Transcriptomics provides a genetic signature of vineyard site with insight into vintage-independent regional wine characteristics

Taylor Reiter [1,2,3], Rachel Montpetit [2], Shelby Byer [2], Isadora Frias [2], Esmeralda Leon [4], Robert Viano [4], Michael Mcloughlin [4], Thomas Halligan [4], Desmon Hernandez [4], Rosa Figueroa-Balderas [2], Dario Cantu [2], Kerri Steenwerth [5], Ron Runnebaum [2,4], and Ben Montpetit [1,2,#].

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

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

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

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

[5] Crops Pathology and Genetics Research Unit, USDA Agricultural Research Service, Davis, CA, 95616, USA

Corresponding author: Ben Montpetit, email benmontpetit@ucdavis.edu.

Abstract

In wine fermentations, the metabolic activity of both *Saccharomyces cerevisiae* and non-*Saccharomyces* organisms impact wine chemistry. Ribosomal DNA amplicon sequencing of grape musts has demonstrated that microorganisms occur non-randomly and are associated with the vineyard of origin, suggesting a role for the vineyard, grape, and wine microbiome in shaping wine fermentation outcomes. We used ribosomal DNA amplicon sequencing of grape must and RNA sequencing of primary fermentations to profile fermentations from 15 vineyards in California and Oregon across two vintages. We find that the relative abundance of fungal organisms detected by ribosomal DNA amplicon sequencing did not correlate with transcript abundance from those organisms within the RNA sequencing data, suggesting that the majority of the fungi detected in must by ribosomal DNA amplicon sequencing are not active during these inoculated fermentations. Additionally, we detect genetic signatures of vineyard site and region during fermentation that are predictive for each vineyard site, identifying nitrogen, sulfur, and thiamine metabolism as important factors for distinguishing vineyard site and region.

29 **Importance**

30 The wine industry generates billions of dollars of revenue annually, and economic
31 productivity is in part associated with regional distinctiveness of wine sensory attributes.
32 Microorganisms associated with grapes and wineries are influenced by region of origin,
33 and given that some microorganisms play a role in fermentation, it is thought that microbes
34 may contribute to the regional distinctiveness of wine. We show that while the presence of
35 microbial DNA is associated with wine region and vineyard site, the presence of microbial
36 DNA is not associated with gene expression of those microorganisms during fermentation.
37 We further show that detected gene expression signatures associated with wine region and
38 vineyard site provide a means to address differences in fermentations that may drive
39 regional distinctiveness.

40 **Introduction**

41 During vinification, grape musts are transformed to wine through microbial
42 metabolism, including fermentation of grape sugars into alcohols. In both inoculated and
43 spontaneous fermentations, *Saccharomyces cerevisiae* often becomes the dominant
44 fermentative organism due to a milieu of adaptations that support the rapid consumption
45 of sugars and production of ethanol (1). However, complex microbial communities
46 consisting of other eukaryotic microorganisms and bacteria are present, active, and make
47 significant contributions to the wine making process and final product (2–6). Referred to
48 collectively as non-*Saccharomyces* organisms, these microbes often originate from the
49 vineyard or the winery itself (7, 8). In recognition of the important role these microbes
50 have in wine making, select non-*Saccharomyces* yeasts are increasingly inoculated into
51 commercial fermentations to impart beneficial properties (e.g. bio-protection, lower
52 ethanol, or distinct sensory characteristics (9)). Grape must treatment with sulfur dioxide
53 (SO_2) is also commonly used to control microbial populations, including spoilage
54 organisms, but many microorganisms survive SO_2 treatment and contribute to
55 fermentation outcomes (6, 10, 11).

56 The persistence of vineyard and winery derived microorganisms throughout the
57 winemaking process, as well as the potential for these organisms to influence grape berry
58 development prior to harvest, has led to the idea that select microorganisms unique to a
59 region or vineyard may contribute to region-specific wine characteristics (12, 13). In
60 support of a role of microbial biogeography in regional wine characteristics,
61 microorganisms in vineyards, wineries, and grape musts are known to be associated with

62 their region of origin (4, 7, 8, 14–21). Moreover, the abundance of some organisms in grape
63 must correlates with metabolite concentrations in finished wine, further associating
64 microbial biogeography to fermentation outcomes and wine quality (15, 22). Still, relatively
65 little is known about the influence of non-*Saccharomyces* microorganisms on wine
66 fermentation outcomes, but an increasing number of studies are tackling this complex
67 problem (23, 24). Recent studies have documented increased glycerol accumulation and
68 aroma profiles using sequential- or co-inoculation of *S. cerevisiae* with a single non-
69 *Saccharomyces* yeast species under enological conditions (25–34). While outcomes are
70 diverse, which may be expected given the variety of starting must and culture conditions
71 used across studies, many report consistent alterations in wine such as a higher glycerol
72 content from fermentations inoculated with *S. cerevisiae* and *Starmerella bacillaris* (29, 30,
73 34).

74 How these altered fermentation outcomes occur remains a difficult question to
75 address, as a given outcome may be the direct result of metabolism by the non-
76 *Saccharomyces* organism, or the result of the organism altering *S. cerevisiae* metabolism via
77 direct or indirect interactions (35–37). In support of the latter, the presence of non-
78 *Saccharomyces* organisms has been shown to increase the rate and diversity of resource
79 uptake by *S. cerevisiae* in early fermentation (36–38). In controlled steady-state bioreactor
80 fermentations, the presence of *Lachancea thermotolerans* was found to increase the
81 expression of *S. cerevisiae* genes important for iron and copper acquisition (39). Such
82 interactions are not limited to fungi—lactic acid bacteria can induce epigenetic changes
83 (e.g. *[GAR+]* prion) in *S. cerevisiae* that alter glucose metabolism (40–42). Such abilities of
84 non-*Saccharomyces* organisms to impact *S. cerevisiae* metabolism and fermentation

85 outcomes raises the question of whether microbial biogeography of vineyard sites persists
86 in fermentations, thereby influencing wine outcomes in a site-specific manner. In addition,
87 microbial diversity changes as primary fermentation progresses and *S. cerevisiae* becomes
88 dominant (43), suggesting a changing microbial community could feedback to impact
89 fermentation progression in multiple distinct ways. Currently, we know relatively little
90 about these inter-species interactions and how this influences *S. cerevisiae*, which as a field
91 must be addressed if we are to understand how microbial community dynamics impact
92 wine fermentation outcomes.

93 Past inquiries into the microbial communities of grape must and wine related to
94 regional distinctiveness have focused on assaying the presence of specific microbes based
95 on ribosomal DNA amplicon sequencing (4, 8, 14–20, 44). DNA sequencing has the
96 advantages of capturing both metabolically active and inactive organisms, due to the
97 relative stability of the DNA molecule, offering evidence of a rich history of the microbial
98 community prior to sampling. Ribosomal DNA amplicon data further provides a measure of
99 what microbes may be active at the time of sampling or may become active in the future.
100 While microbiome DNA sequencing of grape musts supports regionally distinct microbial
101 signatures, what microbes are metabolically contributing to fermentation outcomes
102 remains largely unknown. This information is critical when considering the possibility that
103 a particular microbe influences a wine fermentation outcome via metabolism or inter-
104 species interactions.

105 One measure of metabolic activity that is relatively accessible and can be applied at
106 scale to address this issue is the measurement of gene expression in both *S. cerevisiae* and

107 other non-*Saccharomyces* organisms. An interrogation of the genes that are "on" at a given
108 time using RNA sequencing provides important information about the activities an
109 organism may be performing. In addition, the RNA molecule assessed by transcriptomics is
110 constantly turned over within cells and is relatively unstable compared to DNA, which we
111 propose makes transcriptomics a good indicator of microbial activity at the time of
112 sampling. For example, early in fermentation *S. cerevisiae* turns on genes required for
113 glucose metabolism and represses expression of genes needed for the metabolism of other
114 carbon sources; a pattern that reverses towards the end of fermentation when glucose is
115 depleted and *S. cerevisiae* must find alternative energy sources (45). These patterns of gene
116 expression are easily observed using transcriptomics (45, 46), which is increasingly being
117 applied to understanding wine fermentation outcomes (36–39, 47).

118 Here, we characterize microbial populations present in Pinot noir musts from
119 California and Oregon in multiple vintages using both ribosomal DNA amplicon data from
120 grape must samples and gene expression data from multiple fermentation timepoints. We
121 demonstrate that genetic signatures (i.e., DNA and RNA profiles) of vineyard site and
122 region are captured by these data, with total precipitation during growing season being one
123 vineyard-associated factor identified to correlate with site-specific genetic signatures.
124 While DNA profiles reliably predict both vineyard site and region, these profiles did not
125 correlate with the RNA profiles of the primary fermentations. This finding suggests other
126 characteristics influence site-specific gene expression signatures more than the grape must
127 microbiome as measured by ribosomal DNA amplicon sequencing. Importantly, a
128 comparison of DNA sequencing and gene expression data indicates that the majority of
129 organisms detected by ribosomal DNA sequencing lack detectable gene expression during

130 the primary fermentation, thus limiting the likelihood that many of these organisms
131 significantly impact fermentation outcomes during the primary stage of fermentation.
132 Finally, using *S. cerevisiae* gene expression patterns and the associated functions of the
133 genes identified, we are able to identify candidate factors that contribute to vineyard
134 specific fermentation outcomes and wine sensory characteristics.

135 **Results and Discussion**

136 To investigate associations between grape must microbial communities and
137 regional distinctiveness of resulting wines, we performed standardized fermentations of
138 Pinot noir grapes from 15 vineyard sites in California and Oregon across multiple vintages
139 (**Figure S1A**). In 2016, 2017, and 2019, we performed four inoculated fermentations per
140 vineyard site using the wine yeast RC212, taking microbiome samples for DNA isolation
141 and ribosomal DNA amplicon sequencing prior to inoculation. In the 2017 and 2019
142 vintages, we further profiled two primary fermentations from each site using RNA
143 sequencing approaches to perform gene expression analyses at multiple fermentation
144 timepoints. We performed all grape processing and temperature-controlled fermentations
145 at the UC Davis Teaching & Research Winery to standardize vinification and minimize
146 contributions from other factors (winery and winemaker) to the microbiome and
147 transcriptome (48–50).

148 **DNA abundance by ribosomal amplicon sequencing is a poor predictor of detectable**
149 **gene expression during fermentation**

150 Using ribosomal DNA amplicon sequencing of bacteria and fungi, we detected 3254
151 distinct bacterial sequences and 2452 distinct fungal sequences in grape must (**Figure 1A**
152 **and 1B**), with a greater mean species diversity per vineyard site for bacteria than for fungi
153 (**Figure S1B**). However, the core microbiome – i.e., the species present in 90% of all grape
154 musts across all vintages with at least 1% abundance – was larger for fungi than bacteria.
155 The core microbiome consisted of 11 bacterial variants classified to nine taxonomic ranks
156 and 19 fungal variants classified to 10 taxonomic ranks. All bacteria in the core microbiome

157 belonged to the phylum *Proteobacteria* and were dominated by the genus *Tatumella.*
158 (**Figure S2**). *Tatumella* has previously been identified as a dominant genera in other red
159 wine fermentations where it correlated with total acid (by titration) in grape must (51),
160 however these associations have not been experimentally validated. Three of the most
161 abundant bacterial sequence variants identify to the acetic acid producing genus
162 *Gluconobacter* (**Figure S2**). *Gluconobacter* is one of three genera of acetic acid bacteria
163 associated with wine spoilage and the only genus we identify among dominant organisms
164 (**Figure S2**) (52). *Gluconobacter* are primarily active in grape must as the wine
165 environment restricts growth of organisms in this genus (52). Fungi in the core
166 microbiome belong to a single phylum, *Ascomycota*, with all fermentations dominated by
167 the genus *Hanseniaspora*, in particular *Hanseniaspora uvarum*. *H. uvarum* cannot complete
168 alcoholic fermentation alone, but participates in and can alter the quality outcomes of wine
169 fermentations (53). We also identified the fungal genus *Botrytis* among dominant
170 organisms (**Figure S2**), although we lacked the ability to resolve whether the particular
171 variant we detected belongs to the spoilage organism *Botrytis cinerea* or another species in
172 the *Botrytis* genus. Through this work, we have extended microbiome must sequencing to
173 include the 2019 vintage, with results largely matching findings from previous vintages
174 across these same vineyard sites (50). The observed microbial community composition
175 was consistent with organisms previously shown to be present at the initial stages of the
176 wine making process (4, 15–17, 51).

177 Ribosomal DNA amplicon sequencing of grape must is expected to capture cells that
178 are metabolically active, inactive, or dead due to the stability of the DNA molecule. In
179 contrast, gene expression profiling via RNA sequencing is expected to be biased towards

180 living cells. Moreover, the identity of the gene transcripts present at the time of sampling
181 further provides information about what metabolic activities the cell may be performing.
182 Using 3´ Tag RNA sequencing (3´ Tag-seq), we profiled eukaryotic organisms during
183 fermentation using samples taken at multiple fermentation timepoints (i.e., 16, 64, and 112
184 hours after inoculation in 2017 and 2019, plus 2 and 6 hours post-inoculation in 2019).
185 While traditional RNA-sequencing produces sequencing reads from an entire transcript, 3´
186 Tag-seq produces one molecule per transcript by sequencing approximately 100 base pairs
187 upstream of the 3´-end of a sequence (54). This sequencing chemistry requires a poly(A)
188 tail, limiting the sequenced fraction of the transcriptome almost entirely to polyadenylated
189 eukaryotic mRNAs.

190 From the resulting 3´ Tag-seq data, we observed that relatively few eukaryotic
191 microbes were detected during these Pinot noir fermentations (**Figure 1C**). Considering all
192 15 sites together, only 18 eukaryotic species were detected. Further reflecting this finding,
193 *S. cerevisiae* transcripts accounted for the majority of sequences across all fermentations at
194 all time points. To assess whether non-inoculated *S. cerevisiae* strains were responsible for
195 some fraction of sequence reads, we compared each transcriptome against all annotated *S.*
196 *cerevisiae* genomes in GenBank, as well as a genome assembly of *S. cerevisiae* RC212. While
197 non-RC212 *S. cerevisiae* strains were detectable in every fermentation, this fraction
198 accounted for less than 1% of uniquely identifiable sequences. This demonstrates that the
199 inoculated RC212 strain dominated fermentations at all sampled time points. Interestingly,
200 we also identified *Vitis vinifera* transcripts in all samples (**Figure 1C**). The presence of *V.*
201 *vinifera* transcripts suggests intact grape cells persist throughout fermentation.

