Pyrosequencing reveals regional differences in fruit-associated fungal communities

Michael W. Taylor,[1] Peter Tsai,[2] Nicole Anfang,[1]
Howard A. Ross[1,2] and Matthew R. Goddard[1]*
[1]*The School of Biological Sciences and*
[2]*Bioinformatics Institute, University of Auckland,
Auckland, New Zealand.*

Summary

We know relatively little of the distribution of microbial communities generally. Significant work has examined a range of bacterial communities, but the distribution of microbial eukaryotes is less well characterized. Humans have an ancient association with grape vines (*Vitis vinifera*) and have been making wine since the dawn of civilization, and fungi drive this natural process. While the molecular biology of certain fungi naturally associated with vines and wines is well characterized, complementary investigations into the ecology of fungi associated with fruiting plants is largely lacking. DNA sequencing technologies allow the direct estimation of microbial diversity from a given sample, avoiding culture-based biases. Here, we use deep community pyrosequencing approaches, targeted at the 26S rRNA gene, to examine the richness and composition of fungal communities associated with grapevines and test for geographical community structure among four major regions in New Zealand (NZ). We find over 200 taxa using this approach, which is 10-fold more than previously recovered using culture-based methods. Our analyses allow us to reject the null hypothesis of homogeneity in fungal species richness and community composition across NZ and reveal significant differences between major areas.

Introduction

Microbes are key ecosystem drivers, and understanding how microbial assemblages differ in space and time is

imperative for an integrated understanding of ecosystem function (Hodge *et al.*, 2001; Loreau, 2001; van der Heijden *et al.*, 2008). Our first task is to describe the distribution of microbial species in space but because of the technical challenges involved (Amann *et al.*, 1995), the biogeography of microbial communities is relatively poorly understood compared with metazoans (Gaston, 2000; Hughes Martiny *et al.*, 2006). Next-generation DNA sequencing advances, which circumvent a culture-based bias, aim to describe microbial communities by employing 'deep community sequencing' of DNA directly extracted from samples, and studies employing these approaches have made significant advances (Hugenholtz *et al.*, 1998; Tyson *et al.*, 2004; von Mering *et al.*, 2007). The emerging picture of microbial biogeography is a complicated one, with some evidence for both unstructured as well as spatially structured communities. Most studies have focused on bacterial communities (Hughes Martiny *et al.*, 2006), but less attention has been paid to eukaryotic microbial communities (Bik *et al.*, 2012). One study examining a handful of fungal species suggests that distributions may range from being narrowly endemic to global (Taylor *et al.*, 2006). Significant work has also been conducted on the biogeography of aquatic protists (Finlay, 2002), rotifers (Nemergut *et al.*, 2011), soil fungi (Green *et al.*, 2004), and cacti- and flower-associated yeasts (Lachance *et al.*, 2000; 2001; 2008). However, in general, there are fewer studies examining the biogeography of eukaryotic microbes using deep community sequencing techniques, and those that have been conducted are largely focused on soils (e.g. Jumpponen *et al.*, 2010; Orgiazzi *et al.*, 2013).

Humans have been cultivating fruiting plants, and fermenting their produce, since at least the dawn of civilization, and thus, this pursuit is deeply engrained in our history (McGovern *et al.*, 2004; Landry *et al.*, 2006; Chambers and Pretorius, 2010). It is now known that fungi naturally associated with fruits affect both plant and fruit development (as some are pathogens), as well as drive the fermentation of fruits and dictate wine quality and style (Swiegers and Pretorius, 2005; Barata *et al.*, 2012). Much research has been undertaken to characterize the biochemistry and molecular biology of a number of fungal species, which are primarily from the Ascomycota phylum (Chambers and Pretorius, 2010). Indeed, *Saccharomyces*

sp. evolved into a classic fundamental research model for genetics, cell and molecular biology (Landry *et al.*, 2006) and more recently experimental ecology and evolution (Replansky *et al.*, 2008). However, it is surprising that the complementary work to decipher the ecology of fungi associated with fruits is comparatively lacking. Previous work has indicated that some tens of species are naturally present on grapes (Prakitchaiwattana *et al.*, 2004; Čadež *et al.*, 2010; Cordero-Bueso *et al.*, 2011; Zott *et al.*, 2011; Gayevskiy and Goddard, 2012). Recent comprehensive studies used automated ribosomal intergenic spacer analyses fingerprinting to examine the impact of farming systems on fungi associated with grapes (Pancher *et al.*, 2012; Setati *et al.*, 2012). The fungal and bacterial communities associated with the leaves of Tempranillo vines in Portugal have also been assessed using pyrosequencing, and this also revealed a large community diversity (Pinto *et al.*, 2014). However, we are aware of only two studies that have robustly tested whether fruit-associated fungal communities differ in space. Gayevskiy and Goddard (2012) provided evidence for geographical delineations of fungal communities associated with *Vitis vinifera* on the North Island of New Zealand (NZ) using a culture-based approach. Very recently, Bokulich and colleagues (2014) comprehensively examined the communities of both bacteria and fungi in crushed Chardonnay and Cabernet Sauvignon fruit in California using next-generation sequencing approaches and showed that these differ by region, but are also conditioned by climate, year and cultivar. The handful of culture independent studies reveal a larger vine-associated diversity than culture-based methods (Pancher *et al.*, 2012; Setati *et al.*, 2012; Bokulich *et al.*, 2014; Pinto *et al.*, 2014), unsurprisingly suggesting that many species of fungi that are of importance do not grow, or grow only poorly, on the standard laboratory media that are classically employed. While it is well reported that only a fraction of most environmental bacteria have been cultivated (Amann *et al.*, 1995), at least on standard media, we are not aware of any similar direct estimates for the culturability of eukaryotic microbes.

NZ was inhabited by humans only in the last ~700 years and has only had commercially planted vines in the last ~100 years. Here, we analyse replicate samples from 23 vineyards from four distinct regions throughout NZ to elucidate variance in species richness (species counts) and community composition (species abundances) of fruit-associated fungi using pyrosequencing. We describe the amplification and sequencing of the 26S rRNA gene from DNA directly extracted from fungal communities associated with ripe Chardonnay fruit. We then test for homogeneity in communities of fungi associated with vines in different areas of NZ using community ecology analyses.

