ORIGINAL ARTICLE

Quantifying separation and similarity in a *Saccharomyces cerevisiae* metapopulation

Sarah Knight and Matthew R Goddard

School of Biological Sciences, University of Auckland, Auckland, New Zealand

Eukaryotic microbes are key ecosystem drivers; however, we have little theory and few data elucidating the processes influencing their observed population patterns. Here we provide an in-depth quantitative analysis of population separation and similarity in the yeast *Saccharomyces cerevisiae* with the aim of providing a more detailed account of the population processes occurring in microbes. Over 10 000 individual isolates were collected from native plants, vineyards and spontaneous ferments of fruit from six major regions spanning 1000 km across New Zealand. From these, hundreds of *S. cerevisiae* genotypes were obtained, and using a suite of analytical methods we provide comprehensive quantitative estimates for both population structure and rates of gene flow or migration. No genetic differentiation was detected within geographic regions, even between populations inhabiting native forests and vineyards. We do, however, reveal a picture of national population structure at scales above ~100 km with distinctive populations in the more remote Nelson and Central Otago regions primarily contributing to this. In addition, differential degrees of connectivity between regional populations are observed and correlate with the movement of fruit by the New Zealand wine industry. This suggests some anthropogenic influence on these observed population patterns.
The ISME Journal (2015) 9, 361–370; doi:10.1038/ismej.2014.132; published online 25 July 2014

Introduction

Research to date attempting to elucidate the patterns and processes involved in shaping natural populations has largely focused on readily observed macro-organisms but comparatively little work has been conducted on microbial species (Anderson and Kohn, 1998; Martiny *et al.*, 2006; Prosser *et al.*, 2007). Because of their large population sizes, and ease of transfer, one might expect microbial populations to be well mixed (Finlay, 2002); however, there is increasing evidence showing that many are not homogeneous but structured (Taylor *et al.*, 2006; Whitaker and Banfield, 2006; Hanson *et al.*, 2012). Most microbial ecology studies have focused on bacteria, but eukaryotic microbes, which undergo sex (with recombination), also have key ecosystem roles (Green *et al.*, 2008; Van Der Heijden *et al.*, 2008). It is not clear whether the population patterns estimated for eukaryotic 'macrobes' generally hold for eukaryotic microbes (Hartl and Clark, 1997; Anderson and Kohn, 1998; Halkett *et al.*, 2005; Taylor *et al.*, 2006; Prosser *et al.*, 2007; Tsai *et al.*, 2008).

A metapopulation comprises a number of spatially separated populations of the same species that interact to some extent. To date all studies examining microbial populations have simply examined whether population structure is evident or not (Aa *et al.*, 2006; Achtman, 2008; Liti *et al.*, 2009; Goddard *et al.*, 2010; Anderson and Shearer, 2011; Härnström *et al.*, 2011; Gayevskiy and Goddard, 2012; Wang *et al.*, 2012). Merely defining microbial populations as either structured or homogeneous is highly unlikely to reflect the true biological situation. A more accurate approach is not only to assess the degree to which populations are structured, but also connected by gene flow, and crucially go onto quantify these processes; however, to the best of our knowledge, there are no previous studies that have used a unified framework to do this. Here we analyse the natural population of *Saccharomyces cerevisiae* in New Zealand (NZ), and in doing so take the first steps towards quantifying microbial population structure and similarity.

S. cerevisiae, a budding yeast, has been closely associated with humans since the dawn of civilisation because of its fermentative capabilities, and has come to be of significant commercial importance in the production of bread, wine, beer and other alcoholic beverages (McGovern *et al.*, 1996; Pretorius, 2000; Cavalieri *et al.*, 2003; Chambers and Pretorius, 2010). *S. cerevisiae* is also a classic

Correspondence: S Knight, School of Biological Sciences, University of Auckland, 3A Symonds St Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand.
E-mail: s.knight@auckland.ac.nz