202 In comparing specific organisms detected via DNA sequencing and 3´ Tag-seq RNA
203 sequencing, we see that only four (*Aureobasidium pullulans*, *Hanseniaspora uvarum*,
204 *Hanseniaspora vineae*, and *S. cerevisiae*) of 397 distinct fungal species definitively identified
205 by ribosomal DNA profiling were detected using gene expression data. This was unchanged
206 in the 2019 transcriptome profiling samples taken at 2 and 6 hours after inoculation,
207 suggesting that organisms detected by amplicon sequencing had lost activity prior to or
208 concurrent with inoculation, well before *S. cerevisiae* would begin to produce inhibitory
209 concentrations of ethanol. Of the four detected organisms by 3´ Tag-seq, ribosomal DNA
210 amplicon sequencing data indicated that *H. uvarum* was most abundant in musts prior to
211 inoculation and was detected in every vineyard site (**Figure 2A**). Still, the relative
212 abundance of *H. uvarum* in grape must from ribosomal DNA amplicon sequencing was only
213 weakly correlated with relative abundance of RNA during fermentation ($R^2 = 0.14$, $p <$
214 0.01). Importantly, while these values are weakly correlated, *H. uvarum* had almost no
215 detectable gene expression in fermentations from many sites where it dominated the DNA
216 profile of the grape must (**Figure 2B**). Finally, even when we performed this analysis using
217 samples from the first hours of fermentation after inoculation, relative abundance of *H.*
218 *uvarum* DNA in grape musts remained weakly correlated with relative abundance of RNA
219 (two hours: $R^2 = 0.21$, $p < 0.05$, six hours: $R^2 = 0.28$, $p < 0.01$). In the case of *A. pullulans,*
220 DNA in grape must is not correlated with gene expression during fermentation (two hours:
221 $R^2 = -0.03$, $p = 0.60$, six hours: $R^2 = -0.025$, $p = 0.53$). These results indicate that most
222 identified eukaryotic microorganisms in grape must by DNA profiling likely have little
223 metabolic activity in inoculated fermentations even when the organisms are detected at
224 high abundance and are detectable via both sequencing methods.

225 Given these findings, it is important to consider if a lack of detectable gene
226 expression for non-*Saccharomyces* fungal species could be reflective of some other issue
227 that is technical or biological in nature. We consider this highly unlikely for two reasons.
228 First, both DNA and RNA sequencing require similar protocols for extraction of nucleic
229 acids from cells that should perform approximately equally across samples. Second, RNA
230 sequencing relies on a highly conserved biological processes (mRNA polyadenylation),
231 hence while we could envision RNA sequencing failing for one or a few organisms, it should
232 not fail across many fermentations for the large majority of organisms seen in this work.
233 Moreover, of the 16 non-*Saccharomyces* fungi detected via RNA-sequencing, eight of these
234 organisms were not detected at the genus level by DNA profiling (*Cladosporium sp SL-16*,
235 *Lachancea thermotolerans*, *Metschnikowia fructicola*, *Metschnikowia sp. AWRI3582*, *Pichia*
236 *kudriavzevii*, *Preussia sp. BSL10*, *Rhizopus stolonifer*, *Starmerella bacillaris*). This suggests
237 that transcriptomic profiling is a sensitive assay able to detect organisms present in a
238 population that are missed by ribosomal DNA amplicon sequencing, which is likely due to
239 an inability to resolve genus or species using ribosomal DNA sequences.

240 Notably, some of the organisms detected by RNA sequencing have the ability to
241 influence fermentation outcomes: in mixed fermentations with *S. cerevisiae*, *S. bacillaris* has
242 been shown to lower the final ethanol concentration and increase the concentration of
243 glycerol (55), while *M. fructicola* increased the concentration of esters and terpenes (56).
244 Therefore, the detection of these organisms by RNA sequencing provides valuable
245 information with respect to the potentially active microbial population in these
246 fermentations. Our findings align well with another recent report that showed an RNA-
247 based sequencing strategy is a highly sensitive alternative to amplicon sequencing (57). As

248 such, it may be appropriate to use RNA sequencing as a general method to capture the
249 metabolically active microbial community during wine fermentation, especially when
250 drawing a connection between the wine microbiome and fermentation outcomes.

251 **Genetic signatures differentiate vineyard site, region, and vintage**

252 The region and site from which grapes are harvested can have an important
253 influence over the character of a resulting wine based by on a variety of factors (e.g.,
254 climate, soil type, grape associated microbes). As such, we considered if the data generated
255 using DNA and RNA sequencing strategies during these Pinot noir fermentations is
256 reflective of vineyard site through the generation of unique genetic signatures. To
257 investigate this concept, we grouped DNA and RNA sequencing samples by vineyard site,
258 region, and vintage to see if there were detectable differences among these groups. Using
259 analysis of similarities (ANOSIM; see methods), we determined that all three factors
260 explain differences among groups of samples, with vineyard site explaining the most group
261 similarity (**Figure 3A-D**). This supports the idea that fermentations have a detectable
262 genetic signature that is reflective of vineyard site.

263 To understand which specific organisms and genes contribute to the genetic
264 signatures of both vineyard site and region, we built machine learning classification models
265 using random forests. These models weight the contribution of each feature to predictive
266 accuracy of the model, enabling robust identification of specific genes or organisms that
267 differentiate vineyard sites or regions among fermentations. When we used data from all
268 vintages in model training and testing to predict region, we achieved 87%-95% accuracy
269 (**Table S1-S3**; **Figure S3-S4**). When we instead used data from one vintage in model

270 training and testing to predict region, accuracy dropped across all models, but ranged from
271 57%-75% (**Table S1-S3**; **Figure S3-S4**). This suggests that models built with fermentations
272 from all vintages better capture cross-vintage similarities as these models select predictive
273 variables that are consistent across the vintages studied. However, the accuracy of these
274 models may decrease if the same set of predictive variables is not consistent in future
275 vintages. Conversely, the accuracy of a model built from a single vintage and trained on a
276 separate vintage will likely remain consistent across many vintages. From this, we assumed
277 that models trained using data from a single vintage better reflected model accuracy, but
278 that models trained using data from all vintages better reflected cross-vintage similarities.
279 As we aimed to identify vintage-independent factors, we analyzed cross-vintage models
280 moving forward.

281 When we used the same data to generate vineyard-specific models, predictive
282 accuracy was on average 21.4% less than region-specific models (**Table S1**). However, it is
283 important to note that this decrease in accuracy was driven by within-region
284 misclassification for vineyards in Willamette Valley (31 km separation), Santa Maria Valley
285 (5 km separation), and Arroyo Seco (1 km separation) American Viticultural Areas (AVA)
286 (**Figure S5**). The same misclassifications persisted across many models, highlighting
287 potential within-region similarity that contributes to genetic signatures, which fits well
288 with the concept of AVA and region-associated wine characteristics.

289 Across models, we were surprised to find that bacterial models outperformed or
290 performed as well as fungal models for classification of site and region, as bacterial must
291 samples added the least predictive power in previous models for region prediction (14),

292 including for Pinot noir grapes grown in Australia (8). Bacterial must samples have been
293 shown to be predictive of region in Californian Chardonnay, but not Californian Cabernet
294 Sauvignon (14), suggesting a possible cultivar-specific effect. In previous inquiries, samples
295 were processed in vineyard-specific wineries, providing another variable that could
296 potentially alter the measured microbiomes and the contributions attributed to bacteria
297 and fungi.

298 Given that random forests models estimate the importance of each gene in
299 determining vineyard or region classification, we further used the gene expression models
300 to gain insight into biological differences between vineyard sites and regions. For this, we
301 calculated the percent of total importance attributable to each gene from each eukaryotic
302 organism detected (**Table S2**). Vineyard-specific models weighted non-*Saccharomyces*
303 yeast genes as a whole as most important for predictive accuracy (**Figure 3E, Figure S6**).
304 In particular, genes from *S. bacillaris*, *M. fructicola*, *Metschnikowia sp. AWRI3582*, and *L.*
305 *thermotolerans* were important for vineyard site classification. The ability of non-
306 *Saccharomyces* gene expression to distinguish site is likely related to the unique
307 combination of non-*Saccharomyces* organisms present in each fermentation, which results
308 in these organisms having strong predictive power when detected. In contrast, regional
309 models weighted *S. cerevisiae* and *V. vinifera* genes as higher importance (**Figure 3E,**
310 **Figure S6**). We expect that this may result from changes in *V. vinifera* gene expression
311 across more diverse geographical environments, which leads to differences in the grape
312 must and associated fermentations as detected by *S. cerevisiae* gene expression.

313 To more directly address how environmental factors and grape must chemistry
314 correlate with genetic signatures, we correlated initial must chemical parameters (pH,
315 titratable acidity, malic acid, NOPA, and NH_3) and vineyard site characteristics (total
316 precipitation, growing degree days, and geographic distance between sites) with DNA and
317 RNA profiles using the Mantel test (see methods). From these analyses, we found
318 geographic distance between vineyards correlated with precipitation and growing degree
319 days, indicating that sites that are geographically closer experience more similar weather
320 patterns, as would be expected (**Figure 3F**). Amongst the factors tested, only precipitation
321 correlated with all genetic profiles (**Figure 3G**). Similar to geographic distance, initial
322 chemical profiles of vineyard sites were more similar when sites are geographically closer.
323 However, we found surprisingly few correlates between genetic profiles and initial grape
324 must conditions (**Figure 3G**). While fungal profiles correlate with initial malic acid, NOPA,
325 and NH_3 and bacterial profiles correlate with initial NOPA, gene expression profiles only
326 correlate with initial malic acid levels. The finding that gene expression profiles do not
327 correlate with initial nitrogen concentration, even though nitrogen availability is central to
328 yeast growth and linked to the expression of hundreds of genes (45), may reflect nitrogen
329 additions at ~24 h after inoculation during winemaking so that all fermentations had a
330 minimum of 250 mg/L (see methods). Overall, the poor correlation between gene
331 expression patterns and the factors tested suggest that other unmeasured factors drive
332 gene expression distinctiveness in these fermentations. This raises a clear need for future
333 work that measures many factors within vineyards and fermentations to define the
334 organism-environment interactions responsible for driving gene expression and cellular
335 activities of *S. cerevisiae* and other microbial organisms.

336 ***S. cerevisiae* gene expression provides insight into vineyard site and region features**

337 *S. cerevisiae* is likely the best understood eukaryote based on the use of this
338 organism as a model system for biology, which has provided a rich set of genomic
339 resources and databases (58). As such, *S. cerevisiae* gene expression can be used as a
340 biosensor to provide insight into the fermentation environment based on activities yeast
341 perform. The utility of this data is furthered by the fact that *S. cerevisiae* gene functions are
342 well studied in the context of wine production, *S. cerevisiae* is ubiquitous across all
343 fermentations, and the transcriptomics data is dominated by reads from *S. cerevisiae* (e.g.,
344 data completeness). Consequently, given the data above suggesting unknown factors are
345 directing fermentation outcomes, we queried the *S. cerevisiae* gene expression data to
346 assess what genes were important for predicting region and vineyard site to infer what
347 may be unique about musts produced by grapes from each vineyard site or region. Notably,
348 random forests models are non-deterministic, meaning the each time a model is built the
349 specific genes important for predictive accuracy of that model may change, especially for
350 genes with correlated gene expression values (59). Therefore, we first built 100 random
351 forests models for the prediction of region and vineyard site and investigated the genes
352 that were shared across the majority models (**Table S4**). As discussed above, less than 1%
353 of transcripts in any fermentation were expressed by non-RC212 *S. cerevisiae* and thus the
354 genetic signatures we identified are likely specific to this strain.

355 From this analysis, important predictors of both site and region included flavor-
356 associated genes involved in the formation of higher alcohols and volatile fatty acids
357 through the Ehrlich pathway. Each site-specific and region-specific model included an

358 average of 16 (site SD = 2.9, region SD = 2.4) genes associated with flavor development in
359 wine (**Table S5**). These genes were mostly associated with the Ehrlich pathway (site mean
360 = 8.1 genes, SD = 2; region mean = 9 genes, SD = 1.7) and with volatile sulfur formation (site
361 mean = 6.3 genes, SD = 1.6; region mean = 5.1 genes, SD = 1.4). Given that genes in these
362 pathways were detectable as indicators of both region and site, site-variable expression of
363 these genes could contribute to region- and vineyard-specific wine flavor profiles detected
364 in wines from these vineyards in previous vintages (48). At this time, it remains unknown
365 what factors cause these flavor-associated genes to differ between fermentations.

366 In addition to flavor-associated genes, many *S. cerevisiae* genes that were important
367 for predicting vineyard site and region are members of the Com2 regulon (**Table S4**).
368 Expression of genes within the Com2 regulon is protective against SO_2 stress (60). We
369 treated all fermentations with an equal dose of SO_2 at the beginning of vinification;
370 however, variable application of sulfur-containing fungicides in the vineyard may lead to
371 disparate SO_2 stress during fermentation underlying the signatures of site and region that
372 we observe. Wine strains of *S. cerevisiae* are more tolerant of SO_2 than many non-
373 *Saccharomyces* species, but SO_2 exposure can cause inhibition of key metabolic enzymes
374 like alcohol dehydrogenase, as well as other processes through cleavage of disulfide bonds
375 (61, 62). Of the 511 genes dependent on Com2 activation during SO_2 stress (60), an average
376 of 105 genes (SD = 12.7) were important for differentiating site in our predictive models,
377 while 101 genes (SD = 11.6) were important for predicting vineyard region. Within these
378 gene lists are genes involved in the efflux of sulfite and bisulfite, sulfate assimilation, sulfate
379 assimilation, biosynthesis of methionine, cysteine, arginine, and lysine, and biosynthesis of
380 the sulfur-containing vitamin biotin (**Table S6**)**.** These pathways, and their site-specific

381 signatures, are potential areas of future study given that sulfur metabolism can have a
382 profound impact on the sensory attributes of wine (63). In addition, while the molecular
383 form of SO_2 causes *S. cerevisiae* stress and inactivation of wine spoilage microbes (11, 60),
384 this form is in equilibria with the bisulfite form (HSO_3^-) and this ratio is dependent on wine
385 pH (64). The bisulfite form interacts with anthocyanins and can cause color bleaching (64).
386 This suggests that the SO_2 stress response is a factor that would need to be considered in
387 the context of pH and other aspects of SO_2 wine chemistry.

