydrogenases are also involved in the formation of fusel alcohols within the Ehrlich

396   pathway (56). As such, differences in *ADH4* gene expression could be an important site-

397   specific difference that has a role in *S. cerevisiae* metabolism and wine aroma development.

398   Other genes more highly expressed in SMV sites were involved in cell growth processes,

399   including translation (*MRP2* and *TIF2*), transcription (*MED1*), and cell division (*CLB6*)

400   (**Figure S8C**). In site SRH1, more highly expressed genes along DC7 vs. SMV were involved

401   in oxidative stress (*RCK1*) and sporulation (*SPO74* and *SSP1*). These site-specific differences

402   in gene expression involving growth vs. stress indicate varied fermentation environments

403   leading to altered gene expression at 16 hours. Given that genes associated with the Ehrlich

404   pathway and fusel alcohol anabolism differentiated SMV sites and SRH1 at 2 and 16 hours of

405   fermentation, the Ehrlich pathway may be an important component to consider in the

406   context of site-specific differences in these Pinot noir wines.

407   In the case of DC6, at 64 hours OR1/2 and RRV2 samples segregated to one extrema

408   (**Figure 5**). Genes induced in these samples were associated with nitrogen limitation (*DAL5*,

409   *PUT1*, *PUT2*; (1)), while genes involved in ammonia metabolism (*MEP3*, *SSY1*, *AUA1*) were

410   induced within fermentations from sites at the other extrema (**Table S1**). In line with these

411   patterns that reflect differences in nitrogen availability, DC6 values correlated with initial

412   grape must nitrogen as measured by an o-phthaldialdehyde assay (NOPA) and NH_3

413   measurements (initial NOPA: R^2 = 0.62, p < 0.001, initial NH_3: R^2 = 0.60, p < 0.001), led by

414   low initial nitrogen levels in OR1, OR2, and RRV2 (**Figure 7**). Interestingly, while initial

415   nitrogen levels in OR1, OR2, and RRV2 were the lowest among all fermentations, these

416   fermentations were supplemented approximately 24 hours after inoculation with a

417   combination of DAP and a complex nitrogen sources to adjust total yeast assimilable

418   nitrogen (YAN) levels to 250 mg/L. Yet, these data indicate persistent nitrogen limitation for

419   these sites, suggesting that these nitrogen additions may not have been sufficient to meet

420   nutrient requirements in these fermentations. These findings suggest that more research is

421   needed to understand the impact of nitrogen additions on fermentation, including the timing

422   of addition and the nitrogen source.

423   DC8 separated 112 hour samples with SNC1 and AS2 segregating from other samples

424   (**Figure 5**). Interestingly, 14 of the 24 genes induced in these samples are of unknown

425   function (**Table S1 and Figure S8D**). Among the induced genes with known functions were

426   *DDR2* and *HSP30*, which are stress related genes transcribed in response to a variety of

427 environmental or physiological stresses (57), as well as *YDL218W* that is induced in response

428 to the mycotoxin patulin produced by a variety of molds (58). Associated with these stress-

429 related genes were genes that function in meiosis and sexual reproduction, including *SPO74*,

430 *MFA1*, and *AFB1*. These data are suggestive of stresses in these fermentations that could be

431 driving the diploid wine yeast into meiosis and a sexual reproduction cycle. This is of

432 particular note, since the stresses associated with a wine fermentation environment are

433 thought to impart strong selective pressures that drive adaptive evolution (59). This is

434 reflected by the fact that *S. cerevisiae* strains associated with wine show a propensity for

435 genetic diversity, including many instances of hybridization (60). Future research will be

436 required to understand what particular stresses in SNC1 and AS2 are driving these unique

437 patterns of gene expression, in addition to what outcome this has on fermentation

438 performance.

439 Finally, across diffusion components, it is worth noting that fermentations from the

440 same American Viticultural Area (AVA) were commonly grouped together (**Figure 4, Figure**

441 **5**). For example, fermentations from OR, AV, and SMV sites often grouped together, providing

442 support for the concept of AVA and regional differences from the perspective of *S. cerevisiae*

443 gene expression. However, we did not observe grouping among all fermentations from the

444 same AVA along all diffusion components. For example, samples from AS grouped together

445 along DC5 (2 hours) and DC6 (64 hours), but not in DC8 (112 hours). The AS sites are

446 separated by 1 km and yet separation along DC8 suggests there was detectable variation in

447 *S. cerevisiae* metabolism in primary fermentation (**Figure 4, Figure 5**). Replicates from these

448 same vineyard site co-cluster, suggesting that lack of reproducibility in fermentations was

449 not a factor in this observation. Similarly, fermentations from RRV did not cluster together

450 along any diffusion component, suggesting that sub-appellations within the Russian River

451 Valley are associated with significantly different *S. cerevisiae* gene expression patterns

452 (**Figure 4, Figure 5**). This matches recent findings that show sub-regional variation in

453 elemental profiles of wine from the Russian River Valley (61). Importantly, the gene

454 expression differences we detected across each of these diffusion components provides

455 candidate genes and pathways that may underlie vineyard-specific fermentation outcomes.

456

457

458 **Conclusion**

459   In this study, we paired diffusion mapping with differential expression and captured

460 global shifts in gene expression. This method revealed differences in primary fermentation

461 of Pinot noir wine from 15 vineyard sites, as well as changes in *S. cerevisiae* gene expression

462 induced by hypoxia. Diffusion mapping was especially well suited for these studies as in both

463 cases cells progressed asynchronously through transcriptome changes with respect to

464 sampling time. Through our analysis of wine fermentations, we found vineyard-site specific

465 gene expression patterns that correlated with *H. uvarum* metabolic activity and initial

466 nitrogen composition of grape must, as well as indications of sexual reproduction in select

467 fermentations. Together, these data provide important insight into the wine fermentation

468 environment, including pathways, genes, and environmental factors that should be

469 considered in the context of differential fermentation outcomes. Given the tremendous

470 complexity of gene-environment interactions, we expect these data also serve to highlight

471 the large amount of work to be done to understand both the biological mechanisms at play

472 and how this knowledge can be applied by industry. In particular, we report transcriptomic

473 heterogeneity that arises from the same strain of yeast, fermented in the same facility, using

474 grape must from genetically identical plants. How this variability changes across the diverse

475 landscape of wine yeasts strains and fermentation environments (e.g., grape varieties and

476 associated chemical and microbiological profiles) remains to be seen. Importantly, we expect

477 that the approaches pioneered here for studying *S. cerevisiae* gene expression in a complex

478 environment using diffusion mapping provide an effective tool to probe these questions.

479   **Methods**

480

481   **Sampling, sequencing and preprocessing of wine fermentation samples**

482   The winemaking protocol has been described previously (9, 62). The grapes used in

483   this study originated from 15 vineyards in eight American Viticultural Areas in California

484   and Oregon, U.S.A. All grapes were Vitis vinifera L. cv. Pinot noir clone 667, with rootstock

485   101-14 (AV1, RRV1, SNC1, SNC2, CRN1, AS1, AS2, SMV1, SMV2, SRH1), Riparia Gloire (OR1,

486   OR2), or 3309C (AV2, RRV2, RRV3). Grapes were harvested at approximately 24 Brix and

487   transported to University of California, Davis Pilot Winery for fermentation. We performed

488   separate fermentations for grapes from each site, with two fermentations per site, totaling

489   of 20 fermentations per vintage (40 fermentations total). After harvest, the fruit was

490   separated into half-ton macrobins on harvest day and Inodose SO_2 added to 40 ppm. Bins

491   were stored in a 14°C cold room until destemming and dividing of the fruit into temperature

492   jacket-controlled tanks. N_2 sparging of the tank headspace was performed prior to

493   fermentation and tanks sealed with a rubber gasket. We cold soaked the must at 7°C for three

494   days and adjusted TSO_2 to 40 ppm on the second day. After three days, the must temperature

495   was increased to 21°C and programmed pump overs were used to hold the tank at a constant

496   temperature. We reconstituted *S. cerevisiae* RC212 with Superstart Rouge at 20 g/hL and

497   inoculated the must with 25 g/hL of yeast. At approximately 24 hours after inoculation,

498   nitrogen content in the fermentations was adjusted using DAP (target YAN – 35 mg/L – initial

499   YAN)/2, and Nutristart using 25 g/hL. Nitrogen was adjusted only if YAN was below 250

500   mg/L. Approximately 48 hours after fermentation, fermentation temperatures were

501   permitted to increase to 27°C, and again added DAP using the formula (target YAN - 35 mg/L

502   - initial YAN)/2, and fermentation were then continued until Brix < 0. Fermentation samples

503   were taken for Brix measurements and RNA isolation at 16, 64, and 112 hours relative to

504   inoculation. To ensure uniform sampling, a pumpover was performed ten minutes prior to

505   sampling each tank. For RNA samples, 12mL of juice was obtained, centrifuged at 4000 RPM

506   for 5 minutes, supernatant was discarded, and the cell pellet snap frozen in liquid nitrogen.

507   Samples were stored at -80°C until RNA extraction.

508

509

510 **RNA Extraction and Sequencing**

511   Yeast pellets were thawed on ice, resuspended in 5ml Nanopure water, centrifuged

512 at 2000g for 5min, and aspirated the supernatant. RNA was extracted using the Quick RNA

513 Fungal/Bacterial Miniprep kit including DNase l column treatment (cat#R2014, Zymo

514 Research). RNA was eluted in 30μL of molecular grade water and assessed for concentration

515 and quality via Nanodrop and RNA gel electrophoresis. Sample concentrations were adjusted

516 to 200ng/μl and used for sequencing. 3′ Tag-seq single-end sequencing (Lexogen QuantSeq)

517 was applied in both the 2017 and 2019 vintage, with the addition of UMI barcodes in 2019.

518 The University of California, Davis DNA Technologies Core performed all library preparation

519 and sequencing.

520   Sequencing samples were preprocessed according to manufacturer

521 recommendations. First, we hard-trimmed the first 12 base pairs from each read and

522 removed Illumina TruSeq adapters and poly A tails. Next, STAR was used to align reads

523 against *S. cerevisiae* S288C genome (R64, GCF_000146045.2) with parameters --

524 outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --

525 alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.6

526 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --

527 outSAMattributes NH HI NM MD --outSAMtype BAM SortedByCoordinate (63). For the 2019

528 vintage, UMI tools was used to deduplicate alignments (64). Reads mapping to each open

529 reading frame were quantified using htseq count (65).