362

model organism for research into cell biology, genetics and increasingly ecology and evolution (Chambers and Pretorius, 2010; Dujon, 2010; Gray and Goddard, 2012; Hittinger, 2013; Hyma and Fay, 2013). Recent studies have revealed a large genetic diversity within *S. cerevisiae*, and there is good evidence for population structure at intercontinental scales (Fay and Benavides, 2005; Schuller *et al.*, 2005; Aa *et al.*, 2006; Lopandic *et al.*, 2008; Liti *et al.*, 2009; Goddard *et al.*, 2010; Mercado *et al.*, 2011; Di Maio *et al.*, 2012; Wang *et al.*, 2012). Similar inferences have been made at finer scales with reports that some genotypes were unique to different geographic locations in Austria, although many were also ubiquitous across regions (Lopandic *et al.*, 2008). In addition, Bayesian inference shows genetic differentiation between populations spanning hundreds of kilometres in NZ (Gayevskiy and Goddard, 2012). While the scales of these studies differ, they all commonly report the presence of hybrid or mosaic strains indicative of some levels of connectivity between populations via gene flow.

Global-scale analyses have suggested that ecological function may define population structure to a greater extent than geographic origin (Fay and Benavides, 2005; Legras *et al.*, 2007). Strains associated with wine appear somewhat distinct from those isolated from distilling, bread making, fermented milk, rice wine, ale and lager, with geographic origin only explaining 28% of variability (Legras *et al.*, 2007). Furthermore, whole-genome analyses of a limited number of strains suggest specific *S. cerevisiae* populations associated with vineyards, sake and related ferments, although some of these clusters are confounded with geographic origin (Liti *et al.*, 2009; Schacherer *et al.*, 2009). In contrast, a recent rigorous genome-wide population study has provided evidence of gene flow across small distances (<17 km) between distinct populations inhabiting vineyards and oak trees, showing connectivity between ecological niches at small scales (Hyma and Fay, 2013).
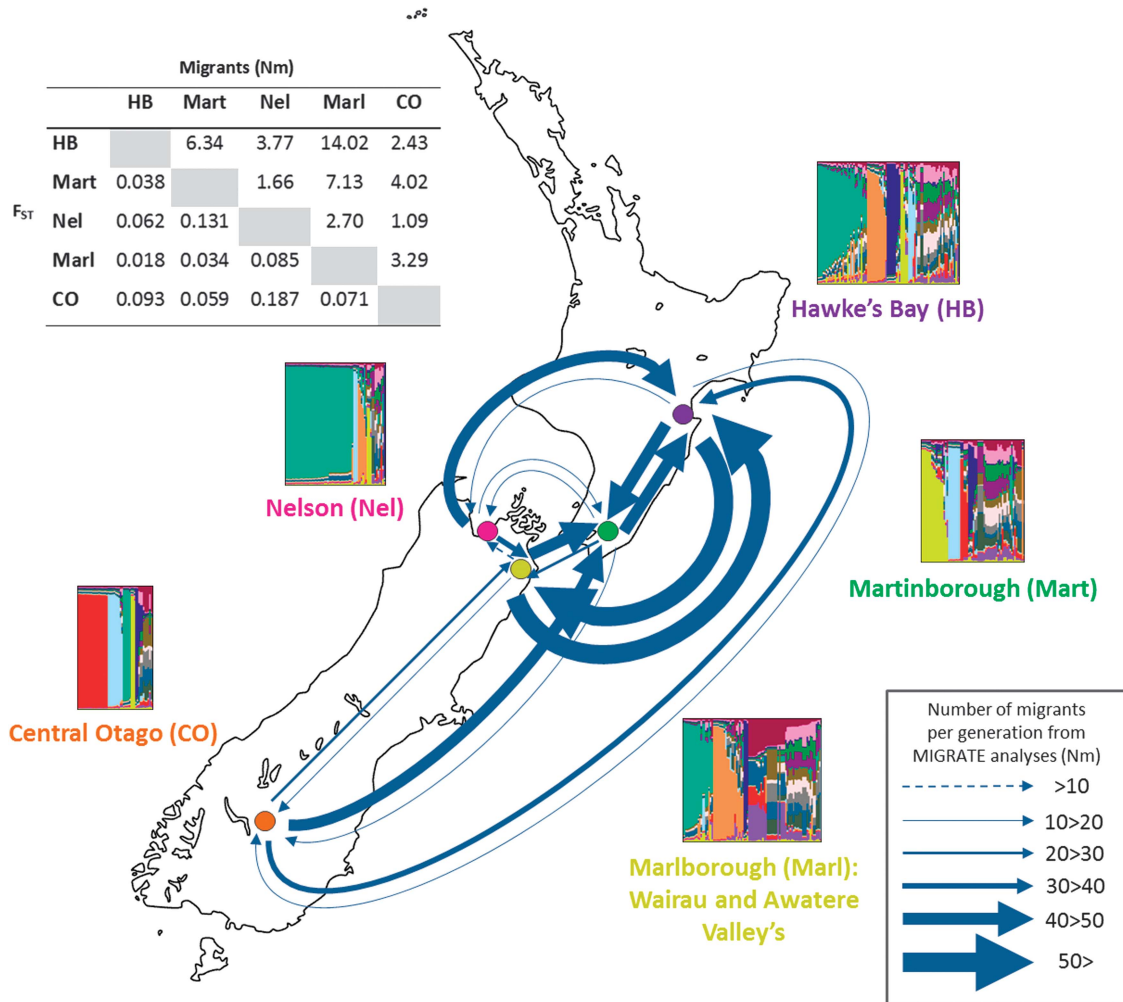
Despite these excellent efforts to date, most studies have been drawn from widely dispersed isolates often from different ecological niches, with relatively small sample sizes from any one discrete population (Fay and Benavides, 2005; Aa *et al.*, 2006; Legras *et al.*, 2007; Liti *et al.*, 2009; Schacherer *et al.*, 2009; Wang *et al.*, 2012) and have thus not afforded adequate power to quantify 'ecological scale' population processes such as gene flow. Previous work using both microsatellite and RAD-seq analyses shows that a distinct *S. cerevisiae* population resides in NZ, suggesting that this population is not subject to rampant inward international gene flow (Goddard *et al.*, 2010; Cromie *et al.*, 2013). Therefore, in addition to its geographic isolation, the NZ *S. cerevisiae* population appears relatively self-contained and thus provides a good population to study the processes we are interested in.

