npg

ORIGINAL ARTICLE

Geographic delineations of yeast communities and populations associated with vines and wines in New Zealand

Velimir Gayevskiy and Matthew R Goddard
School of Biological Sciences, University of Auckland, Auckland Mail Centre, Auckland, New Zealand

Yeasts are a diverse seemingly ubiquitous group of eukaryotic microbes, and many are naturally associated with fruits. Humans have harnessed yeasts since the dawn of civilisation to make wine, and thus it is surprising that we know little of the distribution of yeast communities naturally associated with fruits. Previous reports of yeast community diversity have been descriptive only. Here we present, we believe, the first robust test for the geographic delineation of yeast communities. Humans have relatively recently employed *Saccharomyces cerevisiae* as a model research organism, and have long harnessed its ancient adaption to ferment even in the presence of oxygen. However, as far as we are aware, there has not been a rigorous test for the presence of regional differences in natural *S. cerevisiae* populations before. We combined these community- and population-level questions and surveyed replicate vineyards and corresponding spontaneous ferments from different regions on New Zealand's (NZ's) North Island and analysed the resulting data with community ecology and population genetic tests. We show that there are distinct regional delineations of yeast communities, but the picture for *S. cerevisiae* is more complex: there is evidence for region-specific sub-populations but there are also reasonable levels of gene flow among these regions in NZ. We believe this is the first demonstration of regional delineations of yeast populations and communities worldwide.
The ISME Journal (2012) **6**, 1281–1290; doi:10.1038/ismej.2011.195; published online 22 December 2011
Subject Category: microbial population and community ecology
Keywords: biogeography; community ecology; yeast; population structure; *Saccharomyces cerevisiae*

Introduction

Characterising the distribution of organisms in space and time, and describing the processes that govern these patterns, is central to an integrated description of the natural world (Cox and Moore, 1993). Although the biogeography of plants and animals is reasonably well described (Gaston, 2000), the biogeography of microbes is not (Martiny *et al.*, 2006). Technical advances mean the task of describing microbial diversity is accelerating (Hugenholtz *et al.*, 1998; Green *et al.*, 2004), and not unsurprisingly the emerging picture is more complicated than simply accepting or rejecting the classic Bass–Becking idea that dispersal is not limiting for microbes and thus everything is everywhere but the environment selects (Whitaker *et al.*, 2003; Finlay and Fenchel, 2004; Green *et al.*, 2004; Martiny *et al.*, 2006; Nemergut *et al.*, 2011). Most studies to date have focused on bacterial communities (Martiny *et al.*, 2006), but some notable work has been conducted on the biogeography of aquatic protists (Finlay and Fenchel, 2004), soil fungi (Green *et al.*, 2004), and cacti- and flower-associated yeasts (Lachance *et al.*, 2000, 2001, 2008). However, there are fewer studies that have examined the biogeography of eukaryotic microbes in other niches.

Since the Age of Enlightenment, wine, like many other subjects of human passion, has been investigated from a scientific perspective, and this research has been intertwined with scientific progress generally (Chambers and Pretorius, 2010). Humans have been making wine since at least the dawn of civilisation (McGovern *et al.*, 2004; Chambers and Pretorius, 2010), but it was only within the last 150 years that Pasteur proved that microbes are the agents responsible for fermentation (Pretorius, 2000). Previous work has shown that a number of different yeast (which here we define as single-celled fungi) species are naturally present on grapes and in the resulting juice (Ribereau-Gayon *et al.*, 2006; Ciani *et al.*, 2010), and this includes surveys that span different regions (Schuller *et al.*, 2005). Although rigorous from a technical perspective, the shortcoming of most of these studies is that they are simply descriptive. We require robust ecological approaches in this field, with appropriate replication and the implementation of statistics to test

hypotheses; only recently have a number of studies begun to take this approach (for example, Goddard, 2008; Cadez *et al.*, 2010; Goddard *et al.*, 2010; Cordero-Bueso *et al.*, 2011).

Once in the winery, as a spontaneous (as opposed to artificially inoculated) ferment of grape juice proceeds, the species diversity collapses and *Saccharomyces* species, which are initially rare, take over and complete the ferment (Goddard, 2008; Ciani *et al.*, 2010). Although the species diversity falls, the diversity of different *S. cerevisiae* genotypes may remain high (Cubillos *et al.*, 2009; Goddard *et al.*, 2010). *Saccharomyces* ferment even in the presence of oxygen (the Crabtree effect), which is energetically wasteful, but this complex trait manipulates fruits to create hot, anaerobic, alcoholic environments that are essentially toxic to other microbes; thus *Saccharomyces* eventually dominate (Thomson *et al.*, 2005; Goddard, 2008). Whole genome analyses suggest this trait evolved around the same time as fruiting plants some 100 million years ago (Wolfe and Shields, 1997; Piskur *et al.*, 2006). As well as harnessing this trait to make wine, bread and beer, humans have more recently employed *S. cerevisiae* as a classic research model (Dujon, 1996; Burke *et al.*, 2000; Pretorius, 2000; Chambers and Pretorius, 2010). However, the complementary research to understand the ecology and population biology of *S. cerevisiae* has comparatively lagged behind (Landry *et al.*, 2006; Goddard, 2008; Replansky *et al.*, 2008; Liti *et al.*, 2009; Schacherer *et al.*, 2009; Goddard *et al.*, 2010). As a result of its model status, there has been work describing the diversity and population structure of *S. cerevisiae* (Fay and Benavides, 2005; Ezov *et al.*, 2006; Legras *et al.*, 2007; Liti *et al.*, 2009; Schacherer *et al.*, 2009; Goddard *et al.*, 2010), and its sibling-species *S. paradoxus* (Johnson *et al.*, 2004; Koufopanou *et al.*, 2006; Liti *et al.*, 2009; Zhang *et al.*, 2010b), on intercontinental scales, but very little work has been conducted on the population structure of *S. cerevisiae* associated with fruits and ferments. As well as differences over 3 years, Schuller and Casal (2007) showed significant differences between *S. cerevisiae* populations isolated from ferments of fruit from three different vineyards <100 km apart within the same region in Portugal. However, we are not aware of any previous studies that have conducted robust tests for regional-specific populations of *S. cerevisiae*, which by definition requires population genetic analyses of replicate samples from multiple regions. In sum, there has not yet been a rigorous test of hypotheses regarding the geographic delineations of either fruit-associated yeast communities or wine ferment yeast populations.