530

531 **Hypoxia data set**

532   Gene expression counts data was downloaded using from GEO using accession

533 numbers GSE85595 and GSE115171.

534

535 **Construction of Diffusion Maps**

536   Diffusion maps were built as described previously (66). To build diffusion maps from

537 wine fermentations samples, k = 10 nearest samples was used, while for hypoxia k = 20 was

538 used. We increased the k size for hypoxia given the larger number of samples (n = 146 in

539 2017 vintage, n = 150 in 2019 vintage, and n = 336 in hypoxia). Prior to diffusion map

540 construction, gene counts were to non-mitochondrial mRNA and read counts normalized

541 based on total number of reads per sample (library size).

542

543 **Differential expression**

544 To determine which genes drove separation of samples along each component,

545 differential expression was used to correlate each gene with diffusion component values. The

546 R package limma was used to fit a linear regression model to each gene (20). As input to

547 differential expression, raw sequencing counts were used as input to differential expression,

548 and were filtered and normalized with the limma package using the calcNormFactors()

549 function (20). Using this model, the $\log_2$ fold change is the slope of the line for each unit

550 increase in the diffusion component. $\log_2$ fold change values were normalized by calculating

551 the length of the diffusion component and multiplying all $\log_2$ fold change values by this

552 amount, (max - minimum) * $\log_2$ fold change. $\log_2$ fold change values that were greater than

553 two were analyzed, i.e., genes with a $\log_2$ fold change of at least 2 between the most separated

554 samples along a diffusion component. Gene Ontology and KEGG enrichment analysis were

555 performed using the R clusterProfiler package (67).

556

557 **Data Availability**

558 Data is available in the Sequence Read Archive under accession PRJNA680606. All

559 analysis code is available at github.com/montpetitlab/Reiter_et_al_2020_DiffusionMapping.

560 **Acknowledgements**

References

569

570   1. Rossignol T, Dulau L, Julien A, Blondin B. 2003. Genome-wide monitoring of wine yeast

571   gene expression during alcoholic fermentation. Yeast 20:1369–1385.

572   2. Marks VD, Ho Sui SJ, Erasmus D, Van Der Merwe GK, Brumm J, Wasserman WW, Bryan J,

573   Van Vuuren HJ. 2008. Dynamics of the yeast transcriptome during wine fermentation

574   reveals a novel fermentation stress response. FEMS yeast research 8:35–52.

575   3. Rossouw D, Jolly N, Jacobson D, Bauer FF. 2012. The effect of scale on gene expression:

576   commercial versus laboratory wine fermentations. Applied microbiology and

577   biotechnology 93:1207–1219.

578   4. Reiter T, Montpetit R, Byer S, Frias I, Leon E, Viano R, Mcloughlin M, Halligan T,

579   Hernandez D, Runnebaum R, Montpetit B. 2021. Saccharomyces cerevisiae gene

580   expression during fermentation of Pinot noir wines at industrially relevant scale.

581   bioRxiv 2021.01.11.426308.

582   5. Molina AM, Swiegers JH, Varela C, Pretorius IS, Agosin E. 2007. Influence of wine

583   fermentation temperature on the synthesis of yeast-derived volatile aroma compounds.

584   Applied Microbiology and Biotechnology 77:675–687.

585   6. Curiel JA, Morales P, Gonzalez R, Tronchoni J. 2017. Different non-Saccharomyces yeast

586   species stimulate nutrient consumption in S. cerevisiae mixed cultures. Frontiers in

587   Microbiology 8:2121.

7. Carrau FM, Medina K, Farina L, Boido E, Henschke PA, Dellacassa E. 2008. Production of fermentation aroma compounds by Saccharomyces cerevisiae wine yeasts: effects of yeast assimilable nitrogen on two model strains. FEMS yeast research 8:1196–1207.

8. 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.

9. Cantu A, Lafontaine S, Frias I, Sokolowsky M, Yeh A, Lestringant P, Hjelmeland A, Byer S, Heymann H, Runnebaum RC. 2021. Investigating the impact of regionality on the sensorial and chemical aging characteristics of Pinot noir grown throughout the US West coast. Food Chemistry 337:127720.

10. Roullier-Gall C, Boutegrabet L, Gougeon RD, Schmitt-Kopplin P. 2014. A grape and wine chemodiversity comparison of different appellations in Burgundy: Vintage vs terroir effects. Food chemistry 152:100–107.

11. Duc C, Pradal M, Sanchez I, Noble J, Tesnière C, Blondin B. 2017. A set of nutrient limitations trigger yeast cell death in a nitrogen-dependent manner during wine alcoholic fermentation. PloS one 12:e0184838.

12. Tronchoni J, Curiel JA, Morales P, Torres-Pérez R, Gonzalez R. 2017. Early transcriptional response to biotic stress in mixed starter fermentations involving Saccharomyces cerevisiae and Torulaspora delbrueckii. International journal of food microbiology 241:60–68.

608   13. Alonso-del-Real J, Pérez-Torrado R, Querol A, Barrio E. 2019. Dominance of wine

609   Saccharomyces cerevisiae strains over S. kudriavzevii in industrial fermentation

610   competitions is related to an acceleration of nutrient uptake and utilization.

611   Environmental microbiology 21:1627–1644.

612   14. Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T,

613   Schroeder M, others. 1998. SGD: Saccharomyces genome database. Nucleic acids

614   research 26:73–79.

615   15. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown

616   PO. 2000. Genomic expression programs in the response of yeast cells to environmental

617   changes. Molecular biology of the cell 11:4241–4257.

618   16. Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, Jennings EG, Lee TI, True HL, Lander

619   ES, Young RA. 2001. Remodeling of yeast genome expression in response to

620   environmental changes. Molecular biology of the cell 12:323–337.

621   17. Conesa A, Nueda MJ, Ferrer A, Talón M. 2006. maSigPro: a method to identify

622   significantly differential expression profiles in time-course microarray experiments.

623   Bioinformatics 22:1096–1102.

624   18. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for

625   differential expression analysis of digital gene expression data. Bioinformatics 26:139–

626   140.

627   19. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion

628   for RNA-seq data with DESeq2. Genome biology 15:550.

629   20. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers

630   differential expression analyses for RNA-sequencing and microarray studies. Nucleic

631   acids research 43:e47–e47.

632   21. Jung I, Jo K, Kang H, Ahn H, Yu Y, Kim S. 2017. TimesVector: a vectorized clustering

633   approach to the analysis of time series transcriptome data from multiple phenotypes.

634   Bioinformatics 33:3827–3835.

635   22. Abu-Jamous B, Kelly S. 2018. Clust: automatic extraction of optimal co-expressed gene

636   clusters from gene expression data. Genome biology 19:1–11.

637   23. Bar-Joseph Z, Gitter A, Simon I. 2012. Studying and modelling dynamic biological

638   processes using time-series gene expression data. Nature Reviews Genetics 13:552–

639   564.

640   24. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D,

641   Futcher B. 1998. Comprehensive identification of cell cycle–regulated genes of the yeast

642   Saccharomyces cerevisiae by microarray hybridization. Molecular biology of the cell

643   9:3273–3297.

644   25. Bisson LF. 2019. Gene Expression in Yeasts During Wine Fermentation, p. 165–209. *In*

645   Yeasts in the Production of Wine. Springer.

646   26. Angerer P, Haghverdi L, Buttner M, Theis FJ, Marr C, Buettner F. 2015. destiny: diffusion

647   maps for large-scale single-cell data in R. Bioinformatics 32:1241–1243.

648   27. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F, Zucker SW. 2005.

649   Geometric diffusions as a tool for harmonic analysis and structure definition of data:

650   Diffusion maps. Proceedings of the national academy of sciences 102:7426–7431.

651   28. Coifman RR, Lafon S. 2006. Diffusion maps. Applied and computational harmonic

652   analysis 21:5–30.

653   29. Way GP, Zietz M, Rubinetti V, Himmelstein DS, Greene CS. 2020. Compressing gene

654   expression data using multiple latent space dimensionalities learns complementary

655   biological representations. Genome Biology 21:1–27.

656   30. Moon KR, van Dijk D, Wang Z, Gigante S, Burkhardt DB, Chen WS, Yim K, van den Elzen

657   A, Hirn MJ, Coifman RR, others. 2019. Visualizing structure and transitions in high-

658   dimensional biological data. Nature biotechnology 37:1482–1492.

659   31. Diaconis P, Goel S, Holmes S, others. 2008. Horseshoes in multidimensional scaling and

660   local kernel methods. The Annals of Applied Statistics 2:777–807.

661   32. Reiter T, Montpetit R, Byer S, Frias I, Leon E, Viano R, Mcloughlin M, Halligan T,

662   Hernandez D, Figueroa-Balderas R, Cantu D, Steenwerth K, Runnebaum R, Montpetit B.

663   2021. Transcriptomics provides a genetic signature of vineyard site with insight into

664   vintage-independent regional wine characteristics. bioRxiv 2021.01.07.425830.

665   33. Grainger C, Yeh A, Byer S, Hjelmeland A, Lima MM, Runnebaum RC. 2021. Vineyard site

666   impact on the elemental composition of Pinot noir wines. Food Chemistry 334:127386.

667   34. Cordente AG, Curtin CD, Varela C, Pretorius IS. 2012. Flavour-active wine yeasts.

668   Applied Microbiology and Biotechnology 96:601–618.

669   35. Hansen KD, Wu Z, Irizarry RA, Leek JT. 2011. Sequencing technology does not eliminate

670   biological variability. Nature biotechnology 29:572–573.

671   36. Lohman BK, Weber JN, Bolnick DI. 2016. Evaluation of TagSeq, a reliable low-cost

672   alternative for RNA seq. Molecular ecology resources 16:1315–1321.

673   37. Bendjilali N, MacLeon S, Kalra G, Willis SD, Hossian AN, Avery E, Wojtowicz O, Hickman

674   MJ. 2017. Time-course analysis of gene expression during the Saccharomyces cerevisiae

675   hypoxic response. G3: Genes, Genomes, Genetics 7:221–231.

676   38. Kwast KE, Lai L-C, Menda N, James DT, Aref S, Burke PV. 2002. Genomic analyses of

677   anaerobically induced genes in Saccharomyces cerevisiae: functional roles of Rox1 and

678   other factors in mediating the anoxic response. Journal of Bacteriology 184:250–265.