Results

Species concepts and control populations

Species concepts are notoriously tricky, and with these data, we may only align with concepts that ascribe species boundaries according to genetic homology (De Queiroz, 2007). However, we lack robust estimates for within-species variance at the D1/D2 26S locus for fungi (and most eukaryotes). *Saccharomyces cerevisiae* is one species for which we probably have the most comprehensive set of sequences from different isolates within the same species. We obtained 224 26S rRNA sequences of ~600 bp labelled as *S. cerevisiae* from GenBank, and a multiple alignment of these shows an average pairwise identity of just 86.2%. That species identification is often based on the very sequence we are examining means a degree of circularity is in play; it may well be that some or many of these deposits are incorrectly assigned to this species. The genetic variance at the 26S rRNA regions for most other species of fungi is uncharacterized. However, the empirically determined 98% identity cut-off for the D1/D2 26S region between fungal species is largely based on only a few examples within each species, but these surveys were conducted across a large range of species (Kurtzman and Robnett, 1998; Romanelli *et al.*, 2010). Thus, in line with these estimates, but in acknowledgment of this issue, hereafter, we refer to 98% similarity clusters of sequences we derive as operational genetic units (OGUs). Assignments to taxonomic levels above species are more definitive as they were estimated by comparisons with a reference database using Bayesian approaches.

We included sequence data from two *S. cerevisiae* clonal control populations to assess the degree to which polymerase chain reaction (PCR) and sequencing errors might artificially inflate diversity given the bioinformatic pipeline we employed, and we recovered 1097 and 3526 reads from these two clonal control populations. Seven 98% OGUs were inferred across both populations; however, the distribution of reads was massively skewed. For both control populations, one 98% OGU contained > 99.6% of reads, and this OGU was assigned to *Saccharomyces* with our pipeline, and basic local alignment searching using the representative sequence matched 100% of the query submitted with 100% identity to multiple reliable deposits described as *S. cerevisiae*. Three of the remaining six OGUs, comprising just < 0.4% of control reads, were also assigned to *S. cerevisiae*. The remaining seven reads (comprising 0.0009% of control reads) were split across the remaining three OGUs – one of these was a singleton OGU, the other two were both assigned to the Saccharomycetaceae – the same family as *S. cerevisiae*. Given this near-clonal template, the net rate (accounting for all steps from PCR onwards) of incorrect species

assignment for any one read with our method is ~1 in 700, and the likelihood of incorrect class assignment is lower for this strain of *S. cerevisiae* at least. The inclusion of these control clonal populations along with our environmental samples suggest that PCR and sequencing errors in our system are unlikely to significantly inflate our estimates of diversity from the environmental samples.

Overall fungal diversity

Excluding the controls, we obtained a total of 95 501 post-processed good-quality reads assigned to the fungal kingdom. Five hundred and eighty-four fungal 98% OGUs were revealed, but 331 of these were represented by a single (good quality) read only. We conservatively removed the entirety of these 'singleton OGUs', which removed just 0.004% of the read data. Thus, the remaining 95 170 reads clustered into 253 OGUs. One hundred and thirty-one (52%) OGUs were classified to the Ascomycota, but this phylum comprised 78 665 (83%) of the reads. One hundred and seventeen (46%) OGUs were classified to Basidiomycota, but this phylum comprised only 16 485 (17%) of the reads. Only five OGUs, comprising just 20 reads, remained unclassified at the phylum level. While there was an approximately even split in genetic diversity (number of OGUs) between the two

major fungal phyla, the Ascomycota appear to dominate in this niche, if number of reads allows an estimate of abundance, see Fig. 1.

The frequency distribution of reads was approximately negatively exponentially distributed. Previous work examining short reads at the internal transcribed spacer (ITS) region of mock communities suggests that direct DNA analyses do a reasonable job of estimating differential abundances in fungal communities (Bokulich and Mills, 2013). One OGU dominated overall with 29 821 reads (or 31%), and this was assigned to the *Cladosporidium* genus but has an uncertain species assignment. One hundred and twenty-four of the 253 OGUs were assigned to a genus (the lowest level in the taxonomic reference database) with a high probability, and the remaining one third of reads were unclassified. However, 91% of OGUs, covering 99.99% of reads, were assigned to a class. Figure 1 shows the relationship between abundance (number of reads) and diversity (number of OGUs) at phylum and class level, as well as the breakdown of reads by order. While 53% of reads belonged to Dothideomycetes (the largest and most diverse class of ascomycete fungi), only 26% of OGUs belonged to this class, and there is a range of classes that are diverse but not necessarily abundant. It is of note that Capnodiales (plant epiphytic sooty-moulds) is the most abundant order, comprising 45% of
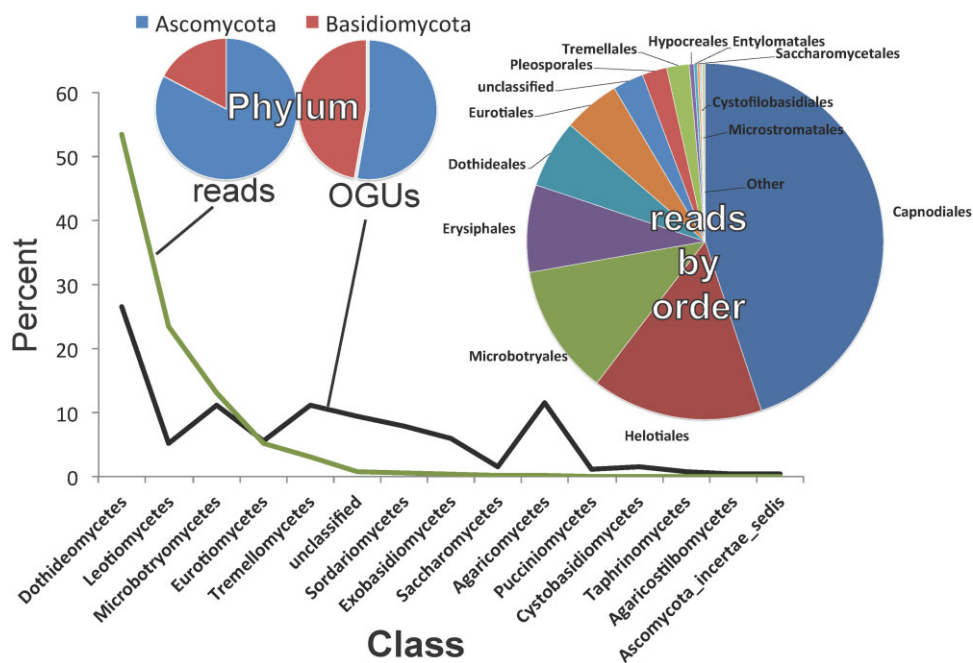
Fig. 1. Number of reads and OGUs at class and phylum level, and the number of reads at order level. The correlation between diversity (number of OGUs) and abundance (number of reads) for all classes is shown in the line plot and with two pie charts for phylum level. The distribution of reads at the order level is shown as a pie chart. Representative sequences that were not assigned to either a class or order given our Bayesian 70% bootstrap cut-off are designated unclassified (but will likely have been classified at higher levels, see Supporting Information Table S1).