388 To further explore connections between *S. cerevisiae* gene expression and region or
389 vineyard site, we identified genes that were predictive for a specific region or site across all
390 models (local importance, see methods). Only one gene was important across all models for
391 predicting the site OR1 (VIT_0003506001; *V. vinifera* pathogenesis-related protein 10.3).
392 This suggests that we have limited resolution into the specific gene expression patterns
393 that differentiate individual sites using this method. Given that gene expression is
394 inherently noisy (65), increasing observations per vineyard site may improve accuracy and
395 inference from site-specific models in the future.

396 In contrast to site-specific models, an average of 22.4 genes per region (SD = 13.5)
397 were predictive across all models, with an average of 14.4 genes (SD = 8.4) expressed by *S.*
398 *cerevisiae* (**Table S7**). Interestingly, many genes that were important for predicting one
399 region were also important for predicting other regions (*BET2, BET3, BIO4, EXG2, FAS2,*
400 *HEM12, LOH1, MEP3, MRX21, NPT1, PSA1, SNZ3, THI11, THI13, THI72, TUB4*), suggesting
401 that expression of these genes differed consistently between regions. These genes encode
402 proteins involved in diverse cellular processes, including heme biosynthesis, cell wall

403 assembly, and synthesis and transport of fatty acids and nitrogen-containing compounds.
404 While the underlying biochemical processes that lead to consistent expression of these
405 genes within regions remains unknown, we investigated whether initial nitrogen content in
406 grape must was related to the importance of *MEP3*, a gene that encodes an ammonia
407 permease, in predicting a region. Interestingly, *MEP3* was important for predicting the
408 three regions with the lowest average initial yeast assimilable nitrogen (OR, AV, RRV) as
409 well as the region with the second highest yeast assimilable nitrogen (SMV) across vintages
410 (**Figure S7**). Given that nitrogen availability plays a fundamental role in shaping
411 fermentations (66), this relationship was expected. We also noted that four genes
412 associated with thiamine availability were important for predicting multiple regions. This
413 suggests that thiamine availability may drive regional differences in wine outcomes, a
414 postulate that could be measured in a future vintage.

415 Taken together, these results identify genes directly linked to wine sensory and
416 chemistry that are strong indicators of vineyard region in Pinot noir fermentations. These
417 findings provide a concrete starting point for future investigation into vineyard specific
418 factors that are responsible for wine fermentation outcomes and wine sensory
419 characteristics.

420

421 **Conclusion**

422 Microbial biogeography of wine has been documented in globally distributed
423 appellations (4, 7, 8, 14–21), and has been correlated with wine fermentation outcomes
424 (15, 22). In inoculated co-cultures, non-*Saccharomyces* microorganisms both contribute to
425 fermentation and change the behavior of the dominant fermenter *S. cerevisiae*, leading to
426 measurable differences in wine aroma and composition (36–38). Here, we show that grape
427 must ribosomal DNA profiles do not correlate with detected eukaryotic gene expression
428 patterns during primary fermentation. Given that we detected little to no correlation
429 between fungal profiles in initial grape must and genes expressed by those organisms
430 during primary fermentation, DNA profiles may not be a robust indicator for inferring
431 contributions from these organisms in wine sensory outcomes in inoculated fermentations.
432 However, DNA profiles, in particular bacterial profiles, are predictive of vineyard site and
433 retain signatures of site-specific processes such as total precipitation during the growing
434 season. These profiles are rich indicators of the patterns that shape the microbial ecology
435 of grapes, and reflect differences among vineyard sites and regions, even when the same
436 clone (e.g., *Vitis vinifera* L. cv. Pinot noir clone 667) is grown on each site.

437 In contrast, the gene expression profiles of *S. cerevisiae* and other detected
438 organisms, retain signatures of vineyard site and region as well as the metabolic
439 transformations that occur during fermentation. Using *S. cerevisiae* gene expression as a
440 biosensor for differences between fermentations, we detected site and region specific
441 signatures linked to nitrogen, sulfur, and thiamine metabolism. While these factors are
442 associated with vineyard-specific differences in gene expression profiles, few vineyard site

443 and initial grape must chemical parameters correlate with the transcriptome, which
444 suggests there are still many variables to discover that underlie the complex metabolic
445 activities and gene expression patterns *S. cerevisiae* displays throughout fermentation. In
446 the future, more comprehensive sequencing approaches (e.g., deeper sequencing with
447 methods that capture the full transcriptome, more samples per site) aimed at the factors
448 and organisms identified in this work would allow for a better understanding of these
449 systems. This will need to be accompanied by measurements of many more vineyard, must,
450 and wine characteristics to provide further predictive power and insights into the
451 complexities and subtleties of vineyard specific wine fermentation outcomes.

452 **Methods**

453 **Grape preparation and fermentation**

454 The winemaking protocol has been described previously (48, 49), but the relevant
455 parts are reproduced with some added details below. The grapes used in this study
456 originated from 15 vineyards in eight American Viticultural Areas in California and Oregon,
457 U.S.A. All grapes were Vitis vinifera L. cv. Pinot noir clone 667, with either rootstock 101-14
458 (AV1, RRV1, SNC1, SNC2, CRN1, AS1, AS2, SMV1, SMV2, SRH1), Riparia Gloire (OR1, OR2),
459 or 3309C (AV2, RRV2, RRV3). Grapes were hand-harvested grapes at approximately 24
460 Brix and transported to the University of California, Davis Teaching & Research Winery for
461 fermentation. Grapes were separated into half-ton macrobins on harvest day and Inodose
462 SO_2 was added to 40 ppm. Upon delivery to the winery, bins were stored at 14°C until the
463 fruit was destemmed and divided into temperature jacket-controlled tanks. N_2 sparging of
464 the tank headspace was performed prior to fermentation and tanks sealed with a rubber
465 gasket. We cold soaked the must at 7°C for three days and adjusted TSO_2 to 40 ppm on the
466 second day. After three days, the must temperature was increased to 21°C and
467 programmed pump overs were used to hold the tank at a constant temperature. Grape
468 must microbiome samples were taken just prior to the increase in temperature. For
469 inoculation, *S. cerevisiae* RC212 was rehydrated with Superstart Rouge at 20 g/hL and
470 inoculated in the must at 25 g/hL. At approximately 24 hours after inoculation, nitrogen
471 content in the fermentations was adjusted using DAP (target YAN – 35 mg/L – initial
472 YAN)/2), and Nutristart (25 g/hL). Nitrogen was adjusted only if YAN was below 250 mg/L.
473 Approximately 48 hours after fermentation, fermentation temperatures were permitted to

474 increase to 27°C, and again added DAP using the formula (target YAN - 35 mg/L - initial
475 YAN)/2, and fermentation were then continued until Brix < 0. Fermentation samples were
476 taken for Brix measurements every twelve hours relative to inoculation and with RNA
477 samples at 2 hours, 6 hours (2019 vintage), 16 hours, 64 hours, and 112 hours (2017 and
478 2019 vintage). To ensure uniform sampling, a pumpover was performed ten minutes prior
479 to sampling each tank. For RNA samples, 12mL of juice was obtained, centrifuged at 4000
480 RPM for 5 minutes, supernatant was discarded, and the cell pellet snap frozen in liquid
481 nitrogen. Samples were stored at -80°C until RNA extraction.

482 **Amplicon sequencing data processing**

483 DNA was extracted for amplicon sequencing and library preparation following (50)
484 and (67). The UC Davis DNA Tech Core performed sequencing using Illumina MiSeq,
485 producing 251 base pair paired-end sequences. We demultiplexed and adapter trimmed
486 libraries by barcode sequences using cutadapt (68). Taxonomically annotated amplicon
487 sequence variant (ASV) counts were generated using DADA2 with the Silva NR database
488 (version 138) for 16S sequences and the UNITE general FASTA release (version 8.2) for ITS
489 sequences (69). All ASVs annotated as "Bacteria,Cyanobacteria,Cyanobacteriia,Chloroplast"
490 and "Bacteria,Proteobacteria,Alphaproteobacteria,Rickettsiales,Mitochondria" were
491 removed as these represent plant mitochondria and chloroplast 16S sequences and not
492 bacterial sequences.

493 **RNA sequencing data processing**

494 Yeast pellets were thawed on ice, resuspended in 5ml Nanopure water, centrifuged
495 at 2000g for 5min, and aspirated the supernatant. RNA was extracted using the Quick RNA

496 Fungal/Bacterial Miniprep kit including DNAsel column treatment (cat#R2014, Zymo
497 Research). RNA was eluted in 30μL of molecular grade water and assessed for
498 concentration and quality via Nanodrop and RNA gel electrophoresis. Sample
499 concentrations were adjusted to 200ng/μl and used for sequencing. 3´ Tag-seq single-end
500 sequencing (Lexogen QuantSeq) was applied in both the 2017 and 2019 vintage, with the
501 addition of UMI barcodes in 2019. The University of California, Davis DNA Technologies
502 Core performed all library preparation and sequencing.

503 The first 12 base pairs from each read were hard trimmed and Illumina TruSeq
504 adapters and poly(A) tails were removed. Sourmash gather was used to determine the
505 organisms present in each sample using parameters -k 31 and --scaled 2000 (70, 71). The
506 GenBank microbial database (https://sourmash-databases.s3-us-west-
507 2.amazonaws.com/zip/genbank-k31.sbt.zip) and eukaryotic RNA database
508 (https://osf.io/qk5th/) was used for these queries.

509 Using results from sourmash, a set of reference genomes was constructed that was
510 representative of all organisms detected within the samples. When different strains of the
511 same species were detected, the one species detected in the largest number of samples was
512 used as a representative species to reduce multi-mapping conflicts. Species present in more
513 than two samples were included because species present in fewer than three samples
514 would have limited predictive power. Species of genus *Saccharomyces* other than *S.*
515 *cerevisiae* S288C were removed to reduce multi-mapping conflicts. Selected genomes were
516 downloaded from NCBI GenBank; however, if no GTF annotation file was available for the
517 species, the genome and GFF3 file was taken from JGI Mycocosm (72), and the GFF3 was

518 converted to GTF using the R package rtracklayer (73). When no annotation file was
519 available on GenBank or JGI Mycocosm, the genome of the closest species-level strain with
520 a GTF annotation file was used. To find closely related organisms, NCBI taxonomy was
521 searched, selected assemblies were downloaded , and sourmash compare was used with a
522 k-size of 31 (70, 71). The organisms with the highest Jaccard similarity were considered the
523 most similar. When no annotation file was available for similar organisms, an annotation
524 file was generated using WebAugustus (74). See **Table S8** for a description of the best
525 matched genome, the genome used for count generation, and the source of genome
526 annotations. Reference genome FASTA files and GTF files were concatenated together to
527 generate a single reference. STAR was then used to align reads against the constructed
528 reference with parameters --outFilterType BySJout, --outFilterMultimapNmax 20, --
529 alignSJoverhangMin 8, --alignSJDBoverhangMin 1, --outFilterMismatchNmax 999, --
530 outFilterMismatchNoverLmax 0.6, --alignIntronMin 20, --alignIntronMax 1000000, --
531 alignMatesGapMax 1000000, --outSAMattributes NH HI NM MD --outSAMtype BAM,
532 SortedByCoordinate (75). For the 2019 vintage, UMI tools was used to deduplicate
533 alignments (76). The number of reads mapping to each gene was quantified using htseq
534 count using the constructed reference GTF file to delineate gene regions (77).

535 **RC212 genome assembly and comparison**

536 The *S. cerevisiae* RC212 genome was assembled to estimate the fraction of RNA-
537 sequencing reads in each fermentation originating from non-RC212 *S. cerevisiae* strains.
538 FASTQ files for accession SRR2967888 were downloaded from the European Nucleotide
539 Archive (78). Reads were k-mer trimmed using the khmer trim-low-abund.py command with

540 parameter -k 20 (79) and the Megahit assembler was used with default parameters to
541 assemble reads (80).

542 **Estimation of non-inoculated yeast in RNA-seq samples**

543 Sourmash gather was used to estimate the fraction of RNA seq reads (k-mers)
544 originating from non-inoculated *S. cerevisiae*. Sourmash gather estimates shared sequence
545 similarity by comparing scaled MinHash signatures derived from k-mer profiles (70, 71).
546 The sourmash Eukaryotic RNA database (https://osf.io/qk5th/) was used, which includes
547 all annotated *S. cerevisiae* genomes in GenBank (e.g., genomes that include
548 *rna_from_genome.fna annotations), as well as our *S. cerevisiae* RC212 genomes assembly.

549 **Correlation between ribosomal DNA amplicon sequencing data and 3´ Tag-seq data**
550 **for non-*Saccharomyces* organisms**

551 Fermentations with fungal ITS amplicon sequencing data and 3´ Tag-seq were
552 compared. First, ribosomal DNA amplicon sequencing read counts from *H. uvarum* were
553 regressed against total 3´ Tag-seq counts from *H. uvarum* using counts from 16 hours, 64
554 hours, and 112 hours of fermentation. 3´ Tag-seq counts were derived from STAR and
555 htseq (see RNA sequencing data processing above). Counts were transformed into
556 compositional counts (relative abundance) prior to linear regression (81). Linear
557 regression was performed using the lm() function in R. This analysis was performed again
558 separately for *H. uvarum* and *A. pullulans* using counts from the 2 hour and 6 hours samples
559 taken in the 2019 vintage. Given that this analysis relied on reads aligned to annotated 3´
560 regions, a separate regression was performed a using proportion of reads assigned to a
561 given organism derived from sourmash gather (see RNA sequencing data processing

562 above). Only results from the first analysis were reported as R^2 values were within 0.01
563 between both analyses.

564 **ANOSIM and NMDS**

565 Compositional data analysis was used for amplicon and transcriptome counts (81).
566 The transform() function in the microbiome bioconductor package was used to transform
567 counts by centered log ratio (82, 83). To test for differences among groups, Aitchinson
568 distance (Euclidean distance on CLR-transformed counts) was used and tested with the
569 anosim() function in the vegan package using parameters distance = "euclidean" and
570 permutations = 9999 (84, 85). A cut off of p = 0.05 was used for statistical significance. To
571 construct NMDS plots, Aitchinson distance was taken using the metaMDS() function in the
572 vegan package with parameter distance = "euclidean". Results were plotted using the ggplot2
573 package (86).