New Zealand (NZ) has only been inhabited by humans in the last ∼700 years (King, 2003), and has only had commercially planted vines in the last ∼100 years (Johnson, 1998). We have recently provided evidence that NZ harbours a genetically distinct population of *S. cerevisiae* using population genetic inference (Goddard *et al.*, 2010), and now we move on to test whether there are regional-specific populations of microbes within distinct wine-growing areas in NZ. Specifically we test two hypotheses: (1) whether there are different communities of yeasts associated with vineyards in different areas; and, (2) whether there are different populations of *S. cerevisiae* conducting spontaneous ferments in these same areas. We have tackled these questions by taking multiple samples of replicate Chardonnay and Syrah vineyards, and the spontaneous ferments deriving from them, from three well known and distinct wine-producing areas on NZ's North Island and tested the two main hypotheses outlined above using community ecology and population genetic analyses.

Materials and methods

Vineyard samples
We sampled 24 distinct vineyards in triplicate, for a total of 72 samples, across three wine-producing regions of NZ's North Island on an approximate N—S cline spanning ∼350 km: West Auckland, Waiheke Island and Hawke's Bay. While Waiheke Island is only ∼40 km away from the West Auckland region, it is an island in the Hauraki Gulf with a distinct microclimate. The mean distance between vineyards within regions is ∼5 km. Five to six Chardonnay vineyards were sampled within each region and these comprised the main experimental regime to test our hypotheses. We also sampled seven Syrah vineyards across Waiheke Island and Hawke's Bay (there is not enough Syrah planted in West Auckland to have included this region). All samples were taken approximately 1 week before harvest around mid-March in 2010: three whole bunches of fruit that appeared rot-free were selected randomly from each of the 24 vineyards and removed aseptically into a sterile plastic bag. The samples were transported back to the laboratory on ice, where the fruit was crushed *in situ* and the contents of the bag mixed thoroughly, and the juice was then immediately recovered into a sterile beaker. Serial dilutions were plated onto yeast extract, peptone, glucose agar (1% yeast extract, 2% peptone and 2% glucose) with chloramphenicol (200 µg ml^{-1}) to prevent bacterial growth and incubated at room temperature for 2–3 days to allow single cells to form colonies; 96 of these colonies were then selected randomly from each sample and stored at −80 °C.

Ferment samples
Commercial spontaneous ferments of fruit deriving from the same vineyards as we initially targeted were also sampled around three quarters of the way through ferment (when ∼75% of sugar had been consumed). In all, 50 ml samples were transported

to the laboratory on ice where they were serially diluted onto Yeast Extract, Peptone, Glucose with chloramphenicol. In total, 96 colonies were again recovered randomly from the resulting plates.

Molecular methods
In total, 2688 colonies were collected from both vineyard and ferment samples. To discriminate between these, the DNA from each purified colony was extracted using the Chelex method (see Serjeant *et al.*, 2008). The genomic internal transcribed spacer (ITS) region was amplified and subjected to restriction fragment length polymorphism (RFLP) analyses, which is a good method to differentiate between species (Esteve-Zarzoso *et al.*, 1999), and we collated the ITS RFLP patterns into cohorts (see Supplementary Material). To assess the confidence between, and identity of, the different cohorts, we determined the sequence of the diagnostic divergent domains 1 and 2 of the 26S nuclear ribosomal gene (D1/D2 26S) for one to three examples from each cohort (some cohorts only had a single representative; see Supplementary Material). The D1/D2 26S region is commonly used to discriminate between yeast species, and empirical analyses suggests that isolates with <98% homology in this region are likely different species (Kurtzman and Robnett, 2003). We reconstructed a bootstrap consensus tree to examine 26S sequence correlation with ITS-RFLP cohorts using Geneious (Drummond *et al.*, 2011). We also compared the 26S sequences with deposits for typical examples for known yeast species in the National Centre for Biotechnology Information database, allowing us to ascribe species names with some degree of confidence. The allele lengths of isolates identified as *S. cerevisiae* from ferment samples were analysed at 10 unlinked variable microsatellite repeat loci as described by Richards *et al.* (2009). The particular combinations of alleles at these loci allow $\sim 10^9$ strains to be differentiated, but more importantly the relatedness between individuals and specific sub-populations to be inferred (Richards *et al.*, 2009; Goddard *et al.*, 2010).

Data analyses
Differences in community composition between regions were tested with an analysis of similarity (ANOSIM) on square root transformed species incidences with PAST, which is freely available (Hammer *et al.*, 2001). An ANOSIM test makes no assumption about the normality of data and this multivariate test is classically used in community ecology and reports the probability of observing the differences in community composition between regions by chance using permutations of a rank Bray–Curtis similarity matrix to create null distributions (Clarke, 1993). The observed rank abundance difference (R) between regions is reported, and this value ranges from −1 to +1: values of R around zero indicate that there are no differences between regions and values above zero indicate that communities differ between regions. We generated null distributions that recalculated R under a framework that assumed no difference between regions by randomising the labels associated with samples across the entire data set a total of one million independent times.

Routine population genetic analyses of the *S. cerevisiae* genotypes, including elucidating matching genotypes and the calculation of genetic distance and various F statistics, were conducted using 'Genetic Analyses in Excel' (Peakall and Smouse, 2006). Outcrossing rates were estimated using Mathematica 7 (code available on request). Allelic richness was estimated using HP-rare (Kalinowski, 2005). Tests for population structure, and assignments to likely sub-populations, employed Bayesian inference as implemented in InStruct (Gao *et al.*, 2007) using a model that incorporated admixture. Five chains of two million Markov Chain Monte Carlo iterations with a burn-in of 100 000 were run for $K = 1$ to 20.