Table 1. Fungal community metrics by region.

	West Auckland	Hawke's Bay	Marlborough	Central Otago
Number of vineyards	5	6	6	6
Total reads	23 075	26 958	21 879	23 258
Mean (± standard deviation) within region OGU richness per sample($\bar{\alpha}$)	51 ± 5 (20%)	56 ± 30 (22%)	48 ± 9 (19%)	26 ± 6 (10%)
Total OGUs (region-γ)	129 (51%)	164 (65%)	128 (51%)	74 (29%)
β_w (uses region-γ)	1.5	1.9	1.6	1.9
β_w (uses national-γ)	3.9	3.5	4.2	8.9
Relative species richness[a]	0.80	1	0.82	0.46
Region-specific OGUs[b]	29	48	9	19
Estimated richness[c]	160	172	155	92
	143–199	167–186	140–187	81–120
Shannon	1.7989	2.1507	2.004	1.6367
Simpson	0.6607	0.8166	0.7949	0.7617

a. Calculated by subsampling 20 000 reads per region, and the rarefied value presented is relative to Hawke's Bay species richness.
b. Number of OGUs found in one region only.
c. Estimated species richness calculated the Chao1 richness estimator, with 95% confidence bounds.
Percentage of the total 253 OGUs are reported in parentheses for region-γ. We use 'Whittaker's' beta diversity: $\beta_w = (\gamma/\bar{\alpha})$, where $\bar{\alpha}$ may either be mean α within region or mean α across all samples, to calculate β_w at the regional level (using the region-γ) and national levels, respectively.

reads; also, the mainly plant parasitic orders Erysiphales and Dothideales are fourth and fifth most dominant, with 8% and 6% of reads, respectively. Saccharomycetales (budding/true yeasts) are the 12th most abundant order, comprising just 0.2% of reads, and *Saccharomyces* spp., which drive the fermentation of fruits into wine (and are model organisms) comprise just five of the 95 170 reads or are present at ~1:20 000 in this community. A comprehensive breakdown of number of reads and OGUs at all taxonomic levels across all the data is provided in the Supporting Information Table S1.
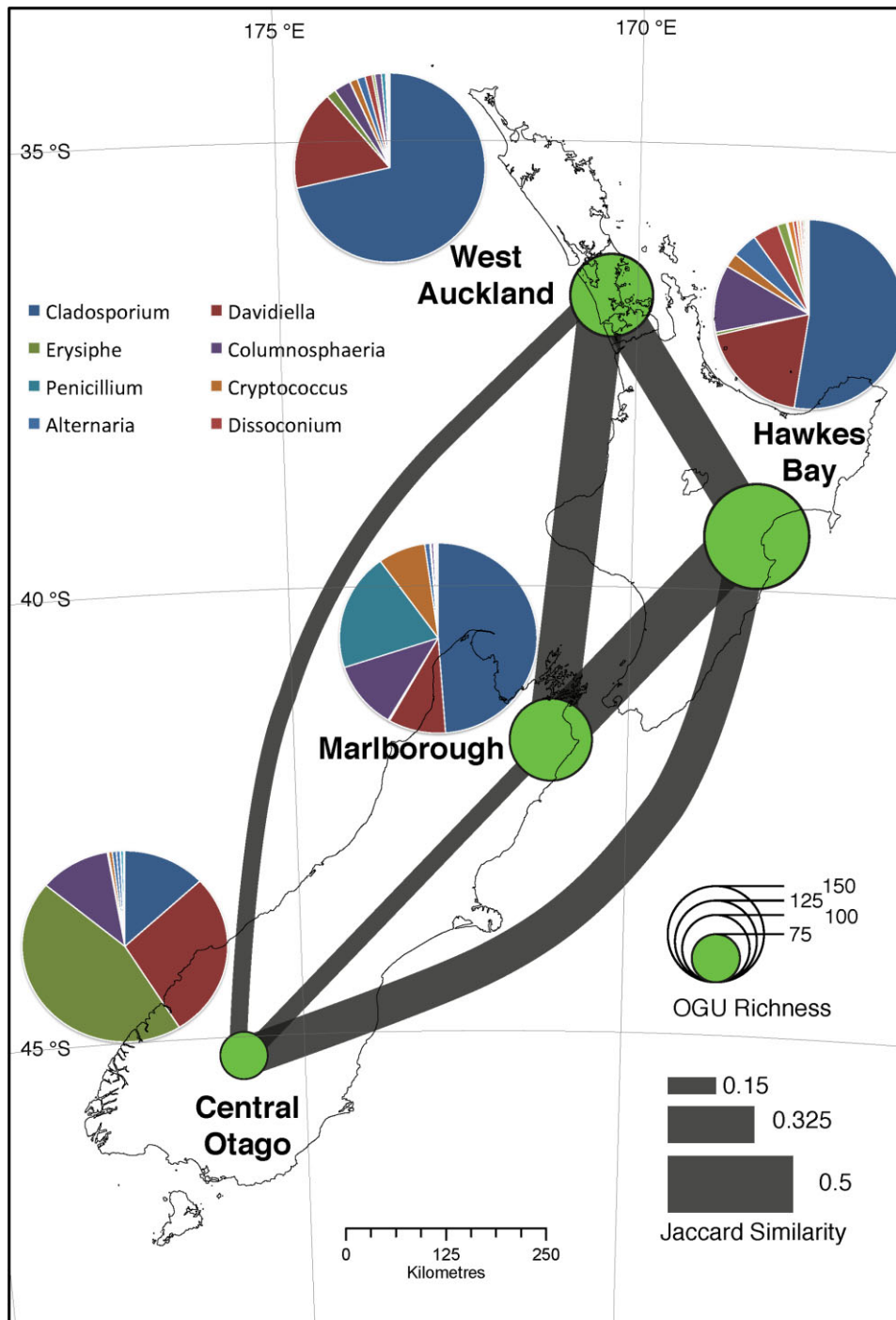
Tests for regional differences in communities