679   39. Stines A, Grubb J, Gockowiak H, Henschke PA, Høj P, Van Heeswijck R. 2000. Proline and

680   arginine accumulation in developing berries of Vitis vinifera L. in Australian vineyards:

681   Influence of vine cultivar, berry maturity and tissue type. Australian Journal of Grape

682   and Wine Research 6:150–158.

683   40. Crabeel M, Lavalle R, Glansdorff N. 1990. Arginine-specific repression in Saccharomyces

684   cerevisiae: kinetic data on ARG1 and ARG3 mRNA transcription and stability support a

685   transcriptional control mechanism. Molecular and cellular biology 10:1226–1233.

686   41. Kwast KE, Burke PV, Poyton RO. 1998. Oxygen sensing and the transcriptional

687   regulation of oxygen-responsive genes in yeast. Journal of Experimental Biology

688   201:1177–1195.

689   42. Abramova N, Sertil O, Mehta S, Lowry CV. 2001. Reciprocal regulation of anaerobic and

690   aerobic cell wall mannoprotein gene expression in Saccharomyces cerevisiae. Journal of

691   bacteriology 183:2881–2887.

692   43. Sertil O, Cohen BD, Davies KJ, Lowry CV. 1997. The DAN1 gene of S. cerevisiae is

693   regulated in parallel with the hypoxic genes, but by a different mechanism. Gene

694   192:199–205.

695   44. Rachidi N, Martinez M-J, Barre P, Blondin B. 2000. Saccharomyces cerevisiae PAU genes

696   are induced by anaerobiosis. Molecular microbiology 35:1421–1430.

697   45. Jordan P, Choe J-Y, Boles E, Oreb M. 2016. Hxt13, Hxt15, Hxt16 and Hxt17 from

698   Saccharomyces cerevisiae represent a novel type of polyol transporters. Scientific

699   Reports 6.

700   46. Quain DE, Boulton CA. 1987. Growth and Metabolism of Mannitol by Strains of

701   Saccharomyces cerevisiae. Microbiology 133:1675–1684.

702   47. Ramakrishnan V, Walker GA, Fan Q, Ogawa M, Luo Y, Luong P, Joseph C, Bisson LF.

703   2016. Inter-kingdom modification of metabolic behavior:[GAR+] prion induction in

704   Saccharomyces cerevisiae mediated by wine ecosystem bacteria. Frontiers in Ecology

705   and Evolution 4:137.

706   48. Bisson LF, Walker G, Ramakrishnan V, Luo Y, Fan Q, Wiemer E, Luong P, Ogawa M,

707   Joseph L. 2017. The two faces of Lactobacillus kunkeei: Wine spoilage agent and bee

708   probiotic. Catalyst: Discovery into Practice 1:1–11.

709   49. Mbuyane LL, de Kock M, Bauer FF, Divol B. 2018. Torulaspora delbrueckii produces

710   high levels of C5 and C6 polyols during wine fermentations. FEMS yeast research

711   18:foy084.

712   50. Luo Z, van Vuuren HJ. 2009. Functional analyses of PAU genes in Saccharomyces

713   cerevisiae. Microbiology 155:4036–4049.

714   51. Jarosz DF, Brown JC, Walker GA, Datta MS, Ung WL, Lancaster AK, Rotem A, Chang A,

715   Newby GA, Weitz DA, others. 2014. Cross-kingdom chemical communication drives a

716   heritable, mutually beneficial prion-based transformation of metabolism. Cell

717   158:1083–1093.

718   52. Barbosa C, Mendes-Faia A, Lage P, Mira NP, Mendes-Ferreira A. 2015. Genomic

719   expression program of Saccharomyces cerevisiae along a mixed-culture wine

720   fermentation with Hanseniaspora guilliermondii. Microbial cell factories 14:124.

721 53. Kosel J, Čadež N, Schuller D, Carreto L, Franco-Duarte R, Raspor P. 2017. The influence

722   of Dekkera bruxellensis on the transcriptome of Saccharomyces cerevisiae and on the

723   aromatic profile of synthetic wine must. FEMS Yeast Research 17:fox018.

724 54. ter Schure EG, Flikweert MT, van Dijken JP, Pronk JT, Verrips CT. 1998. Pyruvate

725   decarboxylase catalyzes decarboxylation of branched-chain 2-oxo acids but is not

726   essential for fusel alcohol production by Saccharomyces cerevisiae. Applied and

727   Environmental Microbiology 64:1303–1307.

728 55. Ciriacy M. 1975. Genetics of alcohol dehydrogenase inSaccharomyces cerevisiac.

729   Molecular and General Genetics MGG 138:157–164.

730 56. Dickinson JR, Salgado LEJ, Hewlins MJ. 2003. The catabolism of amino acids to long

731   chain and complex alcohols in Saccharomyces cerevisiae. Journal of Biological

732   Chemistry 278:8028–8034.

733 57. Kobayashi N, Mcclanahan TK, Simon JR, Treger JM, Mcentee K. 1996. Structure and

734   Functional Analysis of the Multistress Response GeneDDR2fromSaccharomyces

735   cerevisiae. Biochemical and biophysical research communications 229:540–547.

736 58. Iwahashi Y, Hosoda H, Park J-H, Lee J-H, Suzuki Y, Kitagawa E, Murata SM, Jwa N-S, Gu

737   M-B, Iwahashi H. 2006. Mechanisms of patulin toxicity under conditions that inhibit

738   yeast growth. Journal of agricultural and food chemistry 54:1936–1942.

739 59. Conacher CG, Rossouw D, Bauer F. 2019. Peer pressure: evolutionary responses to

740   biotic pressures in wine yeasts. FEMS Yeast Research 19:foz072.

741   60. Marsit S, Dequin S. 2015. Diversity and adaptive evolution of Saccharomyces wine

742   yeast: a review. FEMS Yeast Research 15:fov067.

743   61. Tanabe CK, Nelson J, Boulton RB, Ebeler SE, Hopfer H. 2020. The Use of Macro, Micro,

744   and Trace Elemental Profiles to Differentiate Commercial Single Vineyard Pinot noir

745   Wines at a Sub-Regional Level. Molecules 25:2552.

746   62. Grainger C, Yeh A, Byer S, Hjelmeland A, Lima MM, Runnebaum RC. 2021. Vineyard site

747   impact on the elemental composition of Pinot noir wines. Food Chemistry 334:127386.

748   63. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,

749   Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

750   64. Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in Unique

751   Molecular Identifiers to improve quantification accuracy. Genome research 27:491–

752   499.

753   65. Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-

754   throughput sequencing data. Bioinformatics 31:166–169.

755   66. Fahimipour AK, Gross T. 2020. Mapping the bacterial metabolic niche space. Nature

756   communications 11:1–8.

757   67. Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing

758   biological themes among gene clusters. Omics: a journal of integrative biology 16:284–
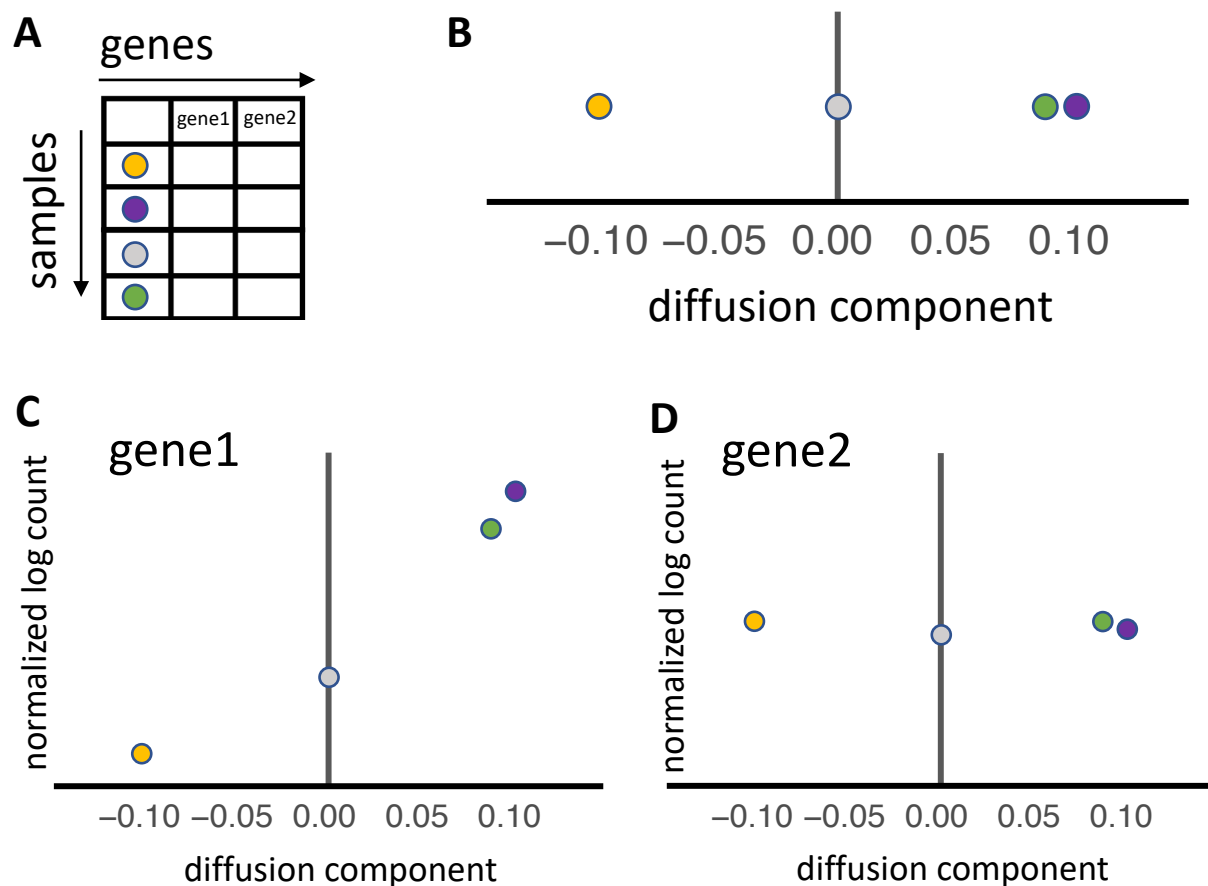
759   287.