Results

We tested the null hypothesis of no difference in vine- and wine-associated yeasts in different regions of NZ with two approaches: (1) the communities associated with vineyards; and, (2) populations of *S. cerevisiae* conducting spontaneous ferments.

Community comparisons—vineyard samples
In total, 22 distinct ITS-RFLP cohorts were delineated among the 2304 isolates deriving from 24 vineyards from West Auckland, Waiheke Island and Hawke's Bay (see Figure 1, and Supplementary Material). The 26S sequence data allowed species assignment based on homology, and the assignment of random examples within ITS-RFLP cohorts were all to the same species with 100% bootstrap support apart from one cohort, which contained matches to both *Sporobolomyces oryzicola* and *Cryptococcus macerans* (#18; see Figure 1). Isolates with 26S sequences matching to *Cryptococcus festucosus* and *C. macerans* derived from cohorts with different RFLP patterns (#4 and #22, and #12, #15 and #18, respectively; see Figure 1) indicating reasonable levels of within species diversity at the ITS region: the respective mean identities at the 26S region of 96% and 90% for these two groupings is below the 98% commonly considered the homology boundary between yeast species (Kurtzman and Robnett, 2003). Thus, these cohorts were considered separately in further analyses. Two other cohorts converged on the same species (#5 and #9 to *Hanseniaspora uvarum,* and #10 and 20 to *H. occidentalis*), but as the average 26S homology for these isolates was >99%, these were comfortably considered the same species and thus

Figure 1 Correlation between ITS-RFLP patterns and D1/D2 26s RNA sequence data. Neighbour-joining consensus tree of 1000 bootstrap replicates reconstructed from an alignment of the D1/D2 26s RNA region for 45 isolates associated with Chardonnay and Syrah fruit (see Supplementary Material). Random isolates from each of the 22 ITS-RFLP cohorts were selected. Isolates from the same cohort whose 26S sequences match to the same type species and are placed together with 100% support are shown in red. Isolates from different ITS-RFLP cohorts whose 26S sequences match to the same species in GenBank and are placed together with 100% support are shown in blue. The two examples from cohort 18 that yielded different 26s RNA sequences are in green. The >98% homology of our query sequences to type-strain deposits in GenBank allow species names to be ascribed to well-supported groups and the abundance of each group among the 2688 isolates is shown. Those species previously reported as associated with juice/wines in NZ are denoted with a superscript.

the cohorts were combined in further analyses. The 20 different final groupings, and their homology to reference sequences at the D1/D2 26S ribosomal DNA region, are shown in Figure 1 and have the National Centre for Biotechnology Information accession numbers JN377440–JN377484 (and see Supplementary Material).

Approximately half of the species identified belong to the Basidiomycota, and the remainder to the Ascomycota. *Rhodotorula glutinis* and *Aureobasidium pullulans* combined made up 75.6% of isolates, so are by far the most common species. Following these two species, there is a tail of rarer species, as is common in ecological sampling (Pedros-Alio, 2007). The most abundant budding yeasts are *Hanseniaspora uvarum* (7.1%) followed by *Candida zemplinina* (5.1%)—these species are commonly reported associated with wine grapes and ferments from other winemaking countries (Cadez *et al.*, 2010; Ciani *et al.*, 2010). Only four of

the species found here have previously been reported as associated with grape juice in NZ (Goddard, 2008; Zhang *et al.*, 2010a; see Figure 1).

Our preliminary analyses showed that the communities associated with Chardonnay and Syrah differed significantly ($R = 0.27$; $P = 0.013$; ANOSIM) regardless of region. The subject of test here although is whether communities differ between regions, and the primary and balanced data set gathered to test this derived from Chardonnay: Table 1 shows the regional distributions. West Auckland and Hawke's Bay both have *Rhodotorula glutinis* as the dominant species, while *Aureobasidium pullulans* is dominant on Waiheke Island. The number of species also differs between regions with West Auckland containing the lowest species diversity and Waiheke Island the greatest. The Shannon and Simpson indices measure diversity and 'evenness' and these metrics suggest that the West Auckland community is the most skewed toward a

dominant species. ANOSIMs report positive R-values for the three possible pair-wise region comparisons, suggesting that each region indeed harbours different communities. Comparison of the observed R-values to null distributions show the communities associated with Chardonnay differ significantly between regions with Waiheke Island having the most distinct community. The West Auckland and Hawke's Bay communities differed moderately ($R = 0.213$; $P = 0.017$). Both comparisons of West Auckland and Hawke's Bay with Waiheke Island produced greater R-values, indicating the communities in these areas differed greatly ($R = 0.66$ and 0.71, respectively; both $P = 0.002$).

S. cerevisiae population structure—ferment samples
We ascertained the microsatellite profiles of 353 individual *S. cerevisiae* isolates from 16 Chardonnay and five Syrah ferments across the three regions. A total of 274 different genotypes were revealed, indicating a highly diverse population of *S. cerevisiae*. Of the 216 alleles we identified, 125 were novel compared with a previous NZ population examined from Auckland ferment and environmental samples

Table 1 Species identities, based on D1/D2 26S region sequence homology to deposits described as 'type strains' in NCBI, and abundances based on ITS RFLP, of isolates from ripe Chardonnay fruit by region

Species	West Auckland	Hawkes Bay	Waiheke Island
Rhodotorula glutinis	67.0%	61.7%	13.8%
Aureobasidium pullulans	26.2%	22.1%	64.2%
Hanseniaspora uvarum	0%	5.7%	17%
Cryptococcus magnus	6.6%	7.0%	0.7%
Candida zemplinina	0%	0.9%	1.8%
Cryptococcus festucosus #1	0%	2.7%	0%
Cryptococcus macerans #1	0%	0%	1.1%
Pichia sporocuriosa	0%	0%	0.9%
Cryptococcus festucosus #2	0.2%	0%	0.5%
Number of species	4	6	8
Number of individuals	564	561	441
Shannon index	0.8092	1.119	1.084
Simpson index	0.4776	0.5619	0.5395

Abbreviations: ITS, internal transcribed spacer; RFLP, restriction fragment length polymorphism.