Regardless of whether or not we can assign names to OGUs, we may still compare diversity between regions. We obtained an average of 23 792 (standard deviation ± 1900) reads per region, with the observed 98% OGU richness (counts) highest in Hawke's Bay, followed by West Auckland and Marlborough, and then Central Otago (Table 1 and Fig. 2, and Supporting Information Fig. S1). Expected and relative OGU richness, estimated using rarefaction analyses, support the trend of Hawke's Bay being twice as rich as Central Otago, with West Auckland and Marlborough intermediate (Table 1). Simpson's and Shannon's indices indicate that the spread of diversity is relatively more even in Hawke's Bay and most skewed in West Auckland (Table 1). Just five OGUs are present in all 23 samples across regions (four Ascomycete : Dothideomycetes and one Basidiomycete : Tremellomycetes), and only 25 OGUs are present in at least one sample in all four regions. Between 4–18% of OGUs appear in one region only (Table 1). It is of note that Central Otago has the greatest ratio of unique OGUs to

richness, suggesting that while this region is relatively OGU poor, it might also harbour a differentiated community.

The earlier observations suggest that there are differences in OGU richness among regions, and next, we employed tests of null models to evaluate this. First, diversity may be calculated at different scales. Alpha diversity accounts for the richness (number) of species in a single sample unit, and the mean ($\bar{\alpha}$) may be calculated from a number of replicate samples in an area. Gamma diversity (γ) describes the total number of species present in a defined larger area. For these first tests, we replace species with OGUs. Our experimental design provides two hierarchical levels for γ-diversity: one at the region level, which is the total number of 'species' present in all samples from a region (regional-γ), and another at the national level, which is the total number present in all four regions spanning NZ (national-γ, see Table 1). Beta (β) diversity provides an estimate of 'species' variance at intermediate scales and here may be calculated at the regional and national scale. Many measures of β-diversity have been developed (Anderson *et al.*, 2011), but here we employ Whittaker's classic measure [$\beta_w = (\gamma/\bar{\alpha}) - 1$], which calculates the proportion by which a given area is richer in species than the average sample within it (Anderson *et al.*, 2011). The greater values for national β_w than regional β_w (Table 1) show that β_w diversity is greater at the national scale: i.e. individual samples on average differ more from national γ-diversity than from their local γ-diversity. This pattern might be expected if communities tend to be region-specific (i.e. are clustered and not homogenized). We employed a hierarchical null model to test whether the observed partition of species richness (numbers of OGUs) could be explained by a random

Fig. 2. The differences in fungal communities on ripe Chardonnay grapes between four NZ regions. The size (area) of the circle representing each region is proportional to the observed OGU richness for that region (refer to Table 1). The thickness of lines connecting regions corresponds to the Jaccard similarity (= 1-dissimilarity) between relative 98% OGU abundances in each region. The pie charts show the breakdown of read abundance by genus, with OGUs unclassified at this level removed. Only the labels for the eight most abundant genera are shown, as rare genera are not readily discernable in the plots.

distribution of individuals in space (Crist *et al.*, 2003). There is significantly lower average 'species richness' in both individual α sampling units (45 OGUs) and regions (123 OGUs) compared with the average richness simu-

lated under a random model (78 and 180, respectively; $P < 0.001$). This analysis rejects the null hypothesis of randomized species distributions across different regions of NZ.

Table 2. Multivariate analysis of variance of differences in fungal community composition and richness by region based on incidence (Jaccard) and presence/absence (Sørensen's) at varying taxonomic levels.

	Jaccard		Sørensen	
Level	No. reads	No. OGUs	No. reads	No. OGUs
OGU	0.001 (0.280)	n.a.	0.0001 (0.359)	n.a.
Genus	0.0014 (0.286)	0.0001 (0.373)	0.0001 (0.415)	0.0001 (0.415)
Family	0.0035 (0.250)	0.0001 (0.358)	0.0001 (0.465)	0.0001 (0.466)
Class	0.0206 (0.247)	0.0001 (0.400)	0.0010 (0.350)	0.0005 (0.350)
Order	0.0061 (0.243)	0.0001 (0.362)	0.0001 (0.515)	0.0001 (0.515)
Phylum	0.014 (0.368)	0.0046 (0.434)	0.7977 (0.053)	0.7998 (0.053)

The P values were determined by comparisons of F-tests based on sequential sums of squares from 10 000 permutations of the raw data, and the R^2 values are in parentheses.
n.a., not applicable.

Second, we performed a complimentary multivariate analysis of variance (MVAOVA) on community composition (i.e. relative abundances) and richness (number of 98% OGUs). We conducted these tests at all taxonomic levels where reads and OGU numbers are summed appropriately (see Supporting Information Table S1). Table 2 shows the P values revealed by these tests. These tests provide strong support for the concept that there is a regional effect on both community composition and richness at most taxonomic levels for these fruit-associated fungi. MVAOVA analyses of individual pairwise comparisons between regions show that all regions significantly differ in terms of OGU richness, but that only Central Otago differs from other regions in terms of abundances (Table 3). The average Jaccard community similarities are displayed in Fig. 2. The R^2 values from MVAOVA analyses indicate that on average, 32% of variance in both community composition and diversity between samples are explained by region at both the OGU level and overall taxonomic levels. Lastly, we also visualized these data by analyses with multidimensional scaling of Jaccard distances generated from the abundances of reads for each OGU, and this reveals a pattern of regional difference that correlates strongly with the results from the suite of other tests (see Fig. 3).

Overall then, the weight of evidence means we accept the alternate hypothesis of significant differences in both fungal community OGU diversity (richness) and commu-

nity composition associated with Chardonnay grapes between different regions of NZ.