Here we analyse close to a thousand *S. cerevisiae* isolates from four niches across six regions spanning over 1000 km. We quantify both the degree to which this population is structured and go on to quantify the extent to which the various regional populations are connected by gene flow in one of the most comprehensive studies of a microbial metapopulation to date. Lastly, this study sheds light on the connection between farmed (managed) and native ecosystems by examining the relationship between microbial populations residing in vineyards and native NZ forest.

Materials and methods

Sample collection and processing

Six to seven *Vitis vinifera* var. Sauvignon Blanc vineyards were selected from each of Hawke's Bay, Martinborough, Nelson, Awatere Valley, Wairau Valley and Central Otago in NZ (Figure 1). Approximately 5 g of soil were aseptically taken from each of these 37 vineyards between 1 and 4 weeks before harvest in mid-March 2011. Ten litres of juice derived from the same vineyards were collected from commercial settling tanks (one vineyard provided juice samples from two pressing tiers, resulting in a total of 38 juice samples). Soil and fruit samples were taken from six native NZ plants located in non-managed native bush reserves within each region (Supplementary Table S2), ranging from 0.1 to 50 km from the vineyard sites, totalling 72 native samples (36 soil and 36 fruit). *S. cerevisiae* is rare in niches other than in actively fermenting fruit, so equivalent selective culturing methods were employed for all samples to control for the effects of high sugar and ethanol (Mortimer and Polsinelli, 1999; Pretorius, 2000; Xufre *et al.*, 2006; Goddard, 2008; Taylor *et al.*, 2014). An enrichment method emulating fermenting selection pressures was employed for all 147 environmental samples (Mortimer and Polsinelli, 1999; Serjeant *et al.*, 2008). Samples were submerged in 10 ml SelMed media (1% yeast extract, 2% peptone, 10% glucose and 5% ethanol) for six days; 500 μl was then transferred to 10 ml fresh SelMed for four additional days, and then dilutions plated onto YPD (1% yeast extract, 2% peptone, 2% glucose) with 50 μg ml^{-1} chloramphenicol to retard bacterial growth. All incubation was at 28 °C. Up to 94 colonies were taken from each sample and stored in 15% glycerol at −80 °C. A total of 7144 individuals were isolated from environmental samples. A natural enrichment of the juice samples was performed by allowing them to ferment spontaneously at 15 °C. In all, 100 ml was concentrated by centrifugation after 21 days, and plated on YPD with 50 μg ml^{-1} chloramphenicol. Again 94 colonies were isolated from each ferment sample totalling 3572 individuals. All niches were thus evenly sampled and in total 10 716 individuals were collected.