574 **Amplicon sequencing random forest models**

575 Random forest classifiers were built using the R ranger package (87). Using ASV
576 counts produced by DADA2, counts were normalized by dividing by total number of aligned
577 reads. The tuneRanger() function was used in the tuneRanger package to optimize each
578 model for parameters m.try, sample.fraction, and min.node.size (88). The ranger() function
579 was then used to build each model with parameters from tuneRanger as well as num.trees =
580 10000, importance = "permutation", and local.importance = TRUE. As a supervised technique,
581 random forest classifiers are trained on a subset of data and tested on a separate subset to
582 calculate predictive accuracy. For models built with samples from all vintages, the
583 createDataPartition() function in the R caret package was used to randomly but equally

584 partition training and testing sets with a 70:30 split, ensuring that all class labels were
585 equally represented in both sets (89). For other models, the classifier was built using all
586 samples from two vintages and validated on the held-out vintage. Accuracy and kappa
587 statistics were calculated for each model.

588 **RNA sequencing random forest model**

589 Counts were imported into R and normalized by dividing by total number of aligned
590 reads (e.g., library size). Given that random forests expects independent samples and RNA-
591 sequencing was conducted in time series over the course of primary fermentation, each
592 gene from each time series set was summarized into mean count, minimum count,
593 maximum count, total count, and standard deviation of counts. Variable selection was
594 performed using the vita method (90) and models were built using the same methods as
595 with amplicon sequencing models.

596 To estimate vineyard- and region-specific gene importance, variable selection and
597 model optimization were performed with 100 different seeds. For each model, gene local
598 importance was averaged for each fermentation from a vineyard site or region in the
599 training set and genes with positive average local permutation importance were retained.
600 The intersection of genes from all models was then taken to determine which genes were
601 predictive for a particular site or region in all models. Although random forests were
602 trained on summarized gene attributes, any genes that were predictive across any attribute
603 were retained as these attributes were often highly correlated.

604 **Mantel tests**

605 Mantel tests were performed to assess the similarity between samples across
606 measurements of bacterial abundance, fungal abundance, transcriptome abundance, initial
607 grape must chemistry, and vineyard site parameters (91, 92). The Mantel test determines
608 the correlation between the same samples in different matrices, testing whether
609 similarities between samples estimated from one measurement type match similarities of
610 the same samples estimated from a different measurement type (91, 92). These tests were
611 performed using complete cases, with microbiome and transcriptome abundances from the
612 2017 and 2019 vintages. Vineyard site parameters total precipitation and growing degree
613 days were estimated using the PRISM climate models including dates April 1 - September
614 30 in 2017 and 2019 (93). Distance matrices were calculated for each matrix using the
615 dist() function in R, with parameters method = "euclidean", with the exception of geographic
616 distance which was calculated using the distm() function in the package geosphere with
617 parameter distHaversine (94). When distances for disparate measurement types were
618 calculated at the same time, values were first scaled and centered using the function scale()
619 with parameters center = TRUE and scale = TRUE. Mantel tests were performed with the
620 mantel() function in the vegan package with parameters method = "spearman", permutations =
621 9999, and na.rm = TRUE (85, 92). p value adjustment were applied using the function
622 p.adjust() with parameter method = "fdr" and a false discovery rate of p = 0.1 used.

623 **Data Availability**

624 RNA sequencing data is available in the Sequence Read Archive under accession number
625 PRJNA680606. Microbiome data is available under accession numbers PRJNA642839 and

PRJNA682452. All analysis code is available at

github.com/montpetitlab/Reiter_et_al_2020_SigofSite.

Acknowledgements

We thank all past and current members of the Steenwerth, Runnebaum, and Montpetit laboratories for their support of this work, as well as the students and staff of the UC Davis Pilot Winery. T.R. was supported by the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative [GBMF4551]; Harry Baccigaluppi Fellowship; Horace O Lanza Scholarship; Louis R Gomberg Fellowship; Margrit Mondavi Fellowship; Haskell F Norman Wine & Food Fellowship; Chaîne des Rôtisseurs Scholarship; Carpenter Memorial Fellowship. The authors would like to recognize support from Jackson Family Wines, in addition to support from Lallemand Inc.

639 **References**

640 1. Querol A, Fernández-Espinar MT, lí del Olmo M, Barrio E. 2003. Adaptive evolution of
641 wine yeast. International journal of food microbiology 86:3–10.

642 2. Jolly N, Augustyn O, Pretorius I. 2003. The occurrence of non-Saccharomyces
643 cerevisiae yeast species over three vintages in four vineyards and grape musts from
644 four production regions of the Western Cape, South Africa. South African Journal of
645 Enology and Viticulture 24:35–42.

646 3. Ghosh S, Bagheri B, Morgan HH, Divol B, Setati ME. 2015. Assessment of wine
647 microbial diversity using ARISA and cultivation-based methods. Annals of
648 microbiology 65:1833–1840.

649 4. Wang C, García-Fernández D, Mas A, Esteve-Zarzoso B. 2015. Fungal diversity in grape
650 must and wine fermentation assessed by massive sequencing, quantitative PCR and
651 DGGE. Frontiers in microbiology 6:1156.

652 5. Bagheri B, Bauer F, Setati M. 2015. The diversity and dynamics of indigenous yeast
653 communities in grape must from vineyards employing different agronomic practices
654 and their influence on wine fermentation. South African Journal of Enology and
655 Viticulture 36:243–251.

656 6. Bagheri B, Bauer FF, Cardinali G, Setati ME. 2020. ecological interactions are a primary
657 driver of population dynamics in wine yeast microbiota during fermentation. Scientific
658 reports 10:1–12.

659 7. Abdo H, Catacchio CR, Ventura M, D'Addabbo P, Alexandre H, Guilloux-Bénatier M,
660 Rousseaux S. 2020. the establishment of a fungal consortium in a new winery.
661 Scientific reports 10:1–12.

662 8. Liu D, Chen Q, Zhang P, Chen D, Howell KS. 2020. The Fungal Microbiome Is an
663 Important Component of Vineyard Ecosystems and Correlates with Regional
664 Distinctiveness of Wine. mSphere 5.

665 9. Roudil L, Russo P, Berbegal C, Albertin W, Spano G, Capozzi V. 2020. Non-
666 Saccharomyces Commercial starter cultures: scientific trends, recent patents and
667 innovation in the wine sector. Recent patents on food, nutrition & agriculture 11:27–
668 39.

669 10. Egli C, Edinger W, Mitrakul C, Henick-Kling T. 1998. Dynamics of indigenous and
670 inoculated yeast populations and their effect on the sensory character of Riesling and
671 Chardonnay wines. Journal of Applied Microbiology 85:779–789.

672 11. Bartowsky EJ. 2009. Bacterial spoilage of wine and approaches to minimize it. Letters
673 in Applied Microbiology 48:149–156.

674 12. Liu D, Zhang P, Chen D, Howell KS. 2019. From the vineyard to the winery: how
675 microbial ecology drives regional distinctiveness of wine. Frontiers in Microbiology
676 10:2679.

677 13. Blanco-Ulate B, Amrine KC, Collins TS, Rivero RM, Vicente AR, Morales-Cruz A, Doyle
678 CL, Ye Z, Allen G, Heymann H, others. 2015. Developmental and metabolic plasticity of

679 white-skinned grape berries in response to Botrytis cinerea during noble rot. Plant
680 physiology 169:2422–2443.

681 14. Bokulich NA, Thorngate JH, Richardson PM, Mills DA. 2014. Microbial biogeography of
682 wine grapes is conditioned by cultivar, vintage, and climate. Proceedings of the
683 National Academy of Sciences 111:E139–E148.

684 15. Bokulich NA, Collins TS, Masarweh C, Allen G, Heymann H, Ebeler SE, Mills DA. 2016.
685 Associations among wine grape microbiome, metabolome, and fermentation behavior
686 suggest microbial contribution to regional wine characteristics. MBio 7.

687 16. Pinto C, Pinho D, Cardoso R, Custódio V, Fernandes J, Sousa S, Pinheiro M, Egas C,
688 Gomes AC. 2015. Wine fermentation microbiome: a landscape from different
689 Portuguese wine appellations. Frontiers in microbiology 6:905.

690 17. Garofalo C, Russo P, Beneduce L, Massa S, Spano G, Capozzi V. 2016. Non-
691 Saccharomyces biodiversity in wine and the 'microbial terroir': A survey on Nero di
692 Troia wine from the Apulian region, Italy. Annals of microbiology 66:143–150.

693 18. Mezzasalma V, Sandionigi A, Bruni I, Bruno A, Lovicu G, Casiraghi M, Labra M. 2017.
694 Grape microbiome as a reliable and persistent signature of field origin and
695 environmental conditions in Cannonau wine production. PLoS One 12:e0184615.

696 19. Mezzasalma V, Sandionigi A, Guzzetti L, Galimberti A, Grando MS, Tardaguila J, Labra
697 M. 2018. Geographical and cultivar features differentiate grape microbiota in Northern
698 Italy and Spain vineyards. Frontiers in microbiology 9:946.

699 20. Singh P, Santoni S, This P, Péros J-P. 2018. Genotype-environment interaction shapes
700 the microbial assemblage in grapevine's phyllosphere and carposphere: an NGS
701 approach. Microorganisms 6:96.

702 21. Vitulo N, Lemos Jr WJF, Calgaro M, Confalone M, Felis GE, Zapparoli G, Nardi T. 2019.
703 Bark and grape microbiome of Vitis vinifera: influence of geographic patterns and
704 agronomic management on bacterial diversity. Frontiers in Microbiology 9:3203.

705 22. Knight S, Klaere S, Fedrizzi B, Goddard MR. 2015. Regional microbial signatures
706 positively correlate with differential wine phenotypes: evidence for a microbial aspect
707 to terroir. Scientific reports 5:1–10.

708 23. Conacher CG, Rossouw D, Bauer F. 2019. Peer pressure: evolutionary responses to
709 biotic pressures in wine yeasts. FEMS Yeast Research 19:foz072.

710 24. Bordet F, Joran A, Klein G, Roullier-Gall C, Alexandre H. 2020. Yeast–Yeast Interactions:
711 Mechanisms, Methodologies and Impact on Composition. Microorganisms 8:600.

712 25. Sadoudi M, Tourdot-Maréchal R, Rousseaux S, Steyer D, Gallardo-Chacón J-J, Ballester
713 J, Vichi S, Guérin-Schneider R, Caixach J, Alexandre H. 2012. Yeast–yeast interactions
714 revealed by aromatic profile analysis of Sauvignon Blanc wine fermented by single or
715 co-culture of non-Saccharomyces and Saccharomyces yeasts. Food Microbiology
716 32:243–253.

717 26. Sadoudi M, Rousseaux S, David V, Alexandre H, Tourdot-Maréchal R. 2017.
718 Metschnikowia pulcherrima influences the expression of genes involved in PDH

719 bypass and glyceropyruvic fermentation in Saccharomyces cerevisiae. Frontiers in
720 Microbiology 8:1137.

721 27. Renault P, Coulon J, de Revel G, Barbe J-C, Bely M. 2015. Increase of fruity aroma
722 during mixed T. delbrueckii/S. cerevisiae wine fermentation is linked to specific esters
723 enhancement. International Journal of Food Microbiology 207:40–48.

724 28. Renault P, Coulon J, Moine V, Thibon C, Bely M. 2016. Enhanced 3-sulfanylhexan-1-ol
725 production in sequential mixed fermentation with Torulaspora
726 delbrueckii/Saccharomyces cerevisiae reveals a situation of synergistic interaction
727 between two industrial strains. Frontiers in microbiology 7:293.

728 29. Englezos V, Torchio F, Cravero F, Marengo F, Giacosa S, Gerbi V, Rantsiou K, Rolle L,
729 Cocolin L. 2016. Aroma profile and composition of Barbera wines obtained by mixed
730 fermentations of Starmerella bacillaris (synonym Candida zemplinina) and
731 Saccharomyces cerevisiae. LWT 73:567–575.

732 30. Englezos V, Pollon M, Rantsiou K, Ortiz-Julien A, Botto R, Segade SR, Giacosa S, Rolle L,
733 Cocolin L. 2019. Saccharomyces cerevisiae-Starmerella bacillaris strains interaction
734 modulates chemical and volatile profile in red wine mixed fermentations. Food
735 research international 122:392–401.

736 31. Lencioni L, Romani C, Gobbi M, Comitini F, Ciani M, Domizio P. 2016. Controlled mixed
737 fermentation at winery scale using Zygotorulaspora florentina and Saccharomyces
738 cerevisiae. International Journal of Food Microbiology 234:36–44.

739 32. Gobert A, Tourdot-Maréchal R, Morge C, Sparrow C, Liu Y, Quintanilla-Casas B, Vichi S,
740 Alexandre H. 2017. Non-Saccharomyces yeasts nitrogen source preferences: impact on
741 sequential fermentation and wine volatile compounds profile. Frontiers in
742 microbiology 8:2175.

743 33. Whitener MB, Stanstrup J, Carlin S, Divol B, Du Toit M, Vrhovsek U. 2017. Effect of non-
744 Saccharomyces yeasts on the volatile chemical profile of Shiraz wine. Australian
745 journal of grape and wine research 23:179–192.

746 34. Binati RL, Junior WJL, Luzzini G, Slaghenaufi D, Ugliano M, Torriani S. 2020.
747 Contribution of non-Saccharomyces yeasts to wine volatile and sensory diversity: A
748 study on Lachancea thermotolerans, Metschnikowia spp. and Starmerella bacillaris
749 strains isolated in Italy. International Journal of Food Microbiology 318:108470.

750 35. Brou P, Taillandier P, Beaufort S, Brandam C. 2018. Mixed culture fermentation using
751 Torulaspora delbrueckii and Saccharomyces cerevisiae with direct and indirect
752 contact: impact of anaerobic growth factors. European Food Research and Technology
753 244:1699–1710.

754 36. Curiel JA, Morales P, Gonzalez R, Tronchoni J. 2017. Different non-Saccharomyces
755 yeast species stimulate nutrient consumption in S. cerevisiae mixed cultures. Frontiers
756 in Microbiology 8:2121.

757 37. Alonso-del-Real J, Pérez-Torrado R, Querol A, Barrio E. 2019. Dominance of wine
758 Saccharomyces cerevisiae strains over S. kudriavzevii in industrial fermentation

759 competitions is related to an acceleration of nutrient uptake and utilization.
760 Environmental microbiology 21:1627–1644.