(Goddard *et al.*, 2010); however, all the genotypes discovered here were distinct. We compared these genotypes with our database of 79 commercial wine strains commonly used in NZ (Richards *et al.*, 2009), and only one of the 82 isolates deriving from the Hawke's Bay Chardonnay ferments matched. Six genotypes differed at just one allele from a commercial strain, but the remaining genotypes differed at many alleles; thus, these data further support the concept that a diverse natural population of *S. cerevisiae* resides in NZ (Goddard *et al.*, 2010). A total of 303 different genotypes were used in further analyses in order to accurately reflect genotype presence–absence across ferments (some genotypes were present in more than one ferment); only a single representative of each genotype from each ferment was included in analyses to remove the effect of within-ferment clonal expansion of genotypes.

When populations are broken down by region and variety, the greatest pairwise F_{st} is between West Auckland Chardonnay and Waiheke Chardonnay (0.12), and the least between Hawke's Bay Chardonnay and Syrah (0.02); the pairwise F_{st} between genotypes when simply grouped by variety (Chardonnay vs Syrah) is only 0.025 indicating very little signal for population delineation by variety, regardless of region. However, the mean pairwise F_{st} is more than doubled to 0.055 when genotypes are instead partitioned by region of isolation, indicating moderate genetic differentiation. Together, this suggests geographic origin accounts for a greater portion of signal for population differentiation than variety in these data (Hartl and Clark, 1997). When the data are partitioned by region, then the largest pairwise difference is between Waiheke Island and West Auckland populations ($F_{st} = 0.084$), which are only ~40 km apart, followed by West Auckland—Hawke's Bay ($F_{st} = 0.056$) and Waiheke Island—Hawke's Bay populations ($F_{st} = 0.025$), which are both ~350 km apart, suggesting isolation by distance *per se* is not a major descriptor of these data. Although some of the genotypes were found in multiple ferments, these were mostly from ferments within the same region and ~94% of genotypes are region specific (see Table 2).

All 10 loci are significantly out of Hardy–Weinberg equilibrium ($P < 0.001$), and this correlates with

Table 2 Population genetic summary for spontaneous NZ Chardonnay and Syrah ferment *S. cerevisiae* genotypes from different regions

Region (n)	Mean allelic richness[a]	F_{is}[b]	Outcrossing rate[c]	Region-specific genotypes (%)[d]
West Auckland (82)	7 ± 1.5	0.73 ± 0.19	0.12–0.16–0.20	100
Waiheke Island (96)	7.6 ± 2.6	0.61 ± 0.20	0.08–0.11–0.14	90
Hawkes Bay (124)	9.9 ± 3.3	0.67 ± 0.12	0.12–0.15–0.19	92

Abbreviations: *n*, the number of genotypes; NZ, New Zealand.
[a]Mean allelic richness was calculated for each locus independently by rarefaction using HP-rare and the mean and s.d. across all loci are reported.
[b]The mean inbreeding coefficients across all ten loci (± s.d.); a value of 1 indicates a completely inbred population.
[c]Average proportion of the mating of spores from independent tetrads per meiotic generation surrounded by 2 log likelihood support limits, which are equivalent to the 95% confidence intervals, calculated following Johnson *et al.* (2004).
[d]Percentage of genotypes found in that region only.

previous reports from natural *S. cerevisiae* populations from South America, Europe and NZ (Schuller *et al.*, 2005; Cubillos *et al.*, 2009; Goddard *et al.*, 2010), and is expected given the life history of *S. cerevisiae* predisposes it to inbreed (Knop, 2006). However, that these populations have inbreeding coefficients (F_{is}) well below 1 (mean 0.67, Table 2) indicates that they are sexually outcrossed (that is, engage in mating between spores from independent parents) to some extent. The inbreeding system that *S. cerevisiae* displays means it is not appropriate to use Wright's inbreeding coefficient as classically calculated to estimate the rates of inbreeding vs outcrossing: this must be modified as matings between the four related spores from the same tetrad lead to a reduction in heterozygosity of one-third per generation, and not one-half as is the case for fusion of gametes from independent meiosis (Falconer and Mackay, 1996; Goddard *et al.*, 2001). Following Johnson *et al.* (2004), we estimated the most likely outcrossing rate with the more appropriate model (Table 2), and this averages 14% outcrossed matings per meiotic generation among all the populations, and strongly suggests that a reasonable extent of genetic mixing is occurring within this population.

The fact that most genotypes were recovered from multiple different ferments within but not between regions (only 10 isolates overlapped between Waiheke and Hawke's Bay) lends support to the idea that there are region-specific sub-populations, but it does not prove it. For example, in outcrossed metazoans populations each individual has a unique genotype (barring identical twins), and thus each genotype is by definition unique to an area: this does not prove that there are region-specific populations. It is more appropriate to test hypotheses concerning the degree of relatedness of individuals within and between regions. If there were no average difference in relatedness between *S. cerevisiae* among different regions, then this population would appear homogenous with no sub-structure: genotypes from different regions would be equally related. However, if different regions contain distinct sub-populations of more closely related individuals then the population would appear structured. The usual methods to infer population structure are sensitive to inbreeding (Falush *et al.*, 2003; Huelsenbeck and Andolfatto, 2007) and thus we employed a Bayesian approach, which accounts for both admixture (we have shown this population is outcrossed to some extent) and *S. cerevisiae*'s tendency for inbreeding, to infer whether population structure exists, and then estimate the relatedness of genotypes among different sub-populations (Gao *et al.*, 2007). Importantly, this approach does not use *a priori* information on the origin of genotypes to detect structure. These analyses unequivocally allow us to reject the null hypothesis of a homogeneous population ($P < 0.0001$): the presence of 17 sub-populations, each containing closely related individuals, is inferred. Multiple Markov Chain Monte Carlo chains (5 runs) converged on identical inferences for the number of sub-populations.