Drivers of differentiation

The 12 most abundant OGUs cover 93.6% of all reads: seven OGUs were classified to *Cladosporium*, *Davidiella*, *Erysiphe*, *Columnosphaeria*, *Penicillium*, *Cryptococcus* and *Alternaria* genera, and five were unclassified at genus level but assigned to Leotiomycetes, Microbotryomycetes, Eurotiomycetes and Tremellomycetes classes. All five ubiquitous OGUs are in this most abundant list and of these four were classified to Dothideomycetes and one to Tremellomycetes classes. Representative sequences for the ubiquitous OGUs BLASTed to uncultured *Cladosporium* and *Davidiella* entries, various *Alternaria* and *Cryptococcus* species entries, as well as uncultured deposits and multiple *Aureobasidium pullulans* deposits. The abundances of these ubiquitous OGUs did not significantly differ among regions according to univariate analysis of variance (ANOVA) analyses on each individually ($F_{[3,19]}$ between 1.0 and 2.8; $P < 0.0697$; one-way ANOVA on number of reads by region). Surprisingly, univariate analyses of individual OGUs reveal that only one of these top 12 (#35, Basidiomycete: *Cryptococcus* sp.) differs significantly by region ($F_{[3,19]} = 14.4799$, $P < 0.0001$), and this is due to a significantly higher number of reads in Marlborough only (revealed by a

Table 3. P values, and R^2 values in parentheses, revealed from pairwise multivariate analysis of variance on number of reads per 98% OGU with Jaccard (below diagonal) or Sørensen's presence/absence dissimilarities (above diagonal) for replicate samples among the four NZ regions examined: WA, West Auckland; HB, Hawke's Bay; MB, Marlborough; CO, Central Otago.

		Sørensen			
		WA	HB	MB	CO
Jaccard	WA	–	0.004 (0.165)	0.0022 (0.389)	0.0022 (0.282)
	HB	0.1259 (0.133)	–	0.0016 (0.191)	0.0019 (0.303)
	MB	0.1539 (0.312)	0.3128 (0.102)	–	0.0023 (0.277)
	CO	0.0033 (0.301)	0.0089 (0.230)	0.0023 (0.262)	–
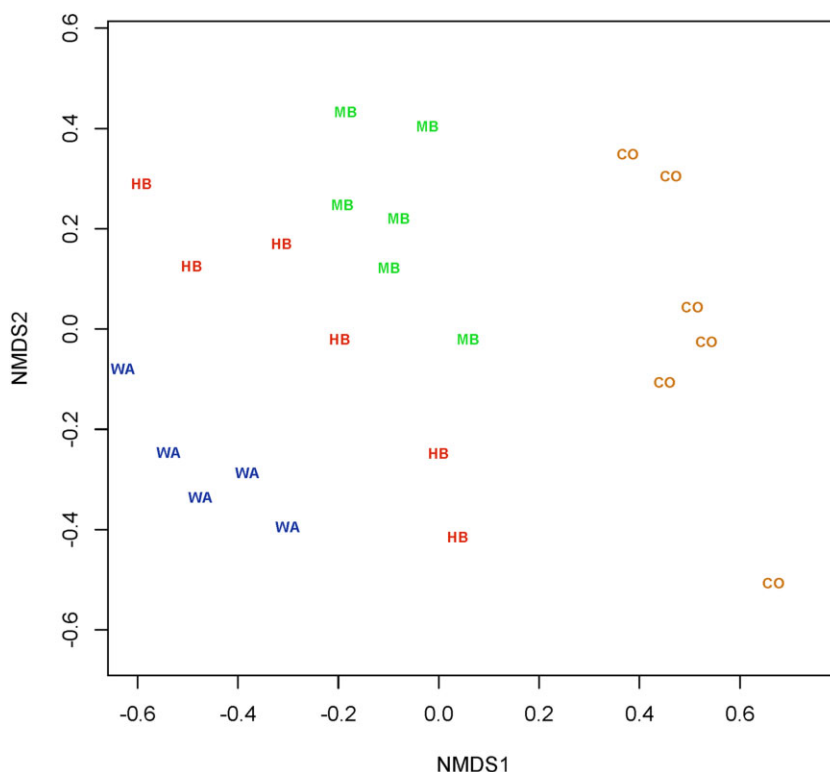
Fig. 3. Non-metric multidimensional scaling (NMDS) plot based on Jaccard dissimilarities calculated from read numbers among the 253 inferred OGUs between replicate samples of communities from the four regions. Stress is 0.19. WA, West Auckland; HB, Hawke's Bay; MB, Marlborough; CO, Central Otago.

Tukey–Kramer honest significant difference post-hoc test, $\alpha = 0.01$). However, while only one of these OGUs differs according to individual univariate analyses, MVAOVA simultaneously analysing the abundances of these OGUs reveals that these 12 provide a sufficiently strong signal for variance in community composition by region ($F_{[3,19]} = 7.0921$, $P < 0.0001$). This analysis is insensitive to the presence of OGU#35 ($F_{[3,19]} = 7.0748$, $P < 0.0001$) and to whether the community composition or OGU richness are analysed ($F_{[3,19]} = 10.267$, $P < 0.0001$). Table 1 details the number of OGUs that are unique to a region, but no one OGU was present in all samples within a region and absent in all others. However, 12, 9, 3 and 3 OGUs were dominant in West Auckland, Hawke's Bay, Marlborough and Central Otago, respectively, i.e. these OGUs were in more than three samples from the respective region and in less than three from all others (see Supporting Information Table S1).

Discussion

This estimate of microbial community diversity associated with vines broadly agrees with the other two recent studies to use direct DNA deep community sequencing approaches (Bokulich *et al.*, 2014; Pinto *et al.*, 2014). Together with these, we reveal approximately an order of magnitude more fungal taxa from this niche compared with the tens of species that have previously been recov-

ered using culture-based approaches (Prakitchaiwattana *et al.*, 2004; Ciani *et al.*, 2010; Čadež *et al.*, 2010; Gayevskiy and Goddard, 2012). The samples from West Auckland and Hawke's Bay analysed here are equivalent subsamples to those used in a parallel culture-based assessment of community diversity using sequence at the same 26S rRNA locus for identification (Gayevskiy and Goddard, 2012); only four and six species were recovered, respectively, from the ~600 colonies analysed from each region. Here, 129 and 164 OGUs were discovered from ~25 000 reads from the same respective West Auckland and Hawke's Bay samples (Table 1), suggesting that culture-based approaches may miss ~95% of the community. Thus, in line with inferences for bacteria inhabiting soil (Keller and Zengler, 2004), culture-based approaches are a poor estimator of fungal species richness in this system.