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

765 39. Shekhawat K, Patterton H, Bauer FF, Setati ME. 2019. RNA-seq based transcriptional
766 analysis of Saccharomyces cerevisiae and Lachancea thermotolerans in mixed-culture
767 fermentations under anaerobic conditions. BMC genomics 20:145.

768 40. Brown JC, Lindquist S. 2009. A heritable switch in carbon source utilization driven by
769 an unusual yeast prion. Genes & development 23:2320–2332.

770 41. Jarosz DF, Brown JC, Walker GA, Datta MS, Ung WL, Lancaster AK, Rotem A, Chang A,
771 Newby GA, Weitz DA, others. 2014. Cross-kingdom chemical communication drives a
772 heritable, mutually beneficial prion-based transformation of metabolism. Cell
773 158:1083–1093.

774 42. Bisson LF, Walker G, Ramakrishnan V, Luo Y, Fan Q, Wiemer E, Luong P, Ogawa M,
775 Joseph L. 2017. The two faces of Lactobacillus kunkeei: Wine spoilage agent and bee
776 probiotic. Catalyst: Discovery into Practice 1:1–11.

777 43. Morrison-Whittle P, Goddard MR. 2018. From vineyard to winery: a source map of
778 microbial diversity driving wine fermentation. Environmental microbiology 20:75–84.

779 44. David V, Terrat S, Herzine K, Claisse O, Rousseaux S, Tourdot-Maréchal R, Masneuf-
780 Pomarede I, Ranjard L, Alexandre H. 2014. High-throughput sequencing of amplicons
781 for monitoring yeast biodiversity in must and during alcoholic fermentation. Journal of
782 industrial microbiology & biotechnology 41:811–821.

783 45. Rossignol T, Dulau L, Julien A, Blondin B. 2003. Genome-wide monitoring of wine yeast
784 gene expression during alcoholic fermentation. Yeast 20:1369–1385.

785 46. Marks VD, Ho Sui SJ, Erasmus D, Van Der Merwe GK, Brumm J, Wasserman WW, Bryan
786 J, Van Vuuren HJ. 2008. Dynamics of the yeast transcriptome during wine
787 fermentation reveals a novel fermentation stress response. FEMS yeast research 8:35–
788 52.

789 47. Barbosa C, Mendes-Faia A, Lage P, Mira NP, Mendes-Ferreira A. 2015. Genomic
790 expression program of Saccharomyces cerevisiae along a mixed-culture wine
791 fermentation with Hanseniaspora guilliermondii. Microbial cell factories 14:124.

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

796 49. Grainger C, Yeh A, Byer S, Hjelmeland A, Lima MM, Runnebaum RC. 2021. Vineyard site
797 impact on the elemental composition of Pinot noir wines. Food Chemistry
798 334:127386.

799 50. Steenwerth KL, Morelan IA, Stahel RJ, Figueroa R, Cantu D, Lee J, Runnebaum RC,
800 Poret-Peterson A. 2020. Fungal and bacterial communities of 'Pinot noir' must: Effects
801 of vintage, growing region, climate, and must chemistry. Accepted, in press.

802 51. Bubeck AM, Preiss L, Jung A, Dörner E, Podlesny D, Kulis M, Maddox C, Arze C, Zörb C,
803 Merkt N, others. 2020. Bacterial microbiota diversity and composition in red and
804 white wines correlate with plant-derived DNA contributions and botrytis infection.
805 Scientific reports 10:1–13.

806 52. Bartowsky EJ, Henschke PA. 2008. Acetic acid bacteria spoilage of bottled red wine—a
807 review. International journal of food microbiology 125:60–70.

808 53. Pietrafesa A, Capece A, Pietrafesa R, Bely M, Romano P. 2020. Saccharomyces
809 cerevisiae and Hanseniaspora uvarum mixed starter cultures: Influence of
810 microbial/physical interactions on wine characteristics. Yeast.

811 54. Lohman BK, Weber JN, Bolnick DI. 2016. Evaluation of TagSeq, a reliable low-cost
812 alternative for RNA seq. Molecular ecology resources 16:1315–1321.

813 55. Englezos V, Cocolin L, Rantsiou K, Ortiz-Julien A, Bloem A, Dequin S, Camarasa C. 2018.
814 Specific phenotypic traits of Starmerella bacillaris related to nitrogen source
815 consumption and central carbon metabolite production during wine fermentation.
816 Applied and environmental microbiology 84.

817 56. Boscaino F, Ionata E, La Cara F, Guerriero S, Marcolongo L, Sorrentino A. 2019. Impact
818 of Saccharomyces cerevisiae and Metschnikowia fructicola autochthonous mixed

819 starter on Aglianico wine volatile compounds. Journal of food science and technology
820 56:4982–4991.

821 57. Cottier F, Srinivasan KG, Yurieva M, Liao W, Poidinger M, Zolezzi F, Pavelka N. 2018.
822 Advantages of meta-total RNA sequencing (MeTRS) over shotgun metagenomics and
823 amplicon-based sequencing in the profiling of complex microbial communities. npj
824 Biofilms and Microbiomes 4:1–7.

825 58. Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T,
826 Schroeder M, others. 1998. SGD: Saccharomyces genome database. Nucleic acids
827 research 26:73–79.

828 59. Gregorutti B, Michel B, Saint-Pierre P. 2017. Correlation and variable importance in
829 random forests. Statistics and Computing 27:659–678.

830 60. Lage P, Belem Sampaio-Marques PL, Mira NP, Mendes-Ferreira A. 2019.
831 Transcriptomic and chemogenomic analyses unveil the essential role of Com2-regulon
832 in response and tolerance of Saccharomyces cerevisiae to stress induced by sulfur
833 dioxide. Microbial Cell 6:509.

834 61. Hinze H, Holzer H. 1986. Analysis of the energy metabolism after incubation of
835 Saccharomyces cerevisiae with sulfite or nitrite. Archives of microbiology 145:27–31.

836 62. Divol B, du Toit M, Duckitt E. 2012. Surviving in the presence of sulphur dioxide:
837 strategies developed by wine yeasts. Applied microbiology and biotechnology 95:601–
838 613.

839 63. Cordente AG, Curtin CD, Varela C, Pretorius IS. 2012. Flavour-active wine yeasts.
840 Applied Microbiology and Biotechnology 96:601–618.

841 64. Scrimgeour N, Nordestgaard S, Lloyd N, Wilkes E. 2015. Exploring the effect of
842 elevated storage temperature on wine composition. Australian journal of grape and
843 wine research 21:713–722.

844 65. Hansen KD, Wu Z, Irizarry RA, Leek JT. 2011. Sequencing technology does not
845 eliminate biological variability. Nature biotechnology 29:572–573.

846 66. Backhus LE, DeRisi J, Brown PO, Bisson LF. 2001. Functional genomic analysis of a
847 commercial wine strain of Saccharomyces cerevisiae under differing nitrogen
848 conditions. FEMS Yeast Research 1:111–125.

849 67. Morales-Cruz A, Figueroa-Balderas R, García JF, Tran E, Rolshausen PE, Baumgartner
850 K, Cantu D. 2018. Profiling grapevine trunk pathogens in planta: A case for
851 community-targeted DNA metabarcoding. BMC microbiology 18:1–14.

852 68. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput
853 sequencing reads. EMBnet journal 17:10–12.

854 69. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2:
855 high-resolution sample inference from Illumina amplicon data. Nature methods
856 13:581–583.

857 70. Brown CT, Irber L. 2016. sourmash: a library for MinHash sketching of DNA. Journal of
858 Open Source Software 1:27.

859 71. Pierce NT, Irber L, Reiter T, Brooks P, Brown CT. 2019. Large-scale sequence
860 comparisons with sourmash. F1000Research 8.

861 72. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X,
862 Korzeniewski F, others. 2014. MycoCosm portal: gearing up for 1000 fungal genomes.
863 Nucleic acids research 42:D699–D704.

864 73. Lawrence M, Gentleman R, Carey V. 2009. rtracklayer: an R package for interfacing
865 with genome browsers. Bioinformatics 25:1841–1842.

866 74. Hoff KJ, Stanke M. 2013. WebAUGUSTUS—a web service for training AUGUSTUS and
867 predicting genes in eukaryotes. Nucleic acids research 41:W123–W128.

868 75. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
869 Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–
870 21.

871 76. Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in Unique
872 Molecular Identifiers to improve quantification accuracy. Genome research 27:491–
873 499.

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

876 78. Borneman AR, Forgan AH, Kolouchova R, Fraser JA, Schmidt SA. 2016. Whole genome
877 comparison reveals high levels of inbreeding and strain redundancy across the

878 spectrum of commercial wine strains of Saccharomyces cerevisiae. G3: Genes,
879 Genomes, Genetics 6:957–971.

880 79. Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, Charbonneau
881 A, Constantinides B, Edvenson G, Fay S, others. 2015. The khmer software package:
882 enabling efficient nucleotide sequence analysis. F1000Research 4.

883 80. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node
884 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
885 Bioinformatics 31:1674–1676.

886 81. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. 2017. Microbiome datasets
887 are compositional: and this is not optional. Frontiers in microbiology 8:2224.

888 82. Lahti L, Shetty S, Blake T, Salojarvi J, others. 2017. Tools for microbiome analysis in R.
889 Version 1:10013.

890 83. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge
891 Y, Gentry J, others. 2004. Bioconductor: open software development for computational
892 biology and bioinformatics. Genome biology 5:R80.

893 84. Aitchison J, Barcelo-Vidal C, Martin-Fernandez JA, Pawlowsky-Glahn V. 2000. Logratio
894 analysis and compositional distance. Mathematical Geology 32:271–275.

895 85. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M.
896 2007. The vegan package. Community ecology package 10:719.

86. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, others. 2019. Welcome to the Tidyverse. Journal of Open Source Software 4:1686.

87. Wright MN, Ziegler A. 2015. ranger: A fast implementation of random forests for high dimensional data in C++ and R. arXiv preprint arXiv:150804409.

88. Probst P. 2018. tuneRanger: Tune random forest of the ranger package. R package version 02.

89. Kuhn M, others. 2008. Building predictive models in R using the caret package. Journal of statistical software 28:1–26.

90. Janitza S, Celik E, Boulesteix A-L. 2018. A computationally fast variable importance test for random forests for high-dimensional data. Advances in Data Analysis and Classification 12:885–915.

91. Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer research 27:209–220.

92. Legendre P, Legendre LF. 2012. Numerical ecology. Elsevier.

93. Daly C, Taylor G, Gibson W, Parzybok T, Johnson G, Pasteris P. 2000. High-quality spatial climate data sets for the United States and beyond. Transactions of the ASAE 43:1957.

94. Hijmans RJ, Williams E, Vennes C, Hijmans MRJ. 2017. Package 'geosphere.' Spherical trigonometry 1:7.

918 **Figure Legends**

919 Figure 1: **Microbial diversity in grape must and fermentation microbiomes from**
920 **different vineyard sites. A, B)** Relative abundance of taxonomic ranks in ribosomal DNA
921 amplicon sequencing data capturing **A** Bacteria and **B** Fungi. Samples taken from
922 fermentations from the same vineyard site and vintage are combined together and reflect
923 relative abundance of organisms from four fermentation tanks. Only three tanks were
924 fermented for AV2 in 2019 due to a smaller harvest. **C)** Relative abundance of all genes
925 expressed by a detected organism during fermentation from the 2017 and 2019 vintages.
926 The top plots show all organisms and bottom plots display only those organisms that
927 account for less than 3% of mapped reads in each sample. Only organisms present in more
928 than one fermentation are plotted.

929 Figure 2: ***H. uvarum* ribosomal DNA amplicon sequencing data does not strongly**
930 **correlate with relative abundance in RNA sequencing data. A** Bar chart of relative
931 abundance of *H. uvarum* compared to other non-*Saccharomyces* species across
932 fermentations from each site based on amplicon sequencing data of ribosomal DNA. **B**
933 Scatter plots relative abundance of *H. uvarum* as determined by amplicon sequencing of
934 ribosomal DNA (x-axis) vs. RNA sequencing (y-axis).

935 Figure 3: **Genetic profiles correlate with vineyard, region, and vintage as well as some**
936 **vineyard site and initial grape must characteristics. A-C)** Non-metric Multi-dimensional
937 Scaling plots of Aitchinson dissimilarity of **A** bacterial communities, **B** fungal communities,
938 and **C** and transcriptomes across vintages. The closer two points are on the graph, the more
939 similar their genetic profiles are. **D)** Vineyard site, region, and vintage account for genetic

940 diversity patterns in Analysis of Similarity (ANOSIM). R values represent strength of
941 association, with higher R values indicating stronger grouping according to the parameter.
942 All values are significant ($p < 0.001$). **E)** Percent of accuracy attributable to different
943 organisms in random forests models. A higher percentage of variable importance was
944 attributable to *S. cerevisiae* and *V. vinifera* in models that predicted region than vineyard
945 site. **F, G)** Correlograms representing similarities between fermentation metrics. **F** Grape
946 must chemical parameters and vineyard site characteristics were correlated in the 2017
947 and 2019 vintages. Squares are labelled with correlation values from Pearson's correlation.
948 Only comparisons with $p < 0.05$ are displayed. **G** Bacterial, fungal, and transcriptome
949 profiles correlated with some vineyard site and grape must chemical characteristics.
950 Squares are labelled with correlation values from Mantel tests. Only comparisons with an
951 FDR < 0.1 are displayed. PPT: precipitation, GDD: growing degree days, MA: malic acid, TA:
952 titratable acidity.

953 Figure S1: ***Diversity of vineyards and ribosomal DNA profiles in this study.*** **A)** Map
954 displaying the 15 vineyard locations across eight American Viticultural Areas (AVAs) in
955 California and Oregon. **B, C)** Bacterial and fungal ribosomal DNA amplicon sequencing Chao
956 1 and Shannon alpha diversity for mean species diversity per vineyard site, averaged
957 across vintages. **B** Bacteria. **C** Fungi.

958 Figure S2: ***Some ribosomal sequencing variants were detected across vineyards and***
959 ***vintages****. Top* 20 most abundant ribosomal DNA amplicon sequencing variants across
960 vintages. Labelled as genus or the next lowest taxonomic rank of classification. **A** Bacteria.
961 **B** Fungi. *Tatumella* was the most abundant bacterial amplicon sequencing variant across

962 vineyards and vintages, while *Hanseniaspora* was the most abundant fungal amplicon
963 sequencing variant.