F_{ST}	Migrants (Nm)				
	HB	**Mart**	**Nel**	**Marl**	**CO**
HB		6.34	3.77	14.02	2.43
Mart	0.038		1.66	7.13	4.02
Nel	0.062	0.131		2.70	1.09
Marl	0.018	0.034	0.085		3.29
CO	0.093	0.059	0.187	0.071	

Hawke's Bay (HB)

Nelson (Nel)

Martinborough (Mart)

Central Otago (CO)

Marlborough (Marl):
Wairau and Awatere
Valley's

Number of migrants
per generation from
MIGRATE analyses (Nm)

- - - → >10
───→ 10>20
───→ 20>30
━━→ 30>40
━━▶ 40>50
━━▶ 50>

Figure 1 The location of NZ regions and analyses of population structure and connectivity. Plots of the ancestry profiles are shown beside each region: each vertical line represents an individual with the different colours showing the proportion of ancestry of each individual to each of the 16 inferred populations. Arrows connecting different regions show directional migration rates as calculated in MIGRATE with the width of the arrows representing the number of migrants per generation as indicated in the scale. Absolute numbers can be found in Supplementary Table S4. The table reports pairwise F_{ST} values below the diagonal and the number of migrants per generation (Nm) as calculated from F_{ST} above the diagonal. All F_{ST} values are significant ($P < 0.01$).

Molecular methods
Genomic DNA was extracted from colonies with 15 µl of 1.25 mg ml^{-1} Zymolyase solution dissolved in 1.2 M sorbitol and 0.1 M KH$_2$PO$_4$ at pH 7.2 and treated with EMA to bind unwanted DNA fragments (Rueckert and Morgan, 2007). We employed a multiplex PCR reaction to distinguish *S. cerevisiae*, and this also identifies *S. uvarum* (de Melo Pereira *et al.*, 2010). DNA from eight *S. cerevisiae* colonies from each sample were initially amplified and scored at 10 unlinked loci as described by Richards *et al.* (2009) using capillary electrophoresis on an ABI3130XL (Applied Biosystems, Life Technologies, Mulgrave, VIC, Australia). If all eight initial isolates were genotypically identical then no further genotyping was performed for that sample; however, if more than one genotype was recovered, another eight were genotyped until either no new genotypes were seen, or all isolates from the sample had been genotyped. A number of control samples

were submitted for the calculation of error rates per allele and per locus as described by Pompanon *et al.* (2005). To further ascertain the reliability of microsatellite loci amplification and scoring, we analysed an additional 96-well control plate replicating the same strain for DNA extraction, PCR amplification and genotyping.

Data analyses
A ± 1-bp error in size calling from run to run variation and plus-A effects was observed and loci were binned accordingly using Genemapper (Version 4). F-statistics, migration estimates (Nm values) and Mantel tests were performed with GenAlEx (Genetic Analyses in Excel) version 6.5 (Peakall and Smouse, 2006; Peakall and Smouse, 2012). Estimates of population diversity were calculated by rarefaction (which controls for unequal sample sizes) using EstimateS (Colwell, 2006).