While there are descriptions of fungal species associated with vines, the ecological functions of fungi in this niche are less well characterized. We clearly cannot make any statements about which fungi are active in this community, just those that are present. While the OGU diversity was approximately equivalent for the Ascomycota and Basidiomycota, the Ascomycota were prevalent in terms of number of reads. Arguably, the best characterized are the species and genera known to cause disease (Pancher *et al.*, 2012). A range of commonly occurring Helotiales, the second most abundant order found here, has also

been described from a recent study examining fungi associated with *Quercus* (oak) tree roots, which are probably root endophytes (Toju *et al.*, 2013). One family in the Helotiales, the Sclerotiniaceae, includes many plant pathogens, and this family comprises 15% of reads here. While the well-known grape pathogen *Botryotinia* is in this family, none of our OGUs were assigned to this genus. Approximately 8% of reads belong to the Erysiphaceae, which contains many species of obligate parasites causing powdery mildew. Of course, the presence of DNA from these species does not necessarily indicate that they are actively causing disease. Indeed, reports from vineyard managers from the year we sampled were that the general health of the fruit was good. This method of analysis might therefore provide an indication of potential disease load. Other common genera are also what one might expect from environmental samples; the most common genus, *Cladosporium*, is known to harbour the most common indoor and outdoor moulds, and these are not especially pathogenic to plants. Another common genus here was the supposedly ubiquitous *Penicillium*. While the effects of some Ascomycota on the fermentation kinetics of fruits and corresponding flavour profile of wines are known (Chambers and Pretorius, 2010; Ciani *et al.*, 2010), this gives us little insight into their ecological role on the vine before the fruit is harvested. Are these passive members of the community, or do they play a more active role? Unsurprisingly, *Saccharomyces'* ecological role is probably the best known in this niche – while rare (only five reads among these samples), it appears extremely well adapted to invade fruits once ripe and damaged through an ecosystem engineering strategy that creates a hot, anaerobic alcoholic environment, which is manifest as fermentation (Goddard, 2008).

Our main motivation for this work was to examine whether there are any geographical patterns for fungal communities associated with fruits in NZ. Our analyses with null models allow us to reject the hypothesis of no geographical patterns in microbial communities associated with grapes. We employed a suite of analyses to evaluate both community composition (relative species abundances) and community diversity (species richness). We show that all regions differ from one another in terms of diversity, but that only Central Otago is distinct in terms of community composition. These findings are in line with those based on culture-based approaches on a subset of the same samples (Gayevskiy and Goddard, 2012), as well as in line with indirect DNA analyses evidence concerning spatial differences between both fungal and bacterial communities associated with Chardonnay and Cabernet fruit across California (Bokulich *et al.*, 2014). The demonstration that certain regions have 'signature' microbial populations is of relevance to the wine industry. It is often suggested that certain wines reflect their geo-graphical origin, and this is encapsulated in the concept of *terroir* (Bokulich *et al.*, 2014). This classically was thought to largely result from the interaction with specific *V. vinifera* varieties and the local soils, geography and climate. However, the limited but increasing evidence showing that the microbes, which may influence vines, also exhibit regional differentiation supports the concept that there might be a microbial aspect to *terroir*.

There may be any number of reasons why communities differ between regions, and it is tempting to speculate. At the highest level, these biogeographical patterns may be a function of neutral processes, where different communities become established by chance, and lack of species range expansion (dispersal) allows these patterns to persist (Hughes Martiny *et al.*, 2006). At the other extreme, in line with the much-touted Baas Becking hypothesis, is the notion that there is no limit to the range of species but that selection sorts these species and defines community composition and diversity in any one area (Hanson *et al.*, 2012). Climate correlates with differences in vine-associated microbial communities in California (Bokulich *et al.*, 2014), and as one moves south down NZ, the climate becomes increasingly cold and dry so the pattern of lower fungal species richness in the southern most regions hints that selection might have a role in determining these patterns. Of course, Central Otago is also the most remote region, and so the greater differentiation of these communities might simply be due to lack of dispersal. Teasing apart the roles of selection and neutral processes in defining variance in community structure is not a trivial task and beyond the scope of this study.

Experimental procedures

Sampling

We sampled fruit from 24 distinct *V. vinifera* var. Chardonnay vineyards among four major and distinct regions (*n* = 6 for each region) across NZ's North and South Islands: (from north to south) West Auckland, Hawke's Bay, Marlborough and Central Otago (see Fig. 2). The mean distance between vineyards within regions is ~10 km, while the mean distance between regions is ~600 km, with West Auckland and Central Otago being 1000 km apart. All samples were taken approximately 1 week before harvest in mid-March in 2010 so that most grapes were at approximately the same ripeness. At each of the 24 vineyards, we randomly selected and collected three entire bunches of grapes, avoiding rotten fruit. These samples were removed with sterile pruning shears into a sterile plastic bag and transported to the laboratory on ice where the fruit was crushed in situ. The resulting juice was briefly washed over the entire bunch before being transferred to sterile centrifuge tubes. The samples were centrifuged for 10 min at 4000 r.p.m., after which the pellet was re-suspended in 10 ml sterile water then stored at −20°C until subsequent analysis.