964 Figure S3: **Accuracy of random forests models using bacterial ribosomal DNA profiles**.
965 Confusion matrices depicting accuracy of random forests models built with bacterial
966 ribosomal DNA amplicon sequencing data to predict **A)** vineyard site and **B)** vineyard
967 region. The models depicted were trained on two vintages and validated on the third.

968 Figure S4: **Accuracy of random forests models using fungal ribosomal DNA profiles**.
969 Confusion matrices depicting accuracy of random forests models built with fungal
970 ribosomal DNA amplicon sequencing data to predict **A)** vineyard site and **B)** vineyard
971 region. The models depicted were trained on two vintages and validated on the third.

972 Figure S5: **Accuracy of random forests models using RNA sequencing**. Confusion
973 matrices depicting accuracy of random forests models built with RNA sequencing data to
974 predict **A)** vineyard site and **B)** vineyard region. The models depicted were trained on one
975 vintage and validated on the other.

976 Figure S6: **Percent of accuracy attributable to different organisms in random forests**
977 **models**. Importance of genes expressed by different organisms in the overall model. A
978 higher percentage of variable importance was attributable to *S. cerevisiae* and *V. vinifera* in
979 models that predicted region than vineyard site.

980 Figure S7: **Initial yeast assimilable nitrogen (YAN) in grape musts across vintages.**
981 Black dots mark the mean initial YAN value calculated from all fermentations in the 2017
982 and 2019 vintages. *MEP3*, which encodes an ammonia permease, was important for

983 predicting the three regions with the lowest average initial YAN (OR, AV, RRV) and the

984 region with the second highest initial YAN (SMV).

985 **Supplemental Data Tables**

986 For Tables S2 – S7: see accompanying supplemental files.

987 Table S1: Accuracy of random forests models built with fungal and bacterial ribosomal DNA
988 amplicon sequencing data and transcriptome data. Validation set "30" indicates models
989 that were trained with 70% of data and validated on the held-out 30%.

Data set	Model type	Validation set	Accuracy
bacteria	AVA	2019	0.74
bacteria	AVA	2017	0.68
bacteria	AVA	2016	0.69
bacteria	AVA	30	0.93
bacteria	site	2019	0.53
bacteria	site	2017	0.5
bacteria	site	2016	0.54
bacteria	site	30	0.93
fungi	AVA	2019	0.67
fungi	AVA	2017	0.6
fungi	AVA	2016	0.67
fungi	AVA	30	0.95
fungi	site	2019	0.33
fungi	site	2017	0.23
fungi	site	2016	0.56
fungi	site	30	0.89
transcriptome	AVA	2019	0.57
transcriptome	AVA	2017	0.63
transcriptome	AVA	30	0.87
transcriptome	site	2019	0.43
transcriptome	site	2017	0.43
transcriptome	site	30	0.67

990

991 Table S8: Species names and accession numbers for genomes used in this study.

species	accession
Aureobasidium pullulans	GCA_004917305.1
Aureobasidium pullulans	GCA_000721785.1
Botrytis cinerea	GCA_000143535.4
Botrytis cinerea	GCA_000349525.1
Cladosporium sp. SL-16	GCA_002921095.1
Hanseniaspora opuntiae	GCA_001749795.1
Hanseniaspora uvarum	GCA_000968475.1
Lachancea thermotolerans	GCF_000142805.1
Metschnikowia fructicola	GCA_000317355.2
Metschnikowia sp. AWRI3582	GCA_002894445.1
Pichia kudriavzevii	GCA_003054445.1
Rhizopus stolonifer	GCA_003325415.1
Saccharomyces cerevisiae	GCA_000146045.2
Starmerella bacillaris	GCA_002024125.1
Vitis vinifera	GCA_000003745.2

992

Figure 1

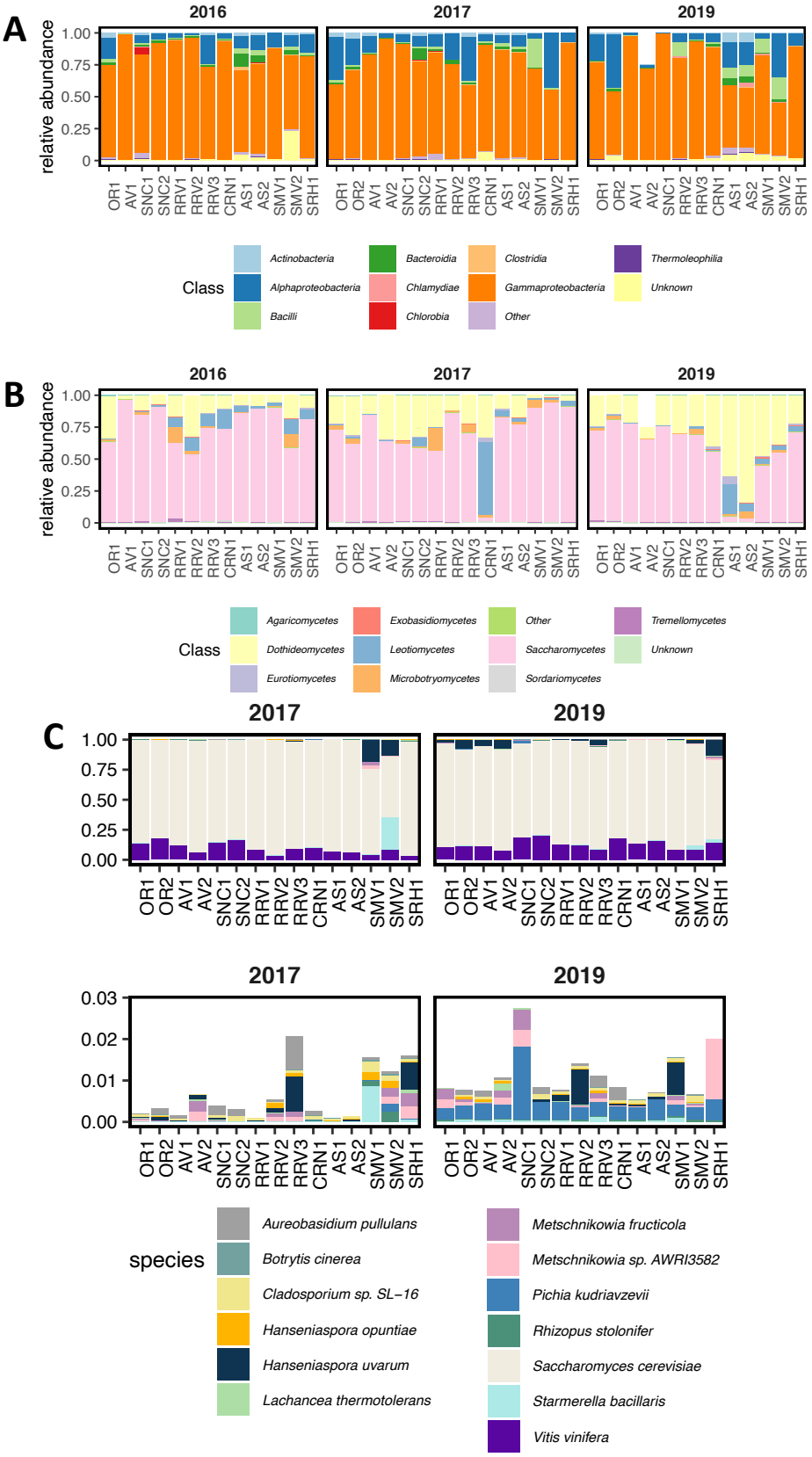

A
2016
2017
2019
relative abundance
1.00
0.75
0.50
0.25
0
OR1 AV1 SNC1 SNC2 RRV1 RRV2 RRV3 CRN1 AS1 AS2 SMV1 SMV2 SRH1
OR1 OR2 AV1 AV2 SNC1 SNC2 RRV1 RRV2 RRV3 CRN1 AS1 AS2 SMV1 SMV2 SRH1
OR1 OR2 AV1 AV2 SNC1 RRV2 RRV3 CRN1 AS1 AS2 SMV1 SMV2 SRH1
Class
Actinobacteria
Alphaproteobacteria
Bacilli
Bacteroidia
Chlamydiae
Chlorobia
Clostridia
Gammaproteobacteria
Other
Thermoleophilia
Unknown
B
2016
2017
2019
relative abundance
1.00
0.75
0.50
0.25
0
OR1 AV1 SNC1 SNC2 RRV1 RRV2 RRV3 CRN1 AS1 AS2 SMV1 SMV2 SRH1
OR1 OR2 AV1 AV2 SNC1 SNC2 RRV1 RRV2 RRV3 CRN1 AS1 AS2 SMV1 SMV2 SRH1
OR1 OR2 AV1 AV2 SNC1 RRV2 RRV3 CRN1 AS1 AS2 SMV1 SMV2 SRH1
Class
Agaricomycetes
Dothideomycetes
Eurotiomycetes
Exobasidiomycetes
Leotiomycetes
Microbotryomycetes
Other
Saccharomycetes
Sordariomycetes
Tremellomycetes
Unknown
C
2017
2019
1.00
0.75
0.50
0.25
0.00
OR1 OR2 AV1 AV2 SNC1 SNC2 RRV1 RRV2 RRV3 CRN1 AS1 AS2 SMV1 SMV2 SRH1
OR1 OR2 AV1 AV2 SNC1 SNC2 RRV1 RRV2 RRV3 CRN1 AS1 AS2 SMV1 SMV2 SRH1
2017
2019
0.03
0.02
0.01
0.00
OR1 OR2 AV1 AV2 SNC1 SNC2 RRV1 RRV2 RRV3 CRN1 AS1 AS2 SMV1 SMV2 SRH1
OR1 OR2 AV1 AV2 SNC1 SNC2 RRV1 RRV2 RRV3 CRN1 AS1 AS2 SMV1 SMV2 SRH1
species
Aureobasidium pullulans
Botrytis cinerea
Cladosporium sp. SL−16
Hanseniaspora opuntiae
Hanseniaspora uvarum
Lachancea thermotolerans
Metschnikowia fructicola
Metschnikowia sp. AWRI3582
Pichia kudriavzevii
Rhizopus stolonifer
Saccharomyces cerevisiae
Starmerella bacillaris
Vitis vinifera

Figure 2

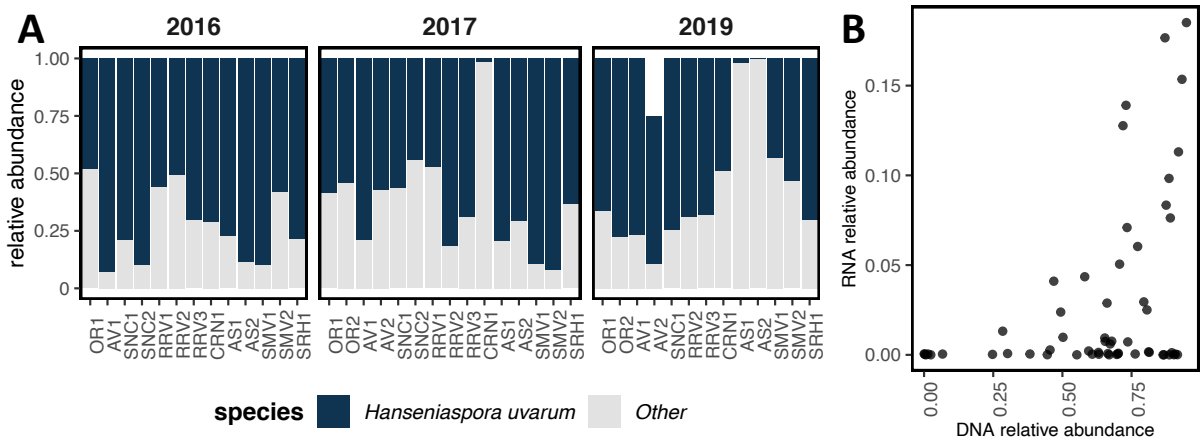
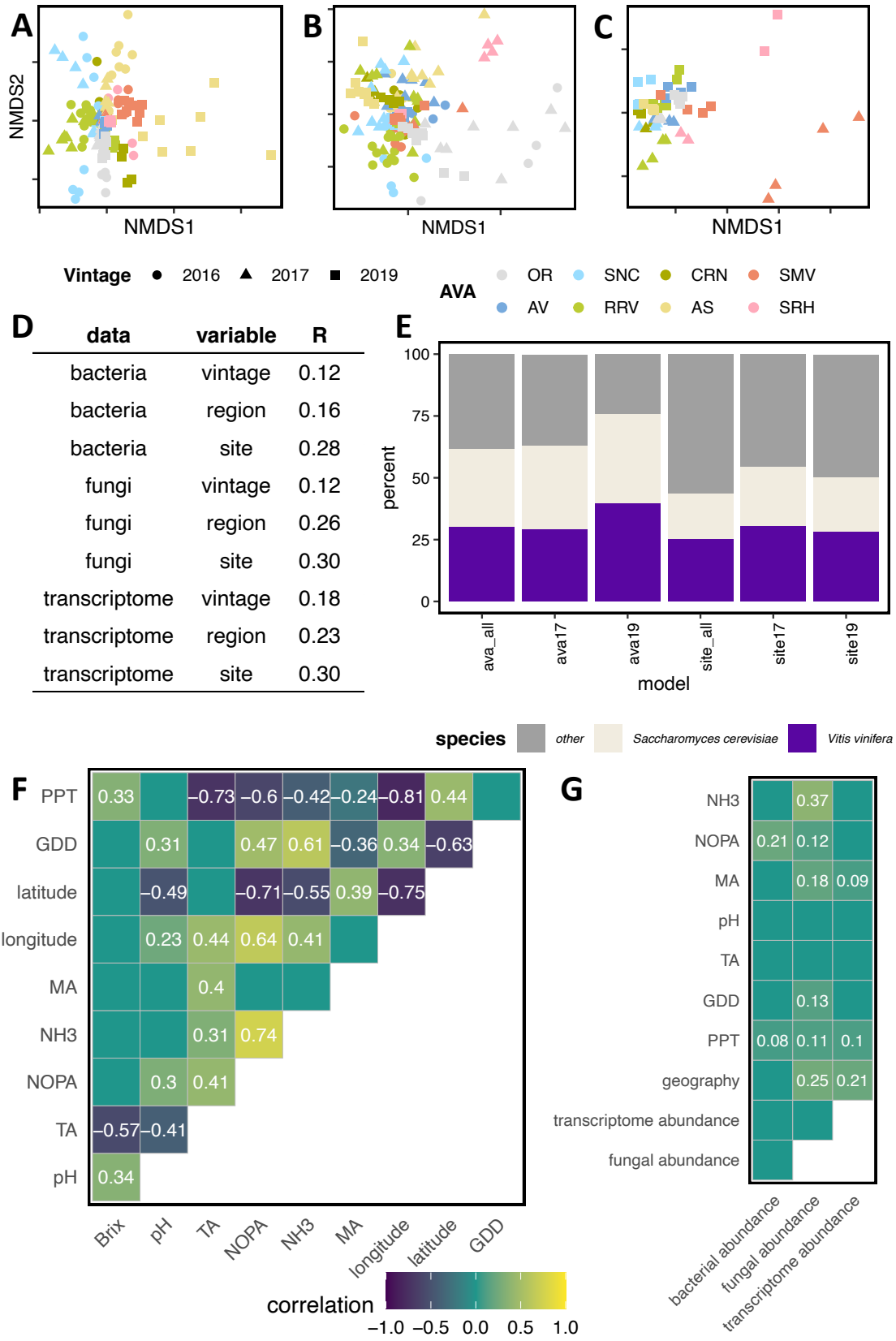