As the optimal inferred number of sub-populations is 17, this means that population sub-division is not as simple as a direct correlation with region (or variety), where only 2–3 sub-populations would be inferred; thus the pattern of population sub-structure is more complicated. There might be any number of reasons why these genotypes are structured into sub-populations other than it being driven by geographic origin, and so having established that population structure exists, the next question is whether any of these sub-populations correlate with geographic location or not. If region-specific populations existed, it would mean that genotypes from the same area would be more closely related to one another than to genotypes from other regions. Is there any evidence for this? The null expectation, if geographic origin has no bearing on population sub-structure, is that isolates from different regions would be equally related within the various sub-populations. We analysed relatedness (proportion of shared ancestry) between genotypes within each sub-population from the different regions and analysis of variances (on arc-sined transformed ancestry proportions) show that while some sub-populations contain genotypes that are equally related between regions, relatedness among other sub-populations significantly differs between regions. The probability of isolates from each region being equally related, and their mean relatedness among the 17 inferred sub-populations are shown in Figure 2. It is clear that the isolates are not assigned to sub-populations randomly with respect to region.
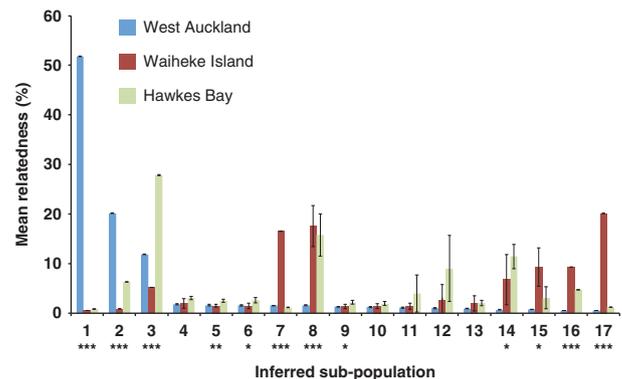
Figure 2 The average relatedness of 303 *S. cerevisiae* genotypes isolated from 21 spontaneous ferments among the 17 inferred sub-populations, partitioned by region. The proportion of ancestry of each genotype was inferred from five Markov Chain Monte Carlo chains with $K = 17$ using InStruct; the grand mean for ancestry of genotypes for each region in each sub-population was calculated from the average of these five runs and is plotted, along with the s.d. Blue bars: genotypes from West Auckland; red bars: Waiheke Island; green bars: Hawkes Bay. The sum of any one isolate's ancestry across all sub-populations is one. Deviation in relatedness between regions within each sub-population was tested with a one-way analysis of variance, and those sub-populations with a significant difference in relatedness between regions are indicated (*$P < 0.01$; **$P < 0.001$; ***$P < 0.0001$).

The between run variance for population assignment and relatedness in each subpopulation for each genotype was generally very small, with the average s.d. for relatedness being just ~1%; a few sub-populations have a larger variance, suggesting a less certain relatedness for these genotypes. There are clear examples of region-specific sub-populations. For example, 55% of the West Auckland genotypes are more closely related to one another than to any genotypes from other regions ($P < 0.001$ by an honestly significant difference test) and these are all classified into sub-population 1. Similarly sub-populations 7 and 17 are made up of closely related genotypes that all come from Waiheke and are more closely related to one another than genotypes from other regions ($P < 0.001$ by a honestly significant difference test). The case is less clear-cut for Hawke's Bay genotypes: isolates that have significantly greater relatedness within sub-population 3 all come from the Hawke's Bay, but there is also a reasonable number of isolates from the other regions in this group, but these have lower relatedness. The interpretation of these data is that groups of highly related genotypes exist, and some of these are strongly described by the region of isolation, but other groups of closely related genotypes comprise strains from different regions. Sixty-three percent of genotypes have at least 90% of the their ancestry inferred to be in one sub-population only, and thus approximately one-third of genotypes have ancestry among multiple sub-populations and accordingly many of these genotypes are hybrids (chimeras) with parents (or grandparents and so on) from different sub-populations. This inference correlates with the earlier suggestion of reasonable levels of outcrossed matings and thus genetic mixing.

Discussion

To our knowledge, there have been no previous reports of the diversity of microbes directly associated with fruit in NZ, and the list of genera, including those that are dominant, appears similar to reports from vineyard microbial communities in Europe (Cadez *et al.*, 2010) and Australia (Prakitchaiwattana *et al.*, 2004) with the exceptions of *Bullera* (interestingly first isolated in NZ from *Melicytus macrophyllus* (Hamamoto and Nakase, 1996)) and *Tetracladium*. All the vineyards sampled are commercially managed, which includes the application of anti-fungal sprays, which does not seem to have a significant effect on yeast communities that contribute to ferments at least (Cadez *et al.*, 2010). Close to 300 different genotypes of *S. cerevisiae* were recovered from the spontaneous ferments; while our sampling effort was large, it seems that we have only scratched the surface of the diversity of *S. cerevisiae* in NZ, and this is in line with a large *S. cerevisiae* diversity recovered from two spontaneous Chardonnay ferments in previous years (Goddard *et al.*, 2010).