The maximum likelihood outcrossing rates were estimated in Mathematica 7 following the method used by Johnson *et al.* (2004) that estimates the proportion of matings between spores from the same meiotic event (that is, that are asci mates), and those from independent meiotic events (code available at http://goddardlab.auckland.ac.nz/data-and-code/). Allelic richness was estimated using rarefaction with HP rare, again controlling for unequal sample sizes, based on the lowest number of 94 observed alleles among sampled populations (Kalinowski, 2005).

Population structure was evaluated using the Bayesian clustering method implemented in InStruct, which does not assume Hardy–Weinberg Equilibrium, accounts for inbreeding, and makes no *a priori* assumptions about the sampling location of the genotypes (Gao *et al.*, 2007). This method estimates the most likely number of populations and assigns genotypes to these probabilistically. Admixture was allowed and the proportion of each genotype's ancestry in each inferred population was estimated. Three chains of one million MCMC iterations with a burn-in of 10 000 were run for $K = 1$–25. Convergence of the MCMC chain was confirmed using the Gelman–Rubin statistic (Gelman and Rubin, 1992). Analyses of the resulting ancestry profiles evaluating and quantifying the contribution of niche and geographic region to population structure was conducted with ObStruct (Gayevskiy *et al.*, 2014).

Directional migration rates were quantified using the Bayesian coalescent approach implemented in MIGRATE that assumes constant population sizes, random mating, a constant mutation rate and that populations are connected only through migration, not population divergence (Beerli and Felsenstein, 2001; Beerli, 2006, 2009; Beerli and Palczewski, 2010). Mutation-scaled population sizes (θ) were calculated using the number of sampled alleles (Haasl and Payseur, 2010). We employed a Brownian motion allele mutation model with starting estimates of the mutation-scaled migration rate derived from F_{ST} calculations to estimate all possible migration routes. Chains of one million steps with a burn-in of 50 000 were run with 10 replicates, sampling every 100 steps (Beerli, 2009). The analysis was run in parallel on the NeSI pan cluster at the University of Auckland.

Results

S. cerevisiae presence, abundance and genetic diversity
PCR analyses revealed that 3900 (36%) of the 10 716 isolates were *S. cerevisiae*. Of the 3780 isolates from spontaneous ferments, 2210 (56%) were *S. cerevisiae* and 1570 (40%) *S. uvarum*, revealing the co-existence of a sister *Saccharomyces* species in this niche. Here we do not pursue the population genetics of *S. uvarum*. *S. cerevisiae* was detected in 13 of the 37 vineyard soils, and four and one of the 36 native

soil and fruit samples, respectively. The breakdown of samples that yielded *S. cerevisiae* is shown in Supplementary Tables S1 and S2.

From control samples, two loci (YOR267C and YBR240C) amplified unreliably and were removed from all analyses. Overall the mean error rates per allele and locus were ± 4.08% and 4.35%, respectively. In total 850 individuals were genotyped, with 681 isolates from spontaneous ferments, 130 from vineyard soil, 31 from native soil and 8 from native fruits. Identical genotypes within the same sample were collapsed to conservatively account for clonal expansion during enrichment and fermentation meaning the data set was compressed to 380 genotype profiles. Just 11 genotypes matched commercially available wine strains commonly used in NZ (Richards *et al.*, 2009) and were removed from further analyses. This resulted in a final data set comprising 369 microsatellite profiles (Supplementary Data set S1). Interestingly, no genotypes matched a genetically and ecologically diverse set of international strains (Liti *et al.*, 2009) genotyped using the same method (Richards *et al.*, 2009; Goddard *et al.*, 2010).