A
2016
2017
2019
relative abundance
1.00
0.75
0.50
0.25
0
OR1
AV1
SNC1
SNC2
RRV1
RRV2
RRV3
CRN1
AS1
AS2
SMV1
SMV2
SRH1
OR2
AV2
species
Hanseniaspora uvarum
Other
B
RNA relative abundance
0.15
0.10
0.05
0.00
DNA relative abundance

Figure 3



data	variable	R
bacteria	vintage	0.12
bacteria	region	0.16
bacteria	site	0.28
fungi	vintage	0.12
fungi	region	0.26
fungi	site	0.30
transcriptome	vintage	0.18
transcriptome	region	0.23
transcriptome	site	0.30

Figure S1

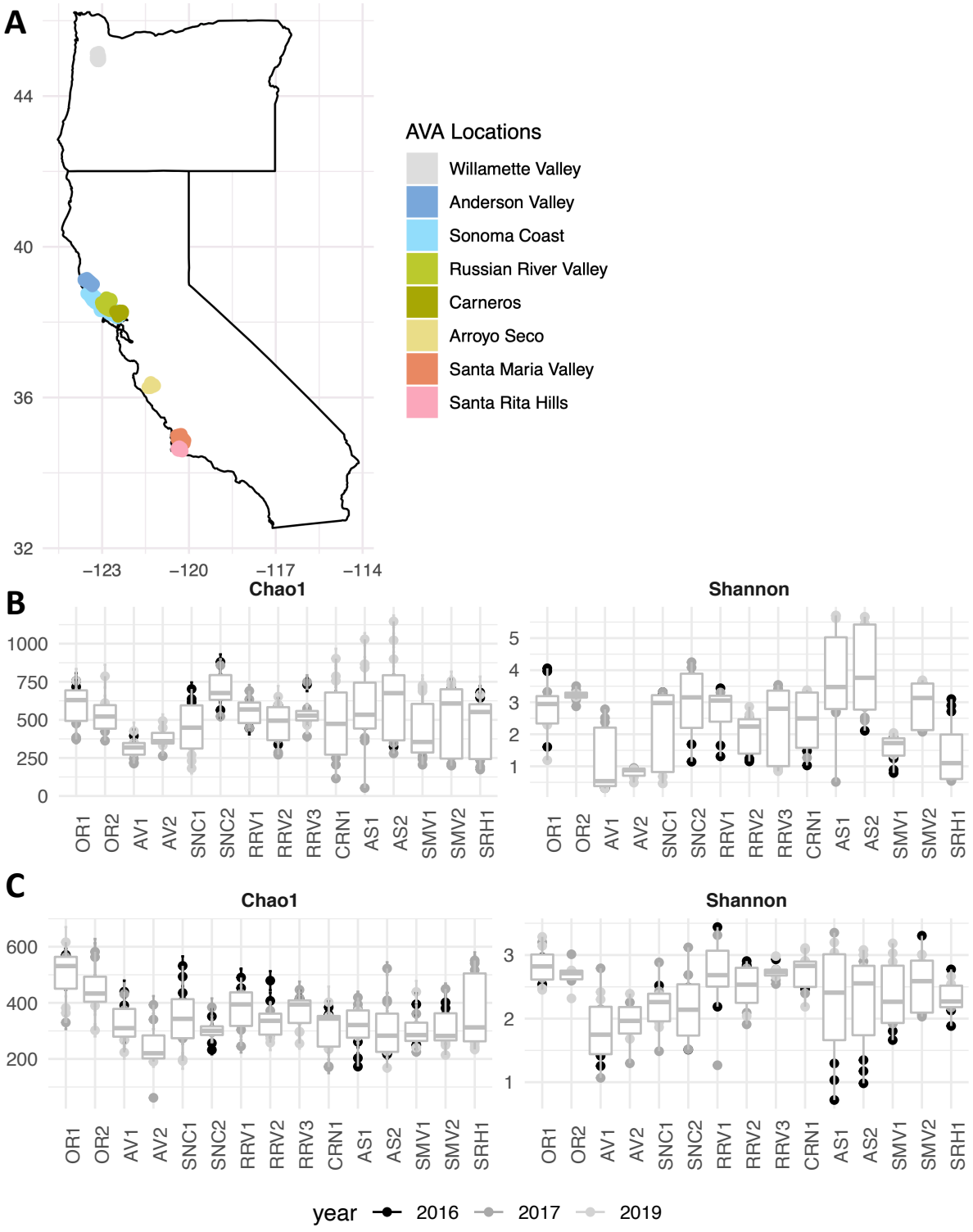

A
44
40
36
32
−123
−120
−117
−114
AVA Locations
Willamette Valley
Anderson Valley
Sonoma Coast
Russian River Valley
Carneros
Arroyo Seco
Santa Maria Valley
Santa Rita Hills
B
Chao1
Shannon
C
Chao1
Shannon
OR1
OR2
AV1
AV2
SNC1
SNC2
RRV1
RRV2
RRV3
CRN1
AS1
AS2
SMV1
SMV2
SRH1
year
2016
2017
2019

Figure S2

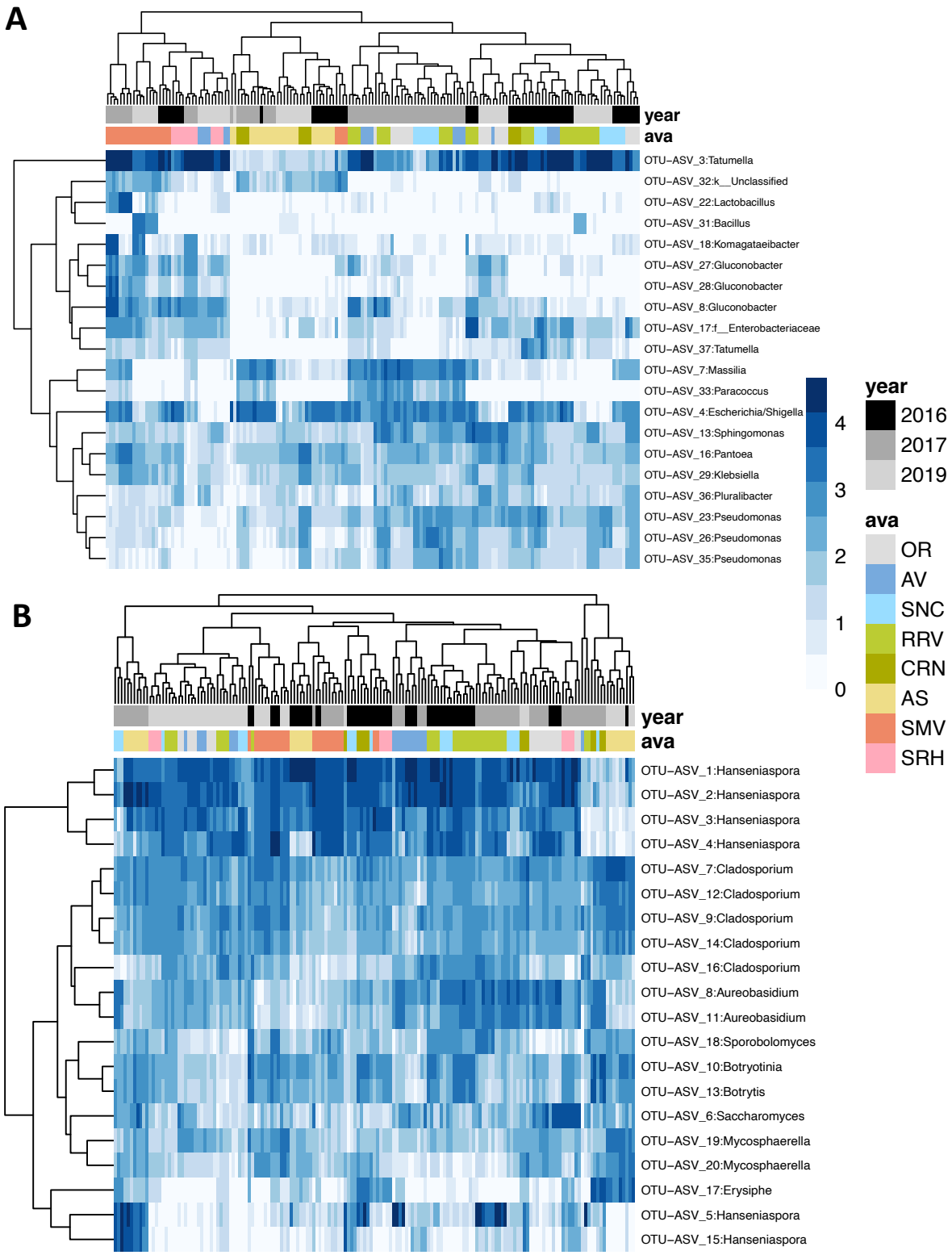

A
year
ava
OTU–ASV_3:Tatumella
OTU–ASV_32:k__Unclassified
OTU–ASV_22:Lactobacillus
OTU–ASV_31:Bacillus
OTU–ASV_18:Komagataeibacter
OTU–ASV_27:Gluconobacter
OTU–ASV_28:Gluconobacter
OTU–ASV_8:Gluconobacter
OTU–ASV_17:f__Enterobacteriaceae
OTU–ASV_37:Tatumella
OTU–ASV_7:Massilia
OTU–ASV_33:Paracoccus
OTU–ASV_4:Escherichia/Shigella
OTU–ASV_13:Sphingomonas
OTU–ASV_16:Pantoea
OTU–ASV_29:Klebsiella
OTU–ASV_36:Pluralibacter
OTU–ASV_23:Pseudomonas
OTU–ASV_26:Pseudomonas
OTU–ASV_35:Pseudomonas
4
3
2
1
0
year
2016
2017
2019
ava
OR
AV
SNC
RRV
CRN
AS
SMV
SRH
B
year
ava
OTU–ASV_1:Hanseniaspora
OTU–ASV_2:Hanseniaspora
OTU–ASV_3:Hanseniaspora
OTU–ASV_4:Hanseniaspora
OTU–ASV_7:Cladosporium
OTU–ASV_12:Cladosporium
OTU–ASV_9:Cladosporium
OTU–ASV_14:Cladosporium
OTU–ASV_16:Cladosporium
OTU–ASV_8:Aureobasidium
OTU–ASV_11:Aureobasidium
OTU–ASV_18:Sporobolomyces
OTU–ASV_10:Botryotinia
OTU–ASV_13:Botrytis
OTU–ASV_6:Saccharomyces
OTU–ASV_19:Mycosphaerella
OTU–ASV_20:Mycosphaerella
OTU–ASV_17:Erysiphe
OTU–ASV_5:Hanseniaspora
OTU–ASV_15:Hanseniaspora

Figure S3

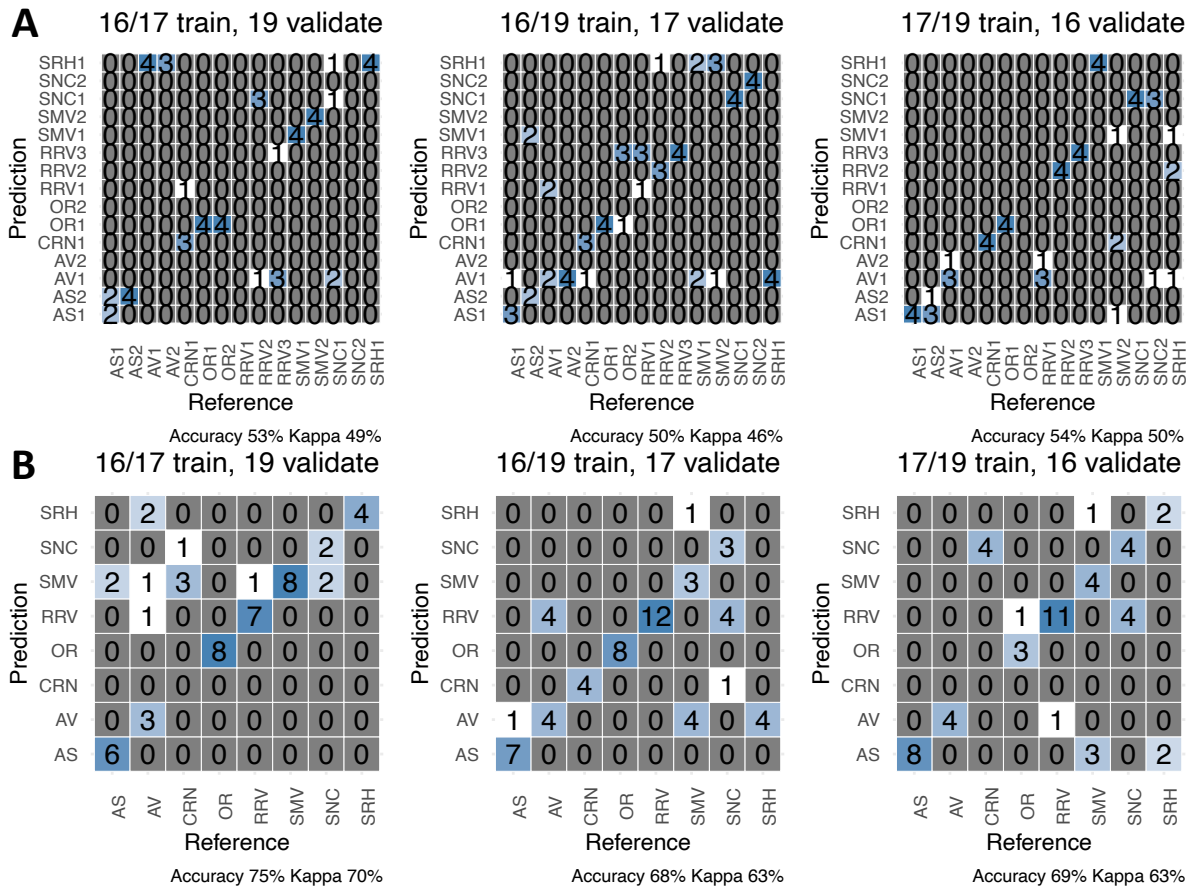

A
16/17 train, 19 validate
16/19 train, 17 validate
17/19 train, 16 validate
Prediction
Reference
SRH1 SNC2 SNC1 SMV2 SMV1 RRV3 RRV2 RRV1 OR2 OR1 CRN1 AV2 AV1 AS2 AS1
AS1 AS2 AV1 AV2 CRN1 OR1 OR2 RRV1 RRV2 RRV3 SMV1 SMV2 SNC1 SNC2 SRH1
Accuracy 53% Kappa 49%
Accuracy 50% Kappa 46%
Accuracy 54% Kappa 50%
B
16/17 train, 19 validate
16/19 train, 17 validate
17/19 train, 16 validate
Prediction
Reference
SRH SNC SMV RRV OR CRN AV AS
AS AV CRN OR RRV SMV SNC SRH
Accuracy 75% Kappa 70%
Accuracy 68% Kappa 63%
Accuracy 69% Kappa 63%