This study's principal aim was to test for geographic differences in microbial communities and populations. The community composition tests we conducted allow us to categorically reject the null hypothesis of no difference in microbial communities on ripe Chardonnay fruit between West Auckland, Waiheke Island and Hawke's Bay. We are not aware of any previous study to objectively show regional community differences of microbes associated with fruit. We are aware of just one previous study attempting to test for regional geographic delineations of *S. cerevisiae* populations: this found no clear correlation between 64 *S. cerevisiae* from South America and Europe based on the clustering of RAPD profiles, but this conclusion was based on observation only and was not objectively tested (Martinez *et al.*, 2007). Population genetic analyses here show that different regions of NZ contain different genetically distinct and highly related sub-populations of *S. cerevisiae* driving spontaneous ferments of Chardonnay and Syrah. However, some genotypes are equally related to genotypes isolated from alternate regions, and this strongly suggests the movement of strains. There are also many hybrid genotypes with ancestry in more than one sub-population. This pattern is similar to the global population structure of *S. cerevisiae* (Liti *et al.*, 2009), as well as other populations from NZ (Goddard *et al.*, 2010). A helpful analogy to interpret this population pattern might be with a human example: within any one region, there are individuals, which belong to families that have resided in that region for a considerable period of time and thus many individuals in that area will be highly related (share ancestors); however, other individuals will have migrated into to the region in the recent past, and share little or no ancestry with, and thus have low relatedness to, the indigenous residents. New migrants may mate with the local population to produce offspring that will thus have relatedness spread among different ancestral groups from different regions. Together, these data and tests allow us to accept the general alternate hypotheses that there are region-specific communities and populations of microbes associated with vines and wines, at least on the North Island of NZ: we believe this is the first demonstration of regional differences worldwide.

There are very few reported observations of changes in vine and wine communities and populations over time (Beltran *et al.*, 2002) and our data clearly do not allow us to make any statements about such changes. One obvious negative manifestation of microbes associated with fruit is the disease or rot some cause. If grapes are infected by organisms such as *Botrytis cinerea* (grey and noble rot) or *Erysiphe necator* (powdery mildew), then the damage caused will allow other microbial populations (including yeasts) to expand. Thus, we were careful to randomly select fruit that was visually rot-free to

minimise this influence on our inference of differences in microbial communities between regions. Healthy fruit was selected at random from replicate vineyards within each region, and all samples processed identically: with the implementation of the ANOSIM we feel this provides a comparable and fair estimate of the differences in communities associated with vines among these regions of NZ. One other potential criticism of this work is that it is culture based. There are no previous reports of direct nucleic acid based estimates for microbial communities associated with vines and wines (but we are pursuing this approach). Although we may have missed some proportion of these communities, because this would have been a systematic bias (the same isolation procedure was used throughout), the replicate estimates from regions allow us to make statistically meaningful comparisons of communities and thus our conclusions are valid.

The population genetic data suggest that there is a degree of migration of *S. cerevisiae* between regions within NZ. This may occur by any number of vectors, but humans, insects and birds seem likely candidates, and the two former have been previously implicated (Goddard *et al.*, 2010). The observation that a diverse natural population of *S. cerevisiae* contributed to these spontaneous ferments is another piece of evidence again showing that spontaneous ferments are not driven by 'escaped' commercial isolates (Blanco *et al.*, 2011), even in large commercial wineries that also run parallel inoculated ferments (as was the case here), but driven by a diversity of natural isolates.

We did not have enough time or resources to go on to test ideas as to why there are differences between regions, but we will look to test some of the main ideas around this in the future. One possibility suggests that dispersal is limiting between these regions, but this seems not to be supported by our data, at least for *S. cerevisiae*: maybe the Bass–Becking idea is correct, and it is local adaptation that determines the differences? Climate is one of the more obvious factors that might explain why certain species are more prevalent in different areas; other factors include the soil and substrate nutrient composition and availability. It is likely that a combination of factors will be responsible, but it will be of interest to ascertain the relative contributions of each.

Finally, it is widely accepted that the particular climatic and geographic/soil conditions within an area interact with different grape varieties differently and serve to make distinctive wines, and this underlies the concept of 'terroir' or a wine's sense of place (Goode, 2006). Previous data suggest that many of the *S. cerevisiae* in spontaneous NZ ferments derive from the local environment (Goddard *et al.*, 2010), but it is also possible that some populations may reside 'in the cellar'. Either way, we provide some evidence for regional delineations of natural *S. cerevisiae* populations and the use of region-specific yeasts to make spontaneously fermented wine is precisely in keeping with the *terroir* concept as these wines more faithfully reflect their sense of place. There is also strong scientific evidence to show that different yeast species and strains act on grape precursors differently to produce different types and amounts of metabolites (Romano *et al.*, 2003; Swiegers and Pretorius, 2005, 2007; Howell *et al.*, 2006; Anfang *et al.*, 2009) and that these may have significant effects on the flavour and aroma of the resulting wine (Lambrechts and Pretorius, 2000; Cordente *et al.*, 2007; Ciani *et al.*, 2010; King *et al.*, 2010). Thus, the logical extension of this is that if different regions harbour different communities and populations, as we have shown, that these may, in part, contribute to the distinctiveness of wines deriving from that area: that is, there is a microbial aspect to *terroir*. However, the extent to which different microbes contribute to the uniqueness of wine from different regions is not yet proven and the stability of these communities and populations over time remains to be tested. Finally, it is well known that fungi and plants have intimate associations (rhizomes being the obvious example). Another aspect of a microbial influence on *terroir* worth considering is the degree to which microbes may manipulate vine and fruit development. It is possible that the different species in different areas modulate vine development differently such that different types or amounts of compounds, including precursors that yeasts act on during the ferment, are present and thus translate into altered sensory properties of wine. This also remains to be tested.

Acknowledgements

References

Anfang N, Brajkovich M, Goddard MR. (2009). Co-fermentation with *Pichia kluyveri* increases varietal thiol concentrations in Sauvignon Blanc. *Aust J Grape Wine R* **15**: 1–8.

Beltran G, Torija MJ, Novo M, Ferrer N, Poblet M, Guillamon JM *et al.* (2002). Analysis of yeast populations during alcoholic fermentation: a six year follow-up study. *Syst Appl Microbiol* **25**: 287–293.