760

Figure 1: Extracting data from diffusion maps. Diffusion maps provide the underlying manifold in gene expression data through non-linear dimensionality reduction. **A)** When applied to many genes across many samples, diffusion maps extract features that represent combinations of genes that drive similarities and differences among samples. **B)** These features are termed diffusion components. Samples at either extreme of the diffusion component are the most different from each other, while samples that fall at the origin are invariant along that component. In the above graphic, the orange and purple dots are the most different, while the purple and green dot are most similar. Diffusion maps do not provide information on which genes lead to separation of samples along each diffusion component. **C)** Performing differential expression using the diffusion component as a continuous variable reveals the genes that significantly contribute to separation of samples. In the graphic, gene 1 has significantly higher expression in samples that fall on the right extrema of the diffusion component, as compared to samples that fall on the left extrema. **D)** While all genes are used to perform differential expression, not all genes are differentially expressed along an individual diffusion component. In this example, gene2 is not differentially expressed along the diffusion component.
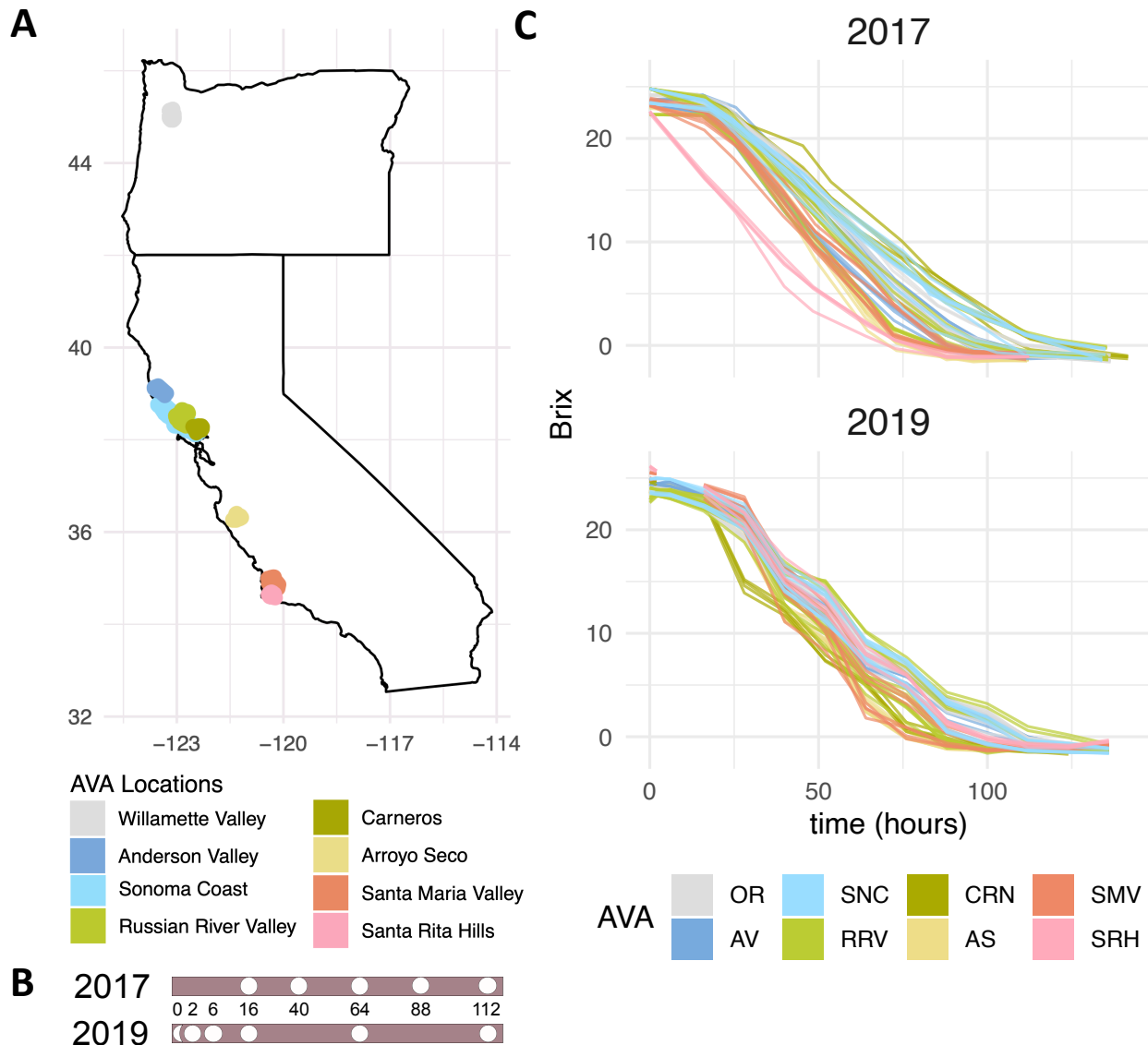
Figure 2: AVA locations of vineyard sites and sampling info across wine fermentations. A) Map of vineyard site American Viticultural Areas represented in this study. **B)** Primary fermentation sampling time points for 2017 and 2019 vintages. Times are shown in hours and are relative to inoculation. **C)** Vineyard site fermentation kinetics where Brix (total soluble solids) is used as a proxy for sugar levels.
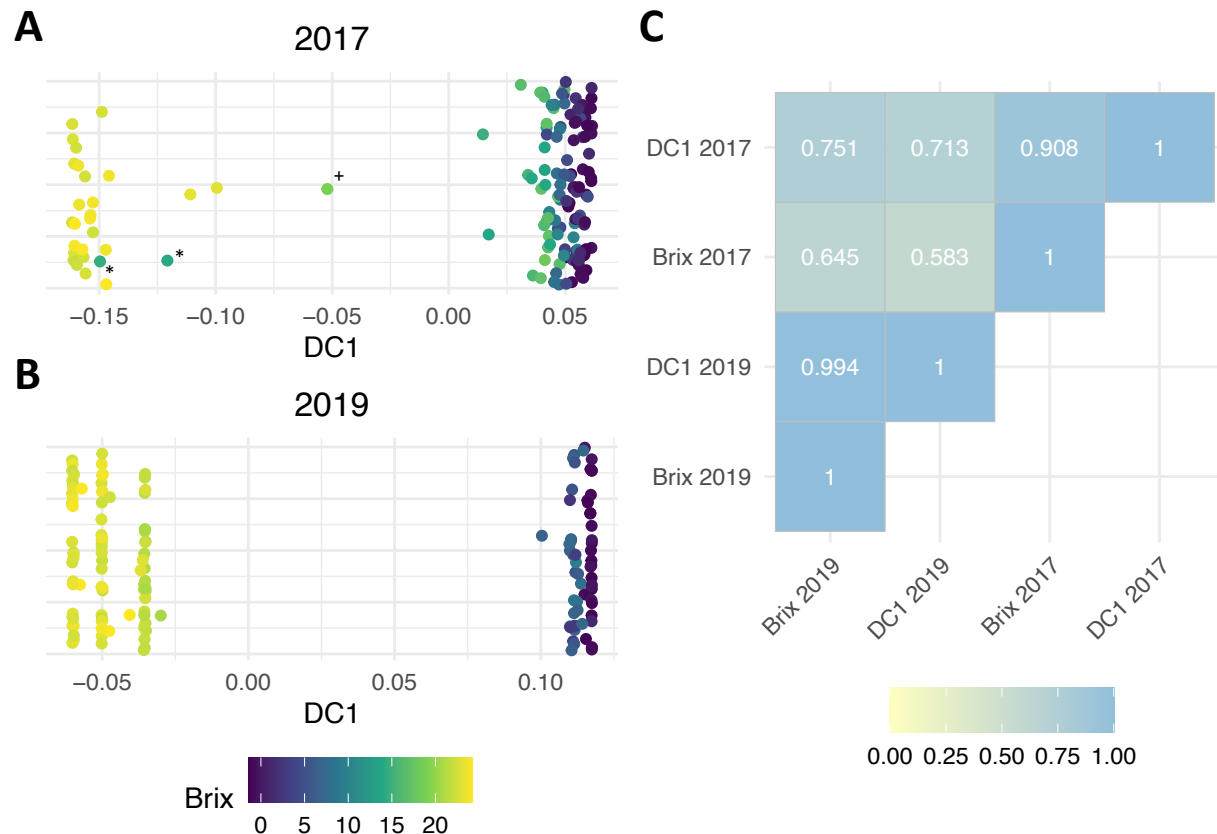
Figure 3: Diffusion mapping captures the core gene expression program of S. cerevisiae during wine fermentation. A, B) DC1 captures the metabolic transition that occurs as Brix decreases during fermentation. The same transition occurs in the 2017 vintage and the 2019 vintage. Each point represents a sample from one time from one fermentation. Points that are closer along the x axis are more similar. The y axis is jittered to allow all points to be visualized. Points are colored by Brix, a proxy for sugar concentration during fermentation, where Brix = 0 indicates end of fermentation. **C**) Graphic displays calculated correlations between differentially expressed genes in diffusion component 1 and genes that are differentially expressed as Brix decreases. Select samples in panel 3A are marked (∗ = SMV and + = OR sites) for ease of identification.
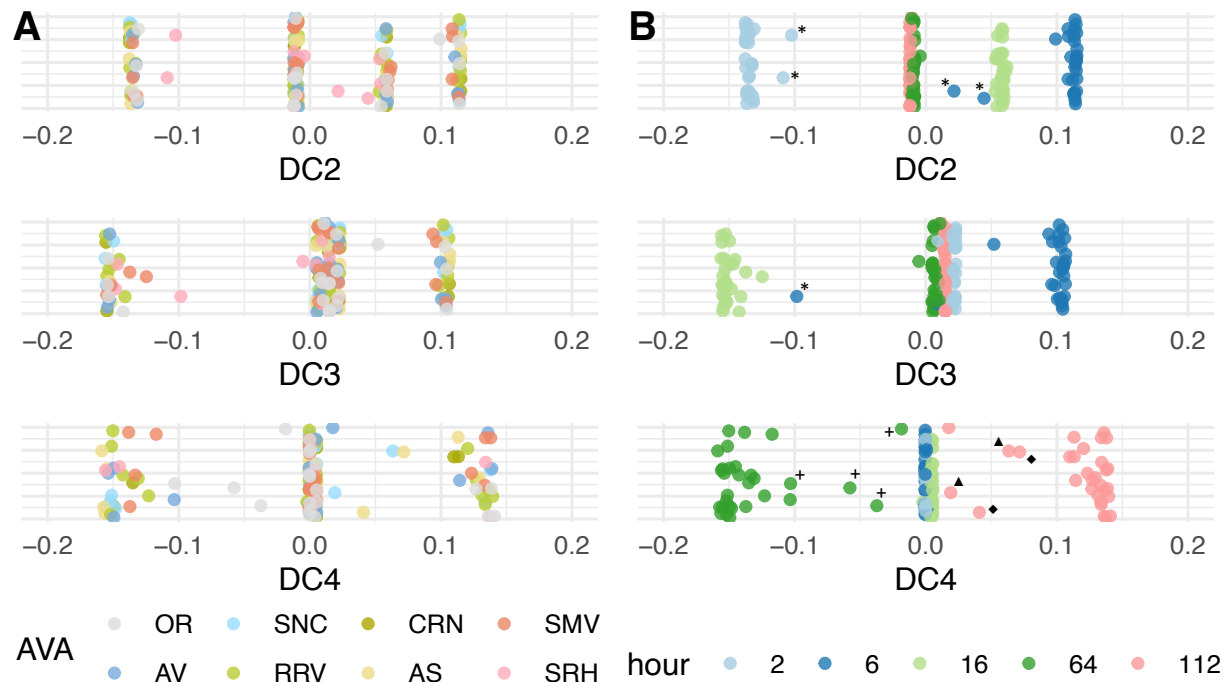
Figure 4: One-dimensional plots of DC2–DC4 for the 2019 vintage. Plots are colored by **A**) American Viticultural Area (AVA) and **B**) hours post inoculation and show that diffusion components capture different latent structure in the dataset. Each diffusion component represents a different relationship among samples with these components appearing to capture shifts between stages of fermentation, not vineyard site. Select samples in panel 4B are marked (∗ = SRH, + = OR, ▲ = SNC, and ◆ = AS sites) for ease of identification.
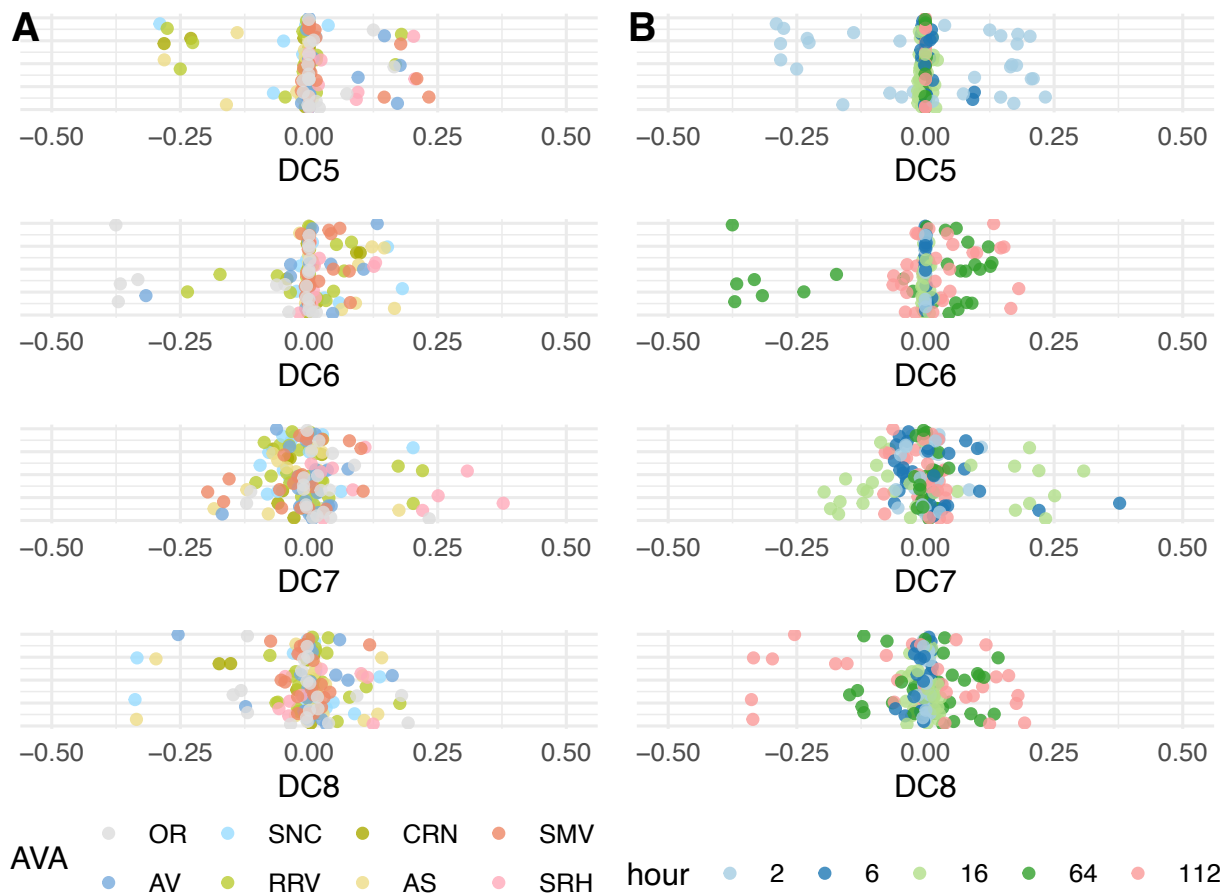
Figure 5: One-dimensional plots of DC5–DC8 for the 2019 vintage. Plots are colored by **A**) American Viticultural Area (AVA) and **B**) hours post inoculation. Higher diffusion components capture differences between vineyard site within a stage of fermentation, as seen by samples clustering together along a single or multiple diffusion components based on AVA, not Brix, indicating similarity in gene expression profiles within an AVA.
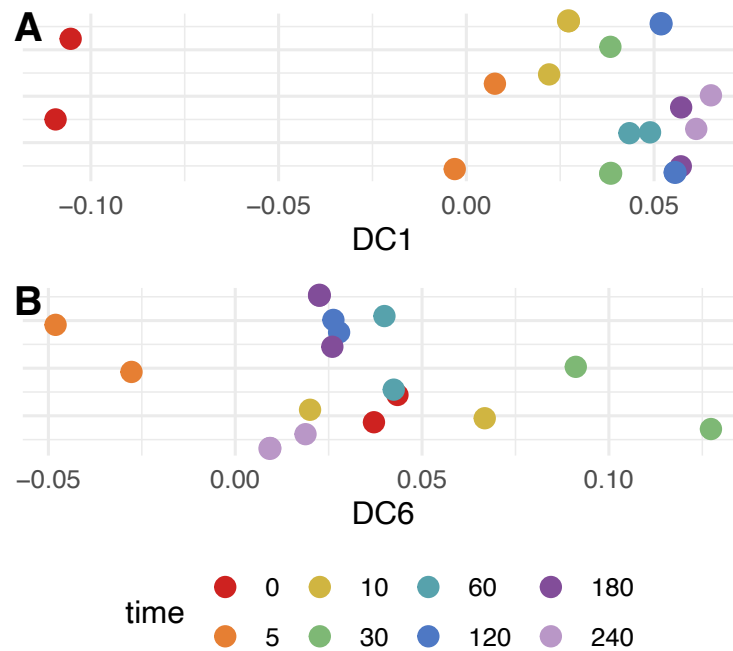
Figure 6: Diffusion map of S. cerevisiae exposed to nitrogen for 0–240 minutes. The trajectory of samples are displayed along **A**) DC1 that captured the transition from aerobic to anaerobic metabolism and **B**) DC6 that captured a transient transcriptome remodeling at 30 min.
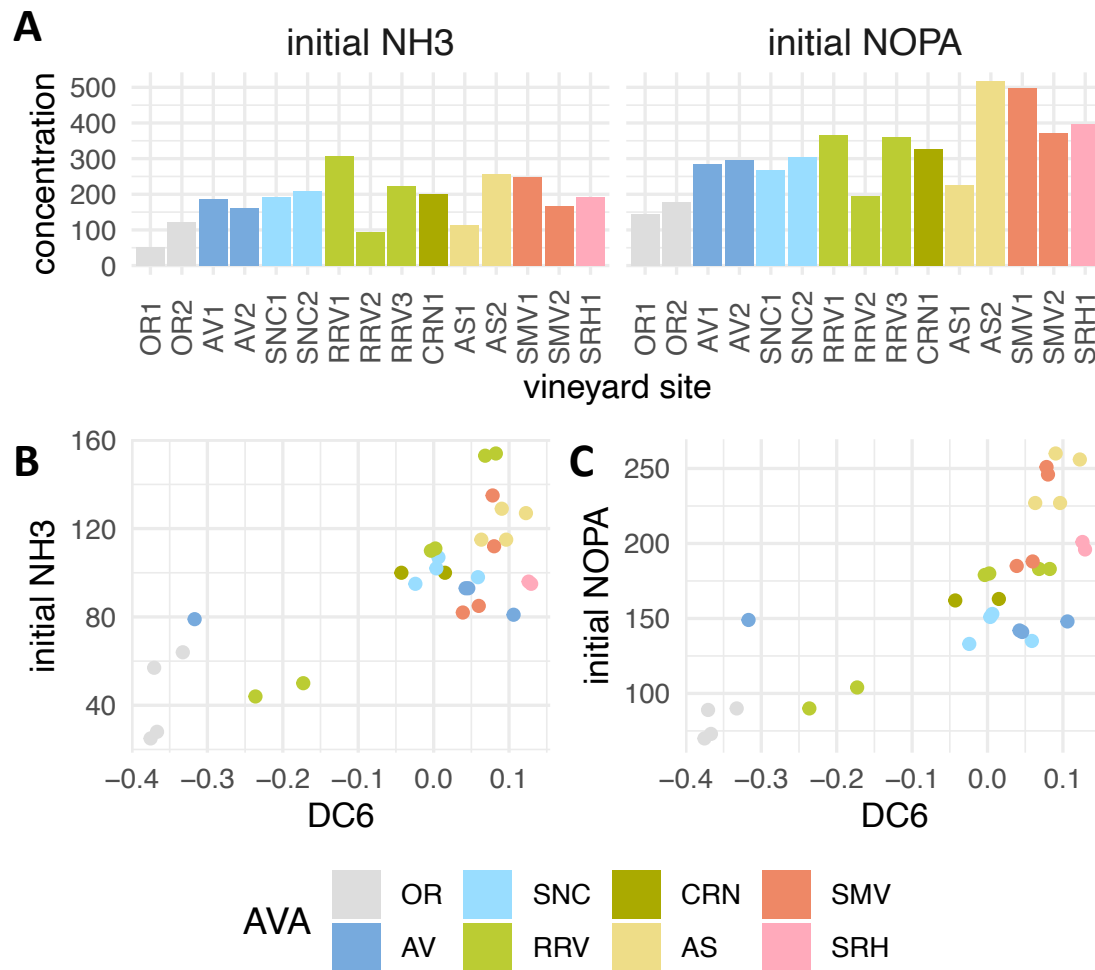
Figure 7: DC6 separates samples by initial nitrogen concentration in the grape must. A) Initial concentration of NH_3 and nitrogen by o-phthaldialdehyde assay (NOPA) was lower in fermentations from OR1, OR2, and RRV2 than other vineyards. **B, C)** Initial NH3 and NOPA in grape must correlates with DC6, driven by low nitrogen in fermentations from vineyard sites OR1, OR2, RRV3.
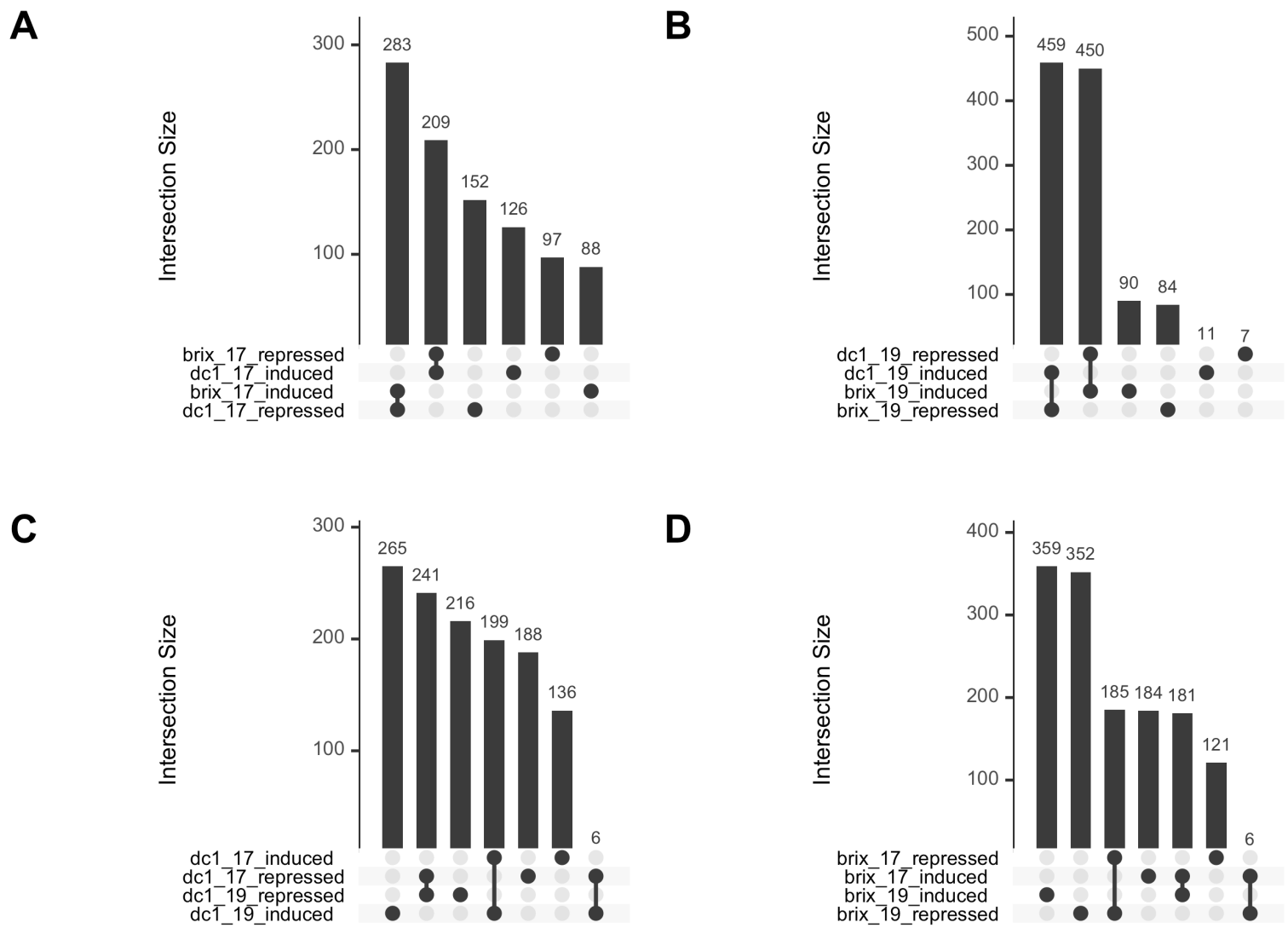
Figure S1: Upset plot depicting intersections between differentially expressed genes using diffusion component 1 or Brix as continuous variable. A) Intersection of differentially expressed genes from the 2017 vintage. **B**) Intersection of genes from the 2019 vintage. **C**) Intersection of genes differentially expressed along diffusion component 1 from the 2017 and 2019 vintages. **D**) Intersection of genes differentially expressed relative to change in Brix from the 2017 and 2019 vintages.
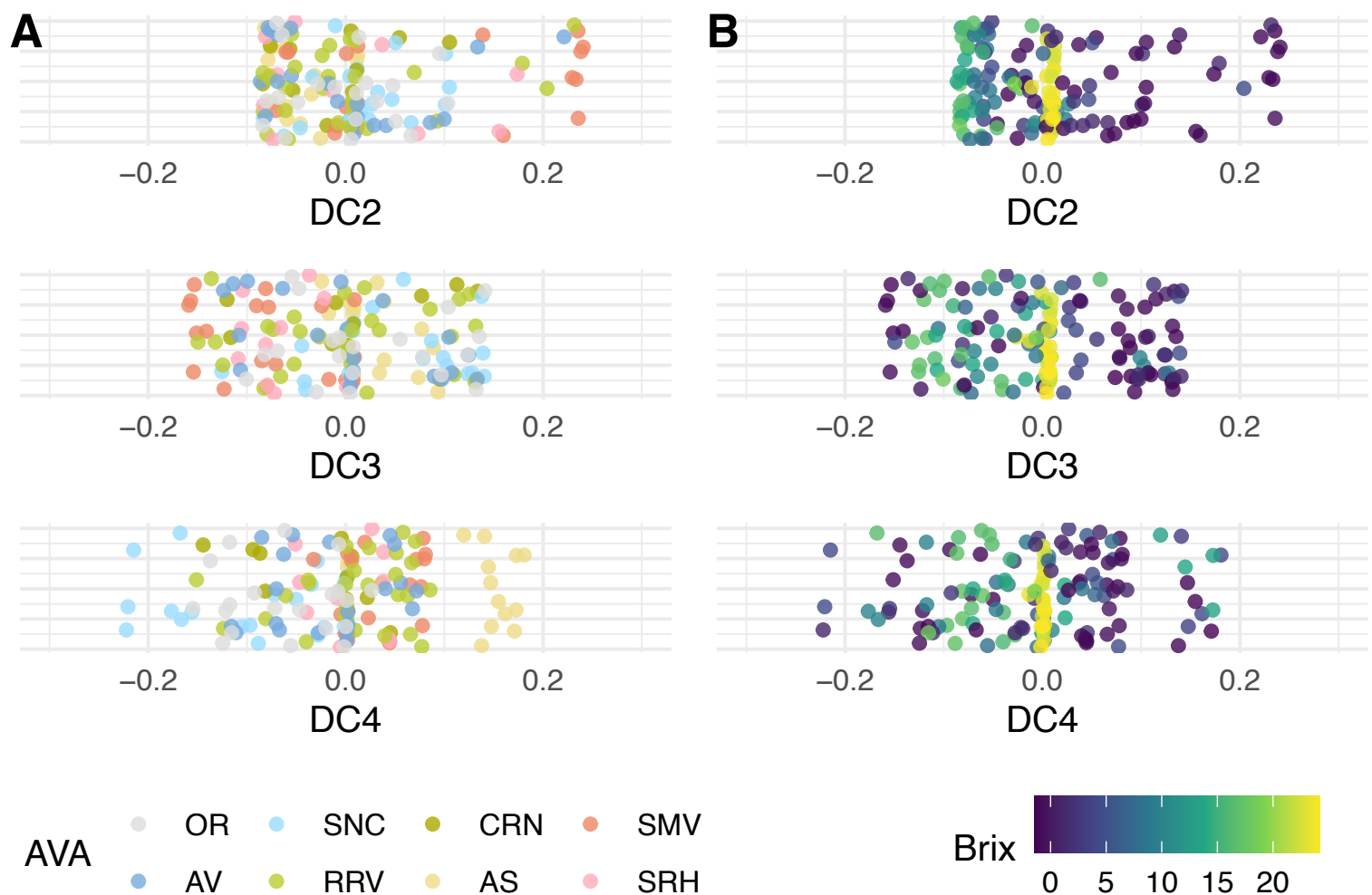
**Figure S2: One-dimensional plots of DC2–DC4 for the 2017 vintage**. Plots are colored by **A**) American Viticultural Area (AVA) and **B**) Brix. Diffusion components capture different latent structure in the data set. Unlike in the 2019 vintage, samples in lower diffusion components do not cluster, potentially indicating a data quality issue that was remediated in the 2019 vintage using UMI barcodes.