DNA extraction, PCR and 26S rRNA gene pyrosequencing of fungal communities

DNA was extracted from juice pellets by bead-beating in an ammonium acetate buffer, as described previously (Taylor *et al.*, 2004). The only significant modification to the published protocol was the addition of four 2.3 mm glass beads to each tube prior to bead-beating. PCR amplification of the ~600 bp divergent domains 1 and 2 (D1/D2) region of the 26S rRNA gene was performed using the primers NL-1 and NL-4 (Kurtzman and Robnett, 2003) and a FastStart High-Fidelity PCR System (Roche). The ITS region is also commonly used to discriminate between and identify fungi (Kurtzman and Robnett, 2003), but the D1/D2 region was chosen over the ITS region as it does not display the length polymorphism of the ITS and is thus less likely subject to PCR and subsequent sequencing bias, as well as producing more reliable alignments. In addition, the D1/D2 region now has a comprehensive reference database available for taxonomic assignment (see later). Primers contained the appropriate adaptor and multiplex identifier (MID) sequences for Titanium pyrosequencing. Approximately 200 ng of extracted DNA was used for PCR amplification, and the PCR product was loaded on a 1% agarose gel and bands of ~650 bp were excised and purified using an illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare). Product quantity was assessed using a Nanodrop 1000 spectrophotometer and purity determined with Agilent 1200 Bioanalyzer DNA 1000 chips. One of the West Auckland samples failed to amplify, leaving $n = 5$ for this region. Pyrosequencing was performed by Macrogen (Korea) on one half of a 454 Life Sciences GS-FLX Titanium instrument split into eight regions, and each region of the pyrosequencing plate contained randomized samples and MIDs. These data are deposited in the NCBI Sequence Read Archive with accession number SRX091268.

Experimentally controlling for sequencing errors

There have been recent reports of high errors in pyrosequencing data (Kunin *et al.*, 2010), and so we also included a biological control to determine if our inference of community diversity might be artificially inflated; we propagated two independent clonal populations of *S. cerevisiae* strain S288C in the laboratory, extracted DNA from these clonal populations and amplified the D1/D2 region. We included these PCR products as two additional independent samples, with separate MIDs, randomized in the 454 run along with samples derived from fruit. In theory, only one OGU/species should be inferred from the analyses of these clonal populations, while more than this would signal the possibility for sequencing error to artificially inflate our estimate of diversity.

Bioinformatic read processing

Initial processing of pyrosequencing reads implemented the denoising PyroNoise algorithm in MOTHUR 1.9.1 (Schloss *et al.*, 2009) using the shhh.flows command, the removal of the MID and primer sequences, as well as removal of any reads < 350 bp in length and any reads containing 'N's and/or a homopolymer stretch of > 8 nucleotides and chimeras. In addition, any sequence read that lacked the correct MID and/or primer sequence was removed from further analyses. Reads were quality-filtered and trimmed using the software Lucy (Li and Chou, 2004), with an error setting of $P = 0.001$. Unique sequences were identified with MOTHUR, and then aligned against the SILVA (Pruesse *et al.*, 2007) large subunit alignment database (References used: 1165 sequences belonging to the Fungi from SILVA Release 106, 23S/28S ribosomal RNA) using the Needleman algorithm with a *k*-mer search. The reads were binned into classes in which the maximum uncorrected pairwise distance, calculated using the average neighbour clustering algorithm, was 2%. This 98% genetic similarity delineation was based upon approximate empirical delineations between species at the 26S rRNA D1/D2 region within both Ascomycota and Basidiomycota (Kurtzman and Robnett, 1998; Romanelli *et al.*, 2010).

Taxonomic classification

A representative sequence for each 98% similarity class was derived by identifying the individual sequence with the minimum distance to all other sequences within that class. This sequence was compared with the fungal rRNA reference database (available via the Ribosomal Database Project), a collection of 8506 taxonomically fully classified (down to genus) reference rRNA gene sequences from fungi (http://www.mothur.org/w/images/3/36/FungiLSU_train_v7.zip). We used the 'wang' *k*-mer method (implemented by 'Classify.seqs' in MOTHUR) with a bootstrap confidence cut-off value of 70% to classify all representative sequences at kingdom, phylum, class, order, family and genus levels. Sequences with bootstrap scores below the 70% value at a given taxonomic level were left as 'unclassified'. Classification at species level was more problematic, as the rRNA reference database is not resolved to this level, and thus we employed BLAST analyses against the entirety of deposits in GenBank to elucidate likely species only for 98% sequence similarity classes of interest.

Statistical analyses of fungal community composition

Tests for whether observed partitions of species richness (i.e. presence/absence) differed from those expected under a neutral model followed (Crist *et al.*, 2003) and were implemented using 999 randomized permutations of the data with the 'adipart' function in the *vegan* package (Oksanen *et al.*, 2013) in R (R Core Development Team, 2013). Estimated richness in each region was computed by subsampling rarefaction in *vegan*, and rarefaction curves were also estimated using Estimate S v9.1.0 (Colwell, 2013).

Tests for community differentiation by region were conducted on dissimilarity distance matrices using MVAOVA (McArdle and Anderson, 2001) as implemented in the 'adonis' function in *vegan*. MVAOVA is directly analogous to a nonparametric MANOVA and is a robust alternative to parametric MANOVA and ordination methods to test how fungal community variance is partitioned by geographical origin. Significance tests are achieved using *F*-tests based on sequential sums of squares from permutations of the raw data to provide pseudo-*F* ratios. The probabilities of observing the pseudo-*F*

ratios were calculated by comparison with null distributions of 10 000 permutations. Community differences between regions were also analysed and visualized using multidimensional scaling plots. Absolute read numbers (abundances) were not analysed as there was approximately a hundred-fold variance in read number between individual samples ranging from ~400 to 10 000. We therefore transformed the data into relative proportions within each sample, i.e. margin totals equalled one. This normalized data matrix allows tests for differences in relative community composition. There are many distance measures to compare communities, but as we were interested in testing community compositions, dissimilarities were estimated using Jaccard's distances, which are based on Bray–Curtis distances but are metric and thus preferable to Bray–Curtis that is semimetric. We also compared species richness using presence/absence Sørensen's index, which is equivalent to a binary Bray–Curtis index.

Acknowledgements

References

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59:** 143–169.

Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L., *et al.* (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecol Lett* **14:** 19–28.

Barata, A., Malfeito-Ferreira, M., and Loureiro, V. (2012) The microbial ecology of wine grape berries. *Int J Food Microbiol* **153:** 243–259.

Bik, H.M., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, R., and Thomas, W.K. (2012) Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol Evol* **27:** 233–243.

Bokulich, N.A., and Mills, D.A. (2013) Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Appl Environ Microbiol* **79:** 2519–2526.

Bokulich, N.A., Thorngate, J.H., Richardson, P.M., and Mills, D.A. (2014) Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *Proc Natl Acad Sci USA* **111:** E139–E148.

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

Ciani, M., Comitini, F., Mannazzu, I., and 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.