Blanco P, Orriols I, Losada A. (2011). Survival of commercial yeasts in the winery environment and their prevalence during spontaneous fermentations. *J Ind Microbiol Biot* **38**: 235–239.

Burke D, Dawson D, Stearns T. (2000). *Methods in Yeast Genetics. A Cold Spring Harbor Laboratory Course Manual.* Cold Spring Harbor Press: Cold Spring Harbor.

Cadez N, Zupan J, Raspor P. (2010). The effect of fungicides on yeast communities associated with grape berries. *FEMS Yeast Res* **10**: 619–630.

Chambers PJ, Pretorius IS. (2010). Fermenting knowledge: the history of winemaking, science and yeast research. *EMBO Rep* **11**: 914–920.

Ciani M, Comitini F, Mannazzu I, Domizio P. (2010). Controlled mixed culture fermentation: a new perspective on the use of non-*Saccharomyces* yeasts in winemaking. *FEMS Yeast Res* **10**: 123–133.

Clarke KR. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* **18**: 117–143.

Cordente AG, Swiegers JH, Hegardt FG, Pretorius IS. (2007). Modulating aroma compounds during wine fermentation by manipulating carnitine acetyltransferases in *Saccharomyces cerevisiae*. *FEMS Microbiol Lett* **267**: 159–166.

Cordero-Bueso G, Arroyo T, Serrano A, Tello J, Aporta I, Velez MD *et al.* (2011). Influence of the farming system and vine variety on yeast communities associated with grape berries. *Intl J Food Micro* **145**: 132–139.

Cox CB, Moore PD. (1993). *Biogeography, an Ecological and Evolutionary Approach.* Blackwell: London.

Cubillos FA, Vasquez C, Faugeron S, Ganga A, Martinez C. (2009). Self-fertilization is the main sexual reproduction mechanism in native wine yeast populations. *FEMS Microbiol Ecol* **67**: 162–170.

Drummond A, Ashton B, Cheung M, Heled J, Kearse M, Moir R *et al.* (2011). Geneious v5.4, Available from http://www.geneious.com/.

Dujon B. (1996). The yeast genome project: what did we learn? *Trends Genet* **12**: 263–270.

Esteve-Zarzoso B, Belloch C, Uruburu F, Querol A. (1999). Identification of yeasts by RFLP analysis of the 5.8S rRNA gene and the two ribosomal internal transcribed spacers. *Int J Syst Bacteriol* **49**: 329–337.

Ezov TK, Boger-Nadjar E, Frenkel Z, Katsperovski I, Kemeny S, Nevo E *et al.* (2006). Molecular-genetic biodiversity in a natural population of the yeast *Saccharomyces cerevisiae* from 'Evolution canyon': microsatellite polymorphism, ploidy and controversial sexual status. *Genetics* **174**: 1455–1468.

Falconer DS, Mackay TFC. (1996). *Introduction to Quantitative Genetics*, 4th edn. Pearson, Prentice Hall: New Jersey.

Falush D, Stephens M, Pritchard JK. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Fay JC, Benavides JA. (2005). Evidence for domesticated and wild populations of *Sacchoromyces cerevisiae*. *PLOS Genetics* **1**: 66–71.

Finlay BJ, Fenchel T. (2004). Cosmopolitan metapopulations of free-living microbial eukaryotes. *Protist* **155**: 237–244.

Gao H, Williamson S, Bustamante CD. (2007). A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* **176**: 1635–1651.

Gaston KJ. (2000). Global patterns in biodiversity. *Nature* **405**: 220–227.

Goddard MR. (2008). Quantifying the complexities of *Saccharomyces cerevisiae*'s ecosystem engineering via fermentation. *Ecology* **89**: 2077–2082.

Goddard MR, Anfang N, Tang R, Gardner RC, Jun C. (2010). A distinct population of *Saccharomyces cerevisiae* in New Zealand: evidence for local dispersal by insects and human-aided global dispersal in oak barrels. *Environ Microbiol* **12**: 63–73.

Goddard MR, Greig D, Burt A. (2001). Outcrossed sex allows a selfish gene to invade yeast populations. *Proc R Soc Lond B-Biol Sci* **268**: 2537–2542.

Goode J. (2006). *The Science of Wine.* University of California Press: Berkeley.

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M *et al.* (2004). Spatial scaling of microbial eukaryote diversity. *Nature* **432**: 747–750.

Hamamoto M, Nakase T. (1996). Ballistosporous yeasts found on the surface of plant materials collected in New Zealand—the genera *Bensingtonia* and *Bullera* with descriptions of five new species. *Antonie Van Leeuwenhoek Int J Gen Mol Microbiol* **69**: 279–291.

Hammer O, Harper DAT, Ryan PD. (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica 4* art. **4**: 9 pp. (online only http://palaeo-electronica.org/2001_1/past/issue1_01.htm).

Hartl DL, Clark AG. (1997). *Principles of Population Genetics*, 3rd edn. Sinauer: Sunderland.

Howell KS, Cozzolino D, Bartowsky EJ, Fleet GH, Henschke PA. (2006). Metabolic profiling as a tool for revealing *Saccharomyces* interactions during wine fermentation. *FEMS Yeast Res* **6**: 91–101.

Huelsenbeck JP, Andolfatto P. (2007). Inference of population structure under a Dirichlet process model. *Genetics* **175**: 1787–1802.

Hugenholtz P, Goebel BM, Pace NR. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**: 4765–4774.

Johnson H. (1998). *The Story of Wine.* Simon and Schuster: London.

Johnson LJ, Koufopanou V, Goddard MR, Hetherington R, Schafer SM, Burt A. (2004). Population genetics of the wild yeast *Saccharomyces paradoxus*. *Genetics* **166**: 43–52.

Kalinowski ST. (2005). HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* **5**: 187–189.