For the entire data set, a large allelic diversity was detected at all loci. YFR028C and YML091C had the greatest diversity with 25 and 30 alleles, respectively, and all other loci had between 11 and 16 alleles. Overall, 295 different genotypes were recovered and only 38 of these were identified in more than one sample. On average samples that yielded *S. cerevisiae* contained 4.7 unique genotypes, although most alleles were shared between populations (Supplementary Data set S1). Rarefaction analyses (Chao, 1987; Colwell, 2006) estimate that these genotypes were sampled from an underlying NZ population containing ∼1700 different genotypes (with 95% confidence limits of 1159–2486).

Testing for ecological drivers of population structure
Only four different genotypes were derived from native fruits and soil, and 21 from vineyard soil (Supplementary S1 and S2). This translates to low statistical power to test how the niche of isolation affects population structure. Despite this, observations of identical genotypes between niches within, but not between, regions are striking. For example, genotypes recovered from native soil and fruit in the Martinborough region (Waiohine Gorge) were identical to spontaneous ferment isolates recovered 20 km away in Martinborough vineyards but were not found in other NZ regions; three of the genotypes isolated from vineyard soil in the Wairau Valley were identical to isolates from spontaneous ferments sourced in the same region, with one of these vineyard soil genotypes being identical to an isolate from the spontaneous ferment from the same vineyard. F_{ST} values between environmental samples (from native soil and fruit and vineyard soil) and spontaneous ferment samples within regions

are extremely low (<0.005) and insignificant ($P>0.33$) with the exception of Martinborough ($F_{ST}=0.046$, $P=0.003$); however, this F_{ST} value is classed as only representing 'low' differentiation (Wright, 1978). There is complete overlap of isolates deriving from all niches in a principal component analysis of genetic distances between genotypes (Supplementary Figure S1) and no significant population differentiation between niches within regions using InStruct and subsequent ObStruct analyses ($P>0.119$, Supplementary Figure S2). Together, this provides no substantial evidence of an effect of niche on population structure within regions. Some of the *S. cerevisiae* genotypes contributing to spontaneous ferments may have derived from wineries, as opposed to the 'environment' (Bokulich *et al.*, 2013). As all of these wineries reside within the same geographic regions the fruit was collected from, these potentially winery derived genotypes form part of the local population we wish to study. Thus, individuals from various niches within regions comprise homogenous populations, and so we combined all genotypes from different niches within regions to form regional populations for further analyses.

Testing for geographic drivers of population structure
There was significant genetic differentiation, as estimated by pairwise F_{ST} values, between populations deriving from all six regions ($P<0.01$), with the exception of those between the Wairau and Awatere Valley's ($F_{ST}=0.001$, $P=0.310$). These two valleys comprise the wider Marlborough region and were thus combined to represent one population residing in Marlborough. The subsequent pairwise F_{ST} values between regions are shown in Figure 1. A low albeit significant correlation was observed between genetic and geographic distance (Mantel Test: $R^2=0.181$, $P<0.001$). Population diversity, as estimated by rarefaction analyses to control for uneven numbers of genotypes, differs by as much as threefold between regions (Table 1). Hawke's Bay and Marlborough

harbour the greatest diversity, whereas Nelson and Central Otago the least. Allelic richness across regions is comparable to estimates within one s.d. of each other (Table 1). All eight loci in all regions are significantly out of Hardy–Weinberg equilibrium ($P<0.001$), and show strong signals for inbreeding; however, outcrossing rates are significantly above zero within each region (Table 1).

Quantifying geographic population structure
InStruct analyses (Gao *et al.*, 2007) indicate the optimal number of populations, given the data are 16. Examination of the ancestry profile plots (Rosenberg, 2004) resulting from this analysis are indicative of population structure by region to some degree (for example, the blocks of green, red and yellow in Nelson, Central Otago and Martinborough, respectively) and are in agreement with the magnitude of the pairwise F_{ST} estimates (Figure 1). Subsequent ObStruct analyses revealed that the inferred population structure is significantly correlated with geographic location ($R^2=0.16$, $P<0.0001$), and this explains about one-sixth of the genetic variability observed. Individuals from the Nelson and Central Otago regions contributed the greatest signal to overall population structure with significant decreases in the R^2 values observed when these are removed ($R^2=\Delta-0.05$ and $\Delta-0.02$, respectively). The R^2 value remained constant when data from Martinborough were removed but increased when Hawke's Bay and Marlborough data were independently removed ($R^2=\Delta+0.03$ for both). Increases in R^2 suggest individuals from these regions add noise to any signal for structure (that is, have homogenised not localised populations). Further, canonical discriminant analysis shows that ~80% of the variation in ancestry profiles can be represented with the first and second axes, suggesting that most of the variation can be visualised in these graphical representations of the data (Supplementary Figure S3). Ancestry profiles from Central Otago and Nelson cluster the most discretely in these plots,