Colwell, R.K. (2013) Estimate S: Statistical estimation of species richness and shared species from samples. URL http://purl.oclc.org/estimates.

Cordero-Bueso, G., Arroyo, T., Serrano, A., Tello, J., Aporta, I., Vélez, M.D., and Valero, E. (2011) Influence of the farming system and vine variety on yeast communities associated with grape berries. *Int J Food Microbiol* **145:** 123–139.

Crist, T.O., Veech, J.A., Gering, J.C., and Summerville, K.S. (2003) Partitioning species diversity across landscapes and regions: a hierarchical analysis of alpha, beta, and gamma diversity. *Am Nat* **162:** 734–743.

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

De Queiroz, K. (2007) Species concepts and species delimitation. *Syst Biol* **56:** 879–886.

Finlay, B.J. (2002) Global dispersal of free-living microbial eukaryote species. *Science* **296:** 1061–1063.

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

Gayevskiy, V., and Goddard, M.R. (2012) Geographic delineations of yeast communities and populations associated with vines and wines in New Zealand. *ISME J* **6:** 1281–1290.

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

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

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10:** 497–506.

van der Heijden, M.G.A., Bardgett, R.D., and van Straalen, N.M. (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol Lett* **11:** 296–310.

Hodge, A., Campbell, C.D., and Fitter, A.H. (2001) An arbuscular mycorrhizal fungus accelerates decomposition and acquires nitrogen directly from organic material. *Nature* **413:** 297–299.

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

Hughes Martiny, J.B., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102–112.

Jumpponen, A., Jones, K.L., David Mattox, J., and Yaege, C. (2010) Massively parallel 454-sequencing of fungal communities in *Quercus* spp. ectomycorrhizas indicates seasonal dynamics in urban and rural sites. *Mol Ecol* **19** (Suppl. 1): 41–53.

Keller, M., and Zengler, K. (2004) Tapping into microbial diversity. *Nat Rev Microbiol* **2:** 141–150.

Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2010) Wrinkles in the rare biosphere: pyrosequencing

errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12:** 118–123.

Kurtzman, C.P., and Robnett, C.J. (1998) Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Anton Leeuw Int J G* **73:** 331–371.

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

Lachance, M.A., Lawrie, D., Dobson, J., and Piggott, J. (2008) Biogeography and population structure of the neotropical endemic yeast species *Metschnikowia lochheadii*. *Anton Leeuw Int J G* **94:** 403–414.

Lachance, M.A., Starmer, W.T., Bowles, J.M., Phaff, H.J., and Rosa, C.A. (2000) Ribosomal DNA, species structure, and biogeography of the cactophilic yeast *Clavispora opuntiae*. *Can J Microbiol* **46:** 195–210.

Lachance, M.A., Starmer, W.T., Rosa, C.A., Bowles, J.M., Barker, J.S., and Janzen, D.H. (2001) Biogeography of the yeasts of ephemeral flowers and their insects. *FEMS Yeast Res* **1:** 1–8.

Landry, C.R., Townsend, J.P., Hartl, D.L., and Cavalieri, D. (2006) Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. *Mol Ecol* **15:** 575–591.

Li, S., and Chou, H. (2004) Lucy 2: an interactive DNA sequence quality trimming and vector removal tool. *Bioinformatics* **20:** 2865–2866.

Loreau, M. (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294:** 804–808.

McArdle, B., and Anderson, M. (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82:** 290–297.

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

von Mering, C., Hugenholtz, P., Raes, J., Tringe, S.G., Doerks, T., Jensen, L.J., *et al.* (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315:** 1126–1130.

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

Oksanen, J., Blanchet, G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, O.B., *et al.* (2013) vegan: Community Ecology Package.

Orgiazzi, A., Bianciotto, V., Bonfante, P., Daghino, S., Ghignone, S., Lazzari, A., *et al.* (2013) 454 pyrosequencing analysis of fungal assemblages from geographically distant, disparate soils reveals spatial patterning and a core mycobiome. *Diversity* **5:** 73–98.

Pancher, M., Ceol, M., Corneo, P.E., Longa, C.M.O., Yousaf, S., Pertot, I., and Campisano, A. (2012) Fungal endophytic communities in grapevines (*Vitis vinifera* L.) respond to crop management. *Appl Environ Microbiol* **78:** 4308–4317.

Pinto, C., Pinho, D., Sousa, S., Pinheiro, M., Egas, C., and Gomes, A. (2014) Unravelling the diversity of grapevine microbiome. *PLoS ONE* **9:** e85622.

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

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.

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

Romanelli, A.M., Sutton, D.A., Thompson, E.H., Rinaldi, M.G., and Wickes, B.L. (2010) Sequence-based identification of filamentous basidiomycetous fungi from clinical specimens: a cautionary note. *J Clin Microbiol* **48:** 741–752.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al.* (2009) Introducing MOTHUR: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Setati, M.E., Jacobson, D., Andong, U.-C., and Bauer, F. (2012) The vineyard yeast microbiome, a mixed model microbial map. *PLoS ONE* **7:** e52609.

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

Taylor, J.W., Turner, E., Townsend, J.P., Dettman, J.R., and Jacobson, D. (2006) Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philos Trans R Soc B* **361:** 1947–1963.

Taylor, M., Schupp, P., Dahllof, I., Kjelleberg, S., and Steinberg, P. (2004) Host specificity in marine sponge-associated bacteria, and potential implications for marine microbial diversity. *Environ Microbiol* **6:** 121–130.

Team, R.C. (2013) R: A Language and environment for statistical computing.

Toju, H., Yamamoto, S., Sato, H., Tanabe, A.S., Gilbert, G.S., and Kadowaki, K. (2013) Community composition of root-associated fungi in a *Quercus*-dominated temperate forest: 'codominance' of mycorrhizal and root-endophytic fungi. *Ecol Evol* **3:** 1281–1293.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., *et al.* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428:** 37–43.

Zott, K., Thibon, C., Bely, M., Lonvaud-Funel, A., Dubourdieu, D., and Masneuf-Pomarède, I. (2011) The grape must non-*Saccharomyces* microbial community: impact on volatile thiol release. *Int J Food Microbiol* **151:** 210–215.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. OGU (OTU) collection curves for the replicate samples in each of the four New Zealand regions.
Table S1. Number of reads for each OGU broken down by both sample location and taxonomic level.