King ES, Kievit RL, Curtin C, Swiegers JH, Pretorius IS, Bastian SEP *et al.* (2010). The effect of multiple yeasts co-inoculations on Sauvignon Blanc wine aroma composition, sensory properties and consumer preference. *Food Chem* **122**: 618–626.

King M. (2003). *History of New Zealand.* Penguin: Auckland.

Knop M. (2006). Evolution of the hemiascomycete yeasts: on life styles and the importance of inbreeding. *Bioessays* **28**: 696–708.

Koufopanou V, Hughes J, Bell G, Burt A. (2006). The spatial scale of genetic differentiation in a model organism: the wild yeast *Saccharomyces paradoxus*. *Philos T R Soc B* **361**: 1941–1946.

Kurtzman CP, Robnett CJ. (2003). Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. *FEMS Yeast Res* **3**: 417–432.

Lachance MA, Lawrie D, Dobson J, Piggott J. (2008). Biogeography and population structure of the

Neotropical endemic yeast species *Metschnikowia lochheadii*. *Antonie Van Leeuwenhoek Int J Gen Mol Microbiol* **94**: 403–414.

Lachance MA, Starmer WT, Bowles JM, Phaff HJ, Rosa CA. (2000). Ribosomal DNA, species structure, and biogeography of the cactophilic yeast *Clavispora opuntiae*. *Can J Microbiol* **46**: 195–210.

Lachance MA, Starmer WT, Rosa CA, Bowles JM, Barker JSF, Janzen DH. (2001). Biogeography of the yeasts of ephemeral flowers and their insects. *FEMS Yeast Res* **1**: 1–8.

Lambrechts MG, Pretorius IS. (2000). Yeast and its importance to wine aroma—a review. *S Afr J Enol Vitic* **21**: 97–129.

Landry CR, Townsend JP, Hartl DL, Cavalieri D. (2006). Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. *Mol Ecol* **15**: 575–591.

Legras JL, Merdinoglu D, Cornuet JM, Karst F. (2007). Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Mol Ecol* **16**: 2091–2102.

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA *et al.* (2009). Population genomics of domestic and wild yeasts. *Nature* **458**: 337–341.

Martinez C, Cosgaya P, Vasquez C, Gac S, Ganga A. (2007). High degree of correlation between molecular polymorphism and geographic origin of wine yeast strains. *J Appl Microbiol* **103**: 2185–2195.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al.* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.

McGovern PE, Zhang J, Tang J, Zhang Z, Hall GR, Moreau RA *et al.* (2004). Fermented beverages of pre- and proto-historic China. *Proc Natl Acad Sci USA* **101**: 17593–17598.

Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK *et al.* (2011). Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13**: 135–144.

Peakall R, Smouse PE. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* **6**: 288–295.

Pedros-Alio C. (2007). Dipping into the rare biosphere. *Science* **315**: 192–193.

Piskur J, Rozpedowska E, Polakova S, Merico A, Compagno C. (2006). How did *Saccharomyces* evolve to become a good brewer? *Trends Genet* **22**: 183–186.

Prakitchaiwattana CJ, Fleet GH, Heard GM. (2004). Application and evaluation of denaturing gradient gel electrophoresis to analyse the yeast ecology of wine grapes. *FEMS Yeast Res* **4**: 865–877.

Pretorius IS. (2000). Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast* **16**: 675–729.

Replansky T, Koufopanou V, Greig D, Bell G. (2008). *Saccharomyces sensu stricto* as a model system for evolution and ecology. *Trends Ecol Evol* **23**: 494–501.

Ribereau-Gayon P, Dobourdieu D, Doneche B, Lonvaud A. (2006). *Handbook of Enology*, Vol. 1 The Microbiology of Wine and Vinifications. John Wiley & Sons, Ltd: Chichetser, UK.

Richards KD, Goddard MR, Gardner RC. (2009). A database of microsatellite genotypes for *Saccharomyces cerevisiae*. *Antonie Van Leeuwenhoek Int J Gen Mol Microbiol* **96**: 355–359.

Romano P, Fiore C, Paraggio M, Caruso M, Capece A. (2003). Function of yeast species and strains in wine flavour. *Intl J Food Micro* **86**: 169–180.

Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L. (2009). Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**: 342–345.

Schuller D, Alves H, Dequin S, Casal M. (2005). Ecological survey of *Saccharomyces cerevisiae* strains from vineyards in the Vinho Verde Region of Portugal. *FEMS Microbiol Ecol* **51**: 167–177.

Schuller D, Casal M. (2007). The genetic structure of fermentative vineyard-associated *Saccharomyces cerevisiae* populations revealed by microsatellite analysis. *Antonie Van Leeuwenhoek Int J Gen Mol Microbiol* **91**: 137–150.

Serjeant K, Tang R, Anfang N, Beggs JR, Goddard MR. (2008). Yeasts associated with the New Zealand *Nothofagus* honeydew system. *New Zeal J Ecol* **32**: 209–213.

Swiegers JH, Pretorius IS. (2005). Yeast modulation of wine flavor. *Adv Appl Microbiol* **57**: 131–175.

Swiegers JH, Pretorius IS. (2007). Modulation of volatile sulfur compounds by wine yeast. *Appl Microbiol Biotechnol* **74**: 954–960.

Thomson JM, Gaucher EA, Burgan MF, De Kee DW, Li T, Aris JP *et al.* (2005). Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat Genet* **37**: 630–635.

Whitaker RJ, Grogan DW, Taylor JW. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**: 976–978.

Wolfe KH, Shields DC. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**: 708–713.

Zhang HY, Lee SA, Bradbury JE, Warren RN, Seth H, Hooks D *et al.* (2010a). Yeasts isolated from New Zealand vineyards and wineries. *Aust J Grape Wine R* **16**: 491–496.

Zhang HY, Skelton A, Gardner RC, Goddard MR. (2010b). *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* reside on oak trees in New Zealand: evidence for migration from Europe and interspecies hybrids. *FEMS Yeast Res* **10**: 941–947.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)