Table 1 Summary of the populations isolated from each region

Region	Number of individuals[a]	Number of different genotypes	Estimated total number of genotypes[b]	Mean allelic richness[c]	Inbreeding Coefficient F_{IS}	Outcrossing rate[d]	Region-specific genotypes
Hawke's Bay	94	87	482: 272–932	10.72 ± 4.86	0.6483	0.12: 0.09–0.16	79/87: 91%
Martinborough	68	56	291: 150–647	11.02 ± 4.29	0.6067	0.11: 0.08–0.16	52/56: 93%
Nelson	66	48	131: 82–252	7.29 ± 1.83	0.2620	0.42: 0.32–0.54	44/48: 92%
Marlborough	91	76	415: 138–1610	10.47 ± 3.64	0.5331	0.18: 0.14–0.24	70/76: 92%
Central Otago	50	44	144: 84–293	8.29 ± 2.84	0.5869	0.14: 0.09–0.20	39/44: 89%

[a]The number of individuals included in the final analysis after the collapsing of identical individuals because of clonal expansion.
[b]The estimated number of genotypes in the population by rarefaction with 95% upper and lower rarefaction limit calculated using EstimateS (Colwell, 2006).
[c]Calculated for each locus independently using HP Rare and based on 94 genes. The mean ± 1 s.d. are reported.
[d]The average proportion of spores mating from independent tetrads per meiotic generation surrounded by 2 log likelihood support limits (equivalent to a 95% confidence interval), calculated following Johnson *et al.* (2004).

recapitulating that populations from these regions provide the strongest signals for differentiation. Pairwise comparisons between regions all significantly differ ($P < 0.001$ or $P = 0.06$ between Hawke's Bay and Marlborough), but the R^2 values vary from 0.02 to 0.23 (Supplementary Table S3).

Quantifying population connectivity and migration
Pairwise estimates of migration between the regions (Nm values) using classic methods derived from F_{ST} values (Hartl and Clark, 1997) suggest that Hawke's Bay and Marlborough are the most connected, closely followed by Marlborough and Martinborough, and Hawke's Bay and Martinborough (Figure 1). Nelson and Central Otago share the lowest number of migrants with an estimate of just one per generation (Figure 1). MIGRATE analyses showed an acceptance ratio for each parameter ranging from 0.38 to 0.65, and an effective sample size of approximately two million, suggesting that the chain length was sufficient. The autocorrelation between parameters and the prior was high and estimated to be around 0.96 overall, indicating a lack of information in data. This is reflected in the high confidence intervals surrounding the estimates (Supplementary Table S4). However, consistent patterns between multiple runs were evident, allowing meaningful estimates of gene flow between regions to be made. Inferred mean rates of movement between regions span an order of magnitude ranging from 6 to 63 migrants per generation (Figure 1 and Supplementary Table S4), and show differential inward and outward movements for some regions. Correlating with the classic Nm estimates, and the analyses of population structure, Nelson and Central Otago show greatest isolation with twofold greater rates of outward than inward migration, and show an average of just 51 inward migrants per generation, 3.2-fold less than the overall average inward migration rate of 164 migrants per generation for all other regions (Figure 1 and Supplementary Table S4). Conversely, Marlborough and the Hawke's Bay, which harbours some of the least distinctive and most diverse populations, experience some of the greatest inward migration rates at an average