Figure S3: Gene set enrichment for differentially expressed genes along DC1 during onset of hypoxia. All enriched categories with p < 0.05 after Bonferroni correction are shown. Pathways on the right side of the figure are induced in samples with a high diffusion component value, while pathways on the left of the figure are induced in samples with a low diffusion component value.

A

Gene	logFC
ANS1	−6.8
YMR244W	−5.4
YPR078C	−3.6
ARG3	−3.0
TAT1	−2.9
ARG8	−2.7
YGL117W	−2.6
ARG1	−2.6
YER085C	−2.6
BIO3	−2.5
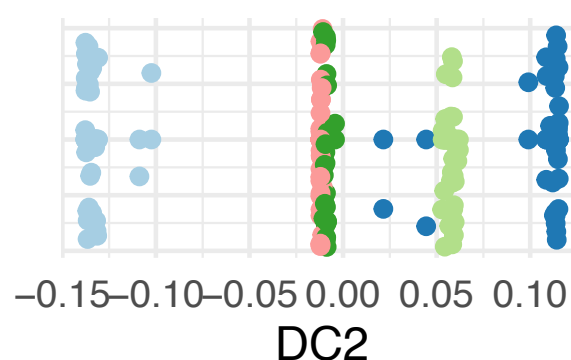
Gene	logFC
TIR1	8.7
TIR2	6.2
TOS6	5.4
TIR4	4.9
TIR3	4.7
SRL1	4.4
CWP1	4.3
ADE17	4.2
ANB1	4.1
HUG1	4.1

DC2

hour   2   6   16   64   112

B

Gene	logFC
PAU17	−8.4
TIR1	−8.0
DAN1	−7.2
TIR3	−6.9
PAU24	−6.1
MET17	−6.0
TIR2	−5.9
PAU5	−5.6
PAU11	−5.2
AUS1	−5.2
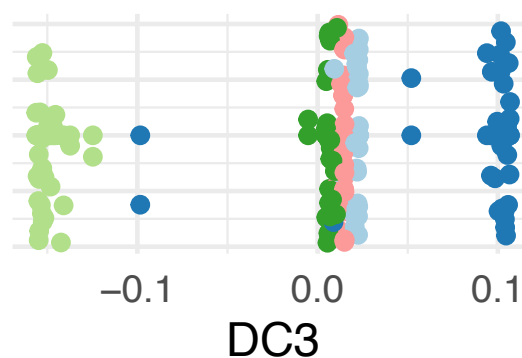
Gene	logFC
TAT1	4.9
FRE4	3.7
ANS1	3.7
FET3	2.8
YMR244W	2.5
FCF2	2.3
YBL028C	2.3
DBP9	2.3
RNT1	2.2
MAK16	2.2

DC3

hour   2   6   16   64   112

C

Gene	logFC
YCL048W−A	−4.1
DUR3	−3.8
OPT2	−3.3
PMA2	−3.2
AAD6	−3.2
BNA4	−3.1
THI72	−2.8
AIM20	−2.8
DAL5	−2.7
HSP33	−2.7
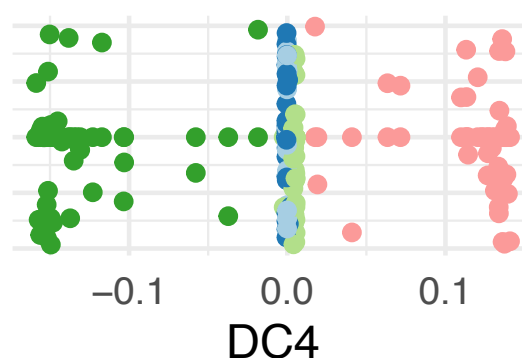
Gene	logFC
HXT13	4.9
MAN2	3.8
ECM11	3.3
DSF1	3.2
YKR045C	3.2
HXT17	3.1
YIR014W	3.1
YJL144W	3.0
INO1	2.9
NPL3	2.8

DC4

hour   2   6   16   64   112

Figure S4: Top 10 genes induced along each diffusion components that separate stages of fermentation in the 2019 vintage. Sites on the diffusion map are colored by hour at which samples were taken. Full gene lists are provided in Supplemental Table 1.

Figure S5: Gene set enrichment for differentially expressed genes along DC2 in the 2019 vintage. All enriched categories with p < 0.05 after Bonferroni correction are shown. Pathways on the right side of the figure are induced in samples with a high diffusion component value, while pathways on the left of the figure are induced in samples with a low diffusion component value.
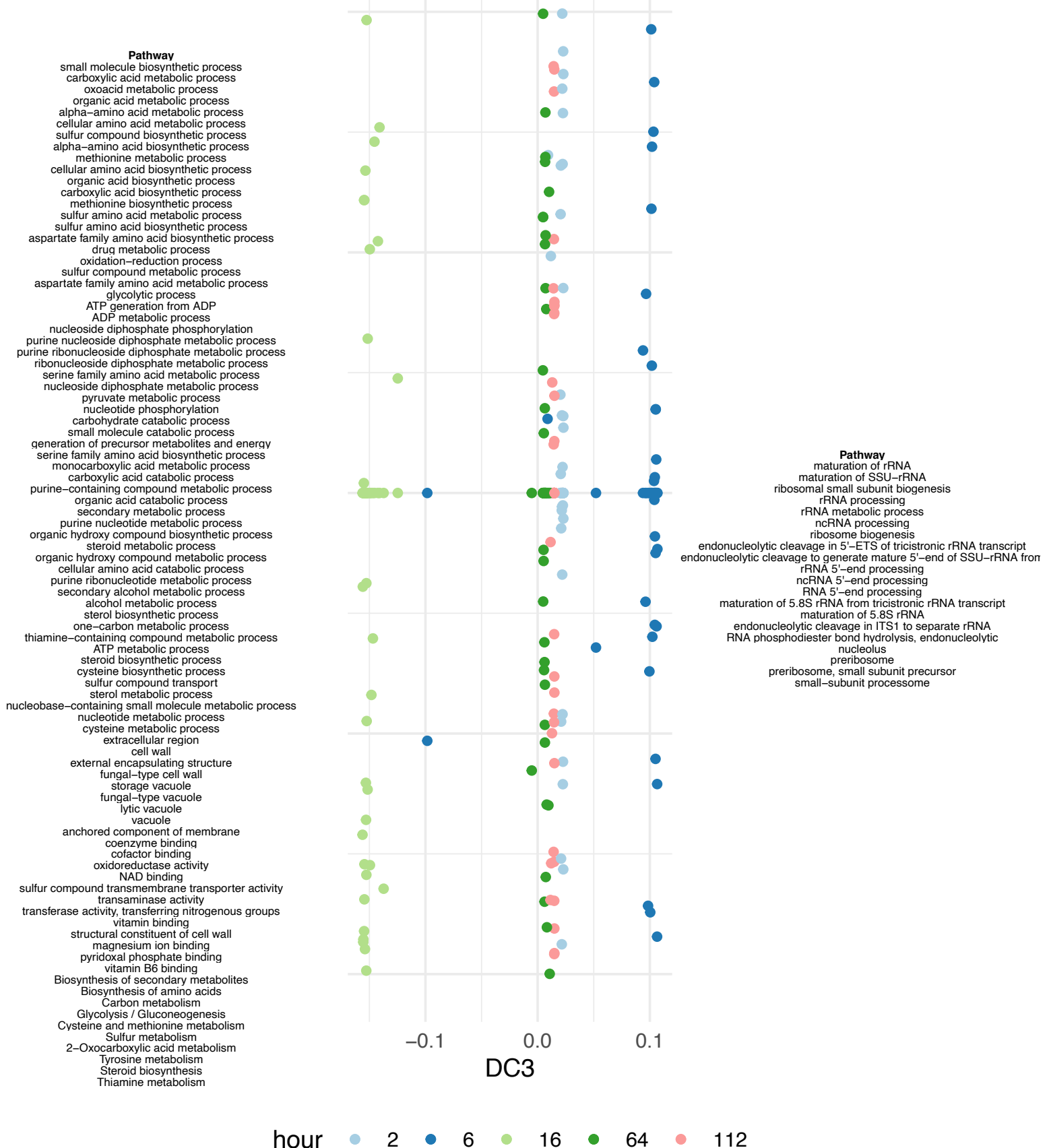
Figure S6: Gene set enrichment for differentially expressed genes along DC3 in the 2019 vintage. All enriched categories with p < 0.05 after Bonferroni correction are shown. Pathways on the right side of the figure are induced in samples with a high diffusion component value, while pathways on the left of the figure are induced in samples with a low diffusion component value.
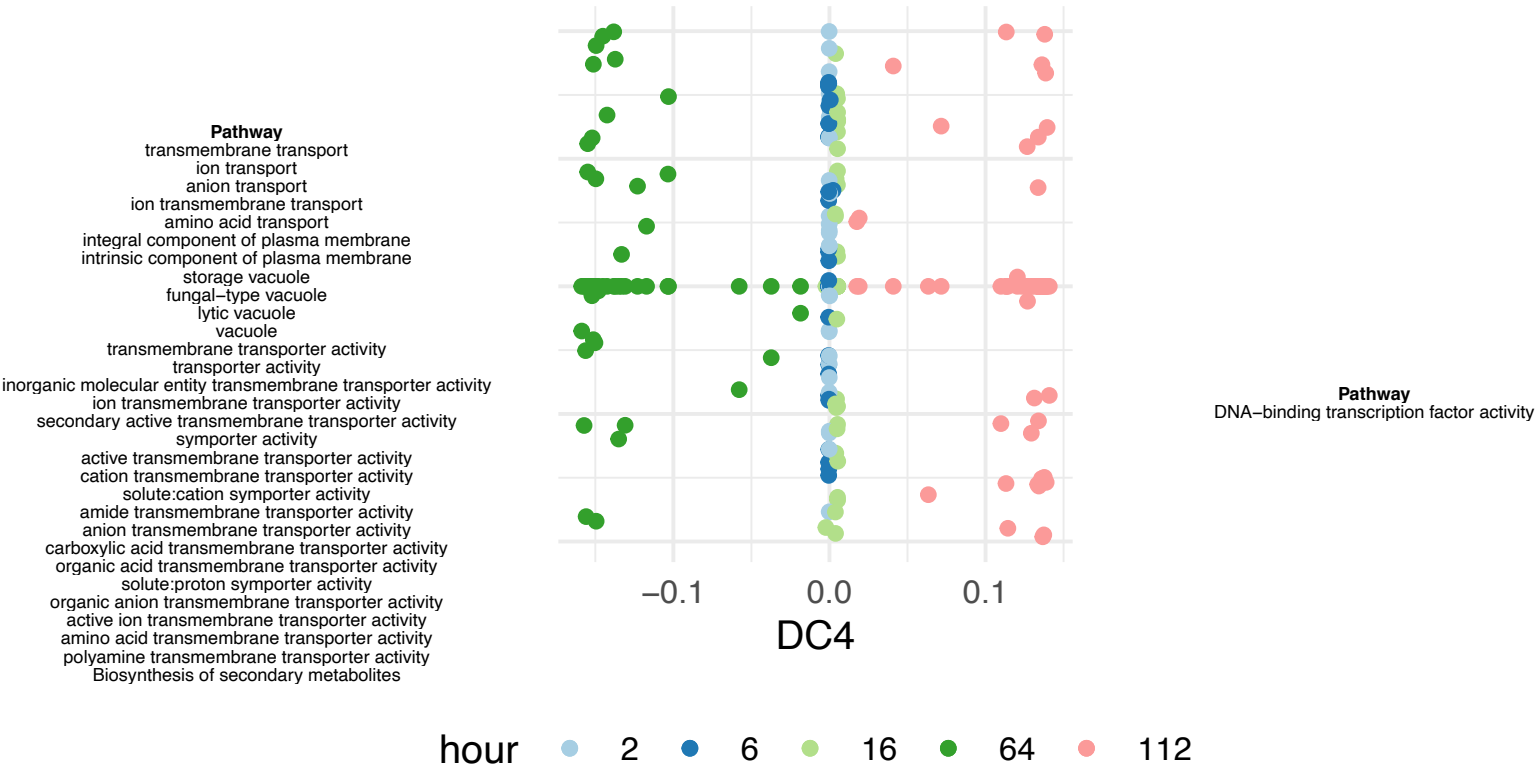
Figure S7: Gene set enrichment for differentially expressed genes along DC4 in the 2019 vintage. All enriched categories with p < 0.05 after Bonferroni correction are shown. Pathways on the right side of the figure are induced in samples with a high diffusion component value, while pathways on the left of the figure are induced in samples with a low diffusion component value.

A

No_sig_genes

Gene	logFC
ANB1	3.6
PDC5	3.5
THI73	2.6
DSE2	2.4
PET9	2.4
SCW11	2.4
MAL12	2.3
BIO3	2.2
YBR056W–A	2.1
PHO3	2.1

DC5

AVA

OR   SNC   CRN   SMV
AV   RRV   AS   SRH

B

Gene	logFC
APE1	−2.9
ADY3	−2.9
WHI5	−2.8
MST28	−2.7
RGM1	−2.7
RRT5	−2.7
YIR014W	−2.6
DAL5	−2.6
PTC2	−2.6
YAP7	−2.6

Gene	logFC
RTA1	3.9
YER181C	3.9
YKR032W	3.9
AGP1	3.5
YEL008W	3.4
HXT13	3.3
YDR034W–B	3.3
SCW10	3.3
YOR314W	3.0
ORM2	3.0

DC6

AVA

OR   SNC   CRN   SMV
AV   RRV   AS   SRH

C

Gene	logFC
ADH4	−2.8
MRP2	−2.7
CLB6	−2.4
CYB5	−2.2
TIF2	−2.2
YOL014W	−2.1
MED1	−2.1

Gene	logFC
PET9	3.2
YER138W–A	3.1
RCK1	2.8
RIP1	2.7
SPO74	2.5
HXT8	2.5
YAR029W	2.4
COS1	2.4
YBR296C–A	2.4
REC104	2.3

DC7

AVA

OR   SNC   CRN   SMV
AV   RRV   AS   SRH

D

Gene	logFC
YKR032W	−3.3
DDR2	−3.0
HSP30	−2.9
YBR201C–A	−2.9
YMR272W–B	−2.9
SPO74	−2.7
YIL054W	−2.7
YDL218W	−2.6
YER053C–A	−2.6
YEL008W	−2.6

DC8

No_sig_genes

AVA

OR   SNC   CRN   SMV
AV   RRV   AS   SRH

Figure S8: Genes induced along each diffusion components that separate fermentations within a stage of fermentation in the 2019 vintage. Sites on the diffusion map are colored by American Viticultural Area (AVA) of each sample. Full gene lists are provided in Supplemental Table 1.

Figure S9: Gene set enrichment for differentially expressed genes along DC5 in the 2019 vintage. All enriched categories with p < 0.05 after Bonferroni correction are shown. Pathways on the right side of the figure are induced in samples with a high diffusion component value, while pathways on the left of the figure are induced in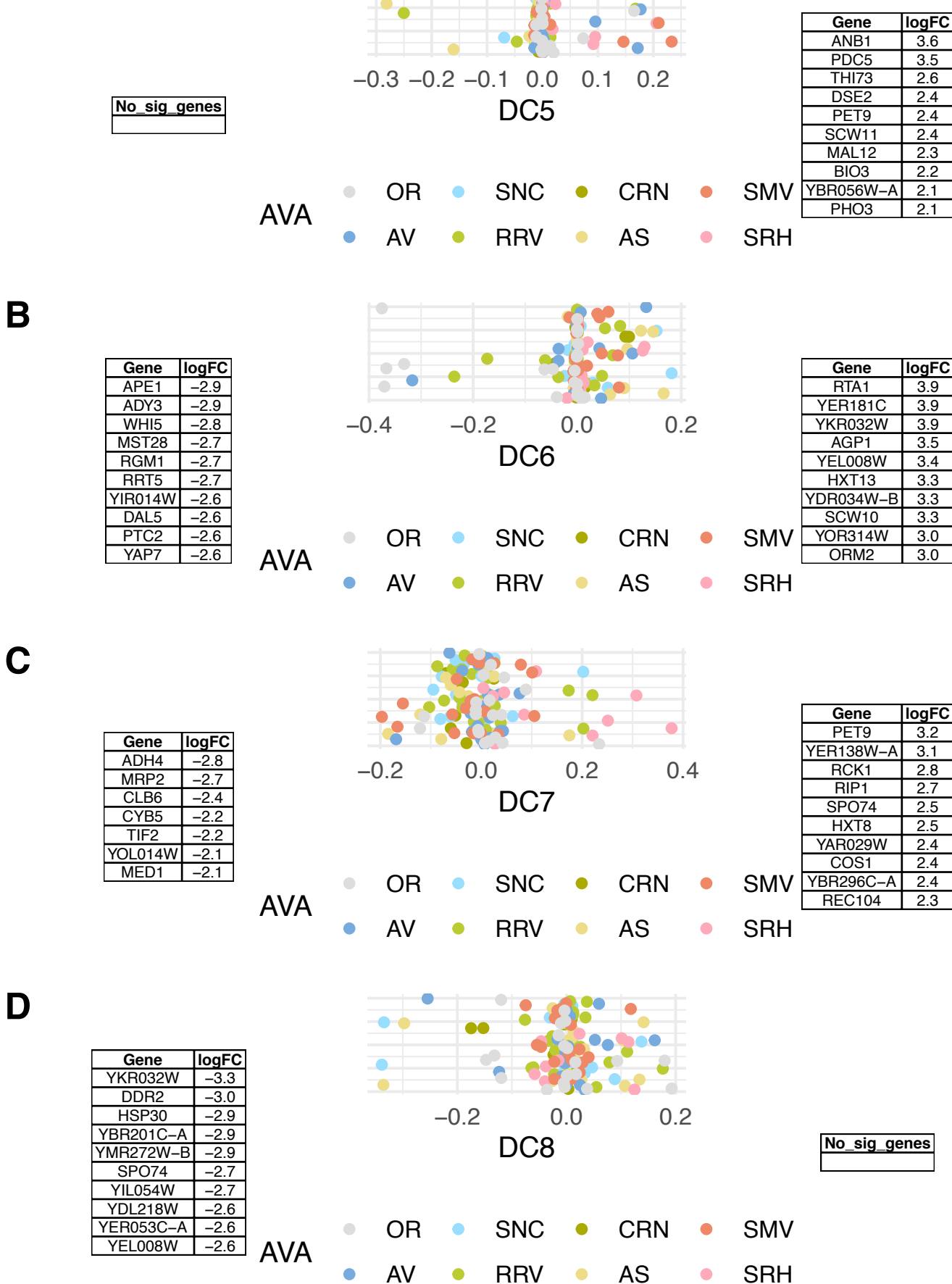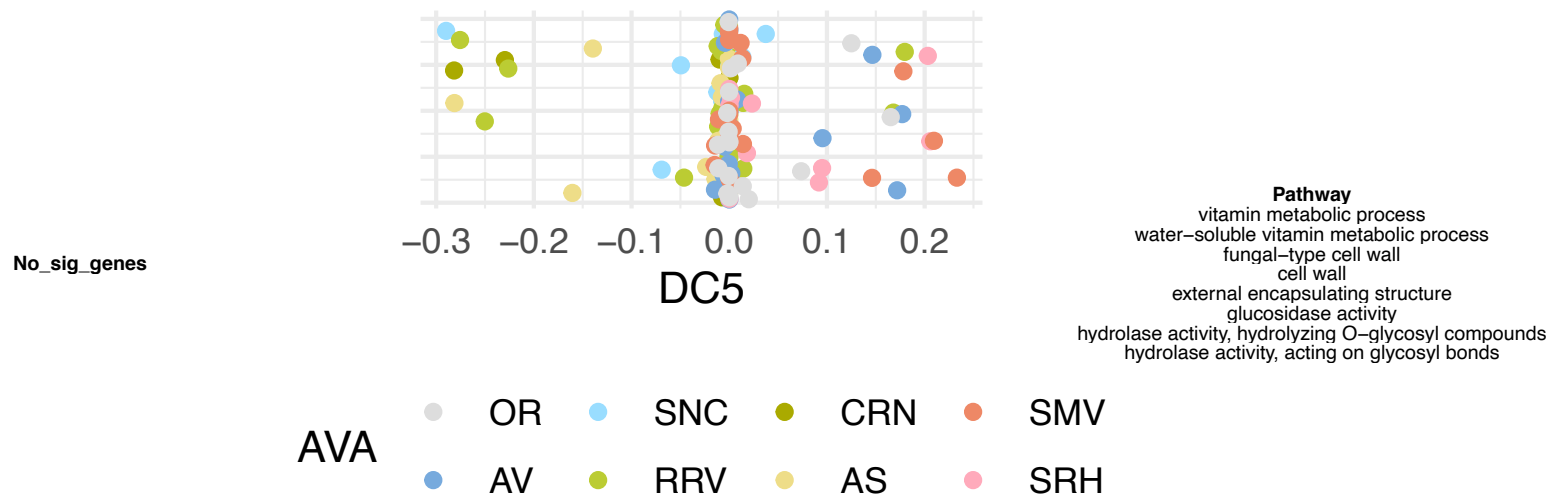 samples with a low diffusion component value.