Figure S4

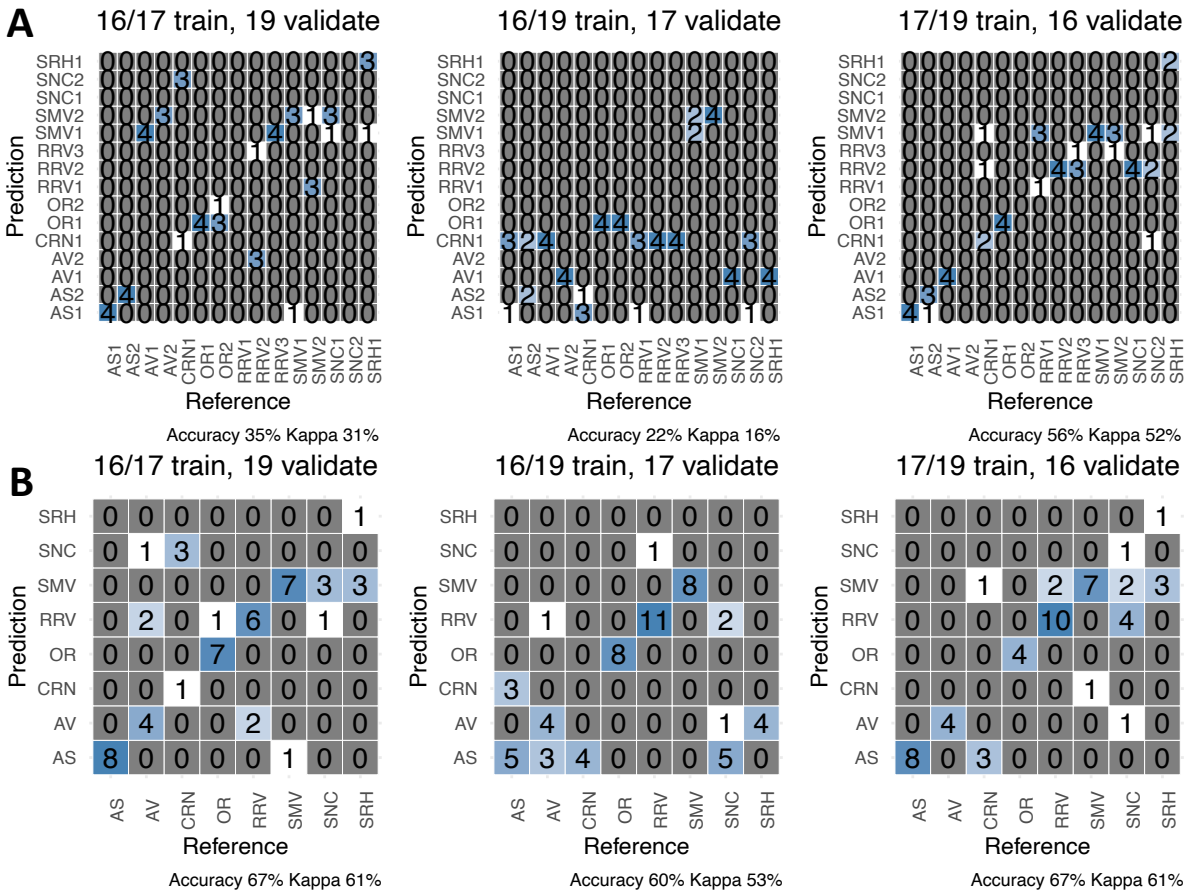

A
16/17 train, 19 validate
16/19 train, 17 validate
17/19 train, 16 validate
Prediction
Reference
Accuracy 35% Kappa 31%
Accuracy 22% Kappa 16%
Accuracy 56% Kappa 52%
B
16/17 train, 19 validate
16/19 train, 17 validate
17/19 train, 16 validate
Accuracy 67% Kappa 61%
Accuracy 60% Kappa 53%
Accuracy 67% Kappa 61%

Figure S5

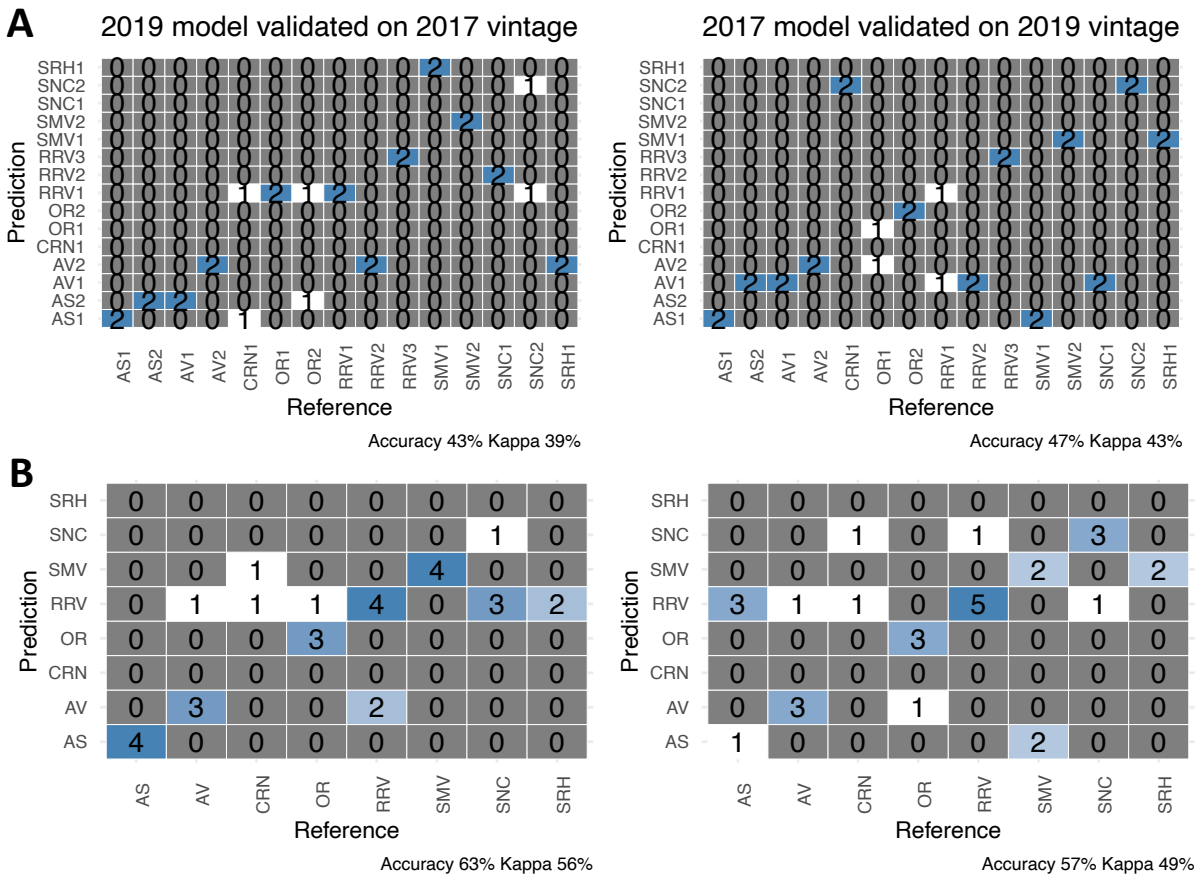

A
2019 model validated on 2017 vintage
Prediction
Reference
SRH1 SNC2 SNC1 SMV2 SMV1 RRV3 RRV2 RRV1 OR2 OR1 CRN1 AV2 AV1 AS2 AS1
AS1 AS2 AV1 AV2 CRN1 OR1 OR2 RRV1 RRV2 RRV3 SMV1 SMV2 SNC1 SNC2 SRH1
Accuracy 43% Kappa 39%
2017 model validated on 2019 vintage
Accuracy 47% Kappa 43%
B
SRH SNC SMV RRV OR CRN AV AS
AS AV CRN OR RRV SMV SNC SRH
Accuracy 63% Kappa 56%
Accuracy 57% Kappa 49%

Figure S6

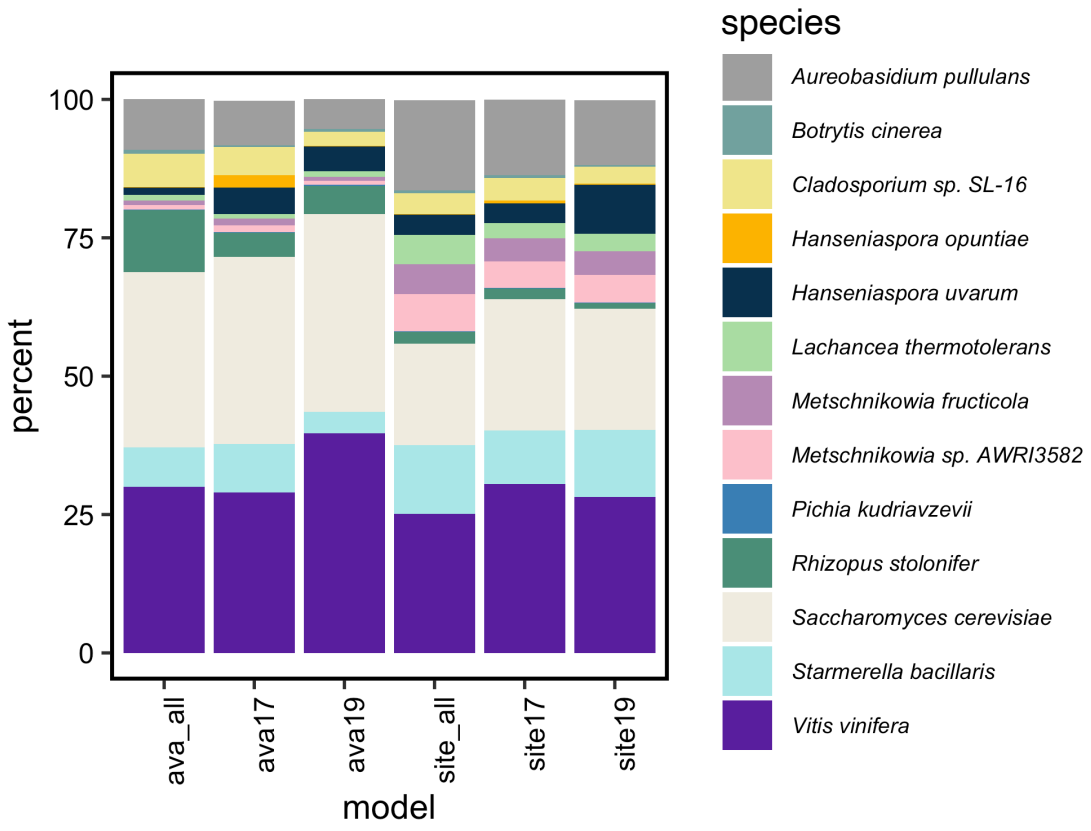
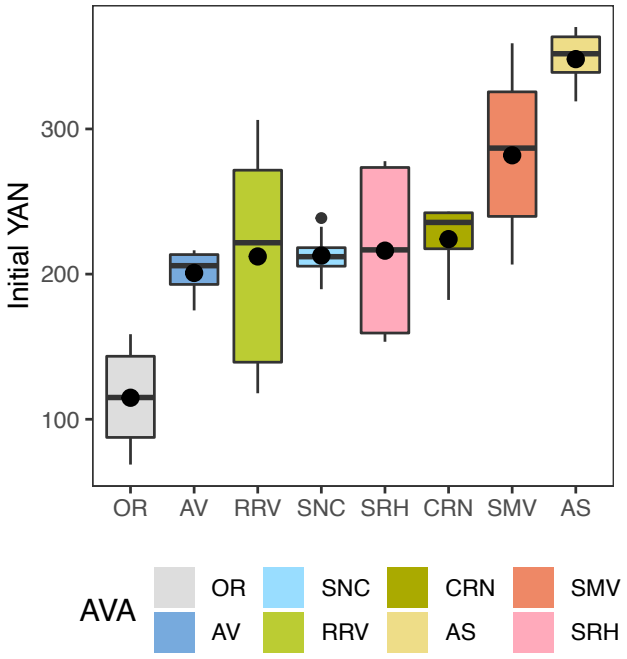

species
Aureobasidium pullulans
Botrytis cinerea
Cladosporium sp. SL-16
Hanseniaspora opuntiae
Hanseniaspora uvarum
Lachancea thermotolerans
Metschnikowia fructicola
Metschnikowia sp. AWRI3582
Pichia kudriavzevii
Rhizopus stolonifer
Saccharomyces cerevisiae
Starmerella bacillaris
Vitis vinifera
percent
100
75
50
25
0
ava_all
ava17
ava19
site_all
site17
site19
model

Figure S7


Initial YAN
300
200
100
OR
AV
RRV
SNC
SRH
CRN
SMV
AS
AVA
OR
AV
SNC
RRV
CRN
AS
SMV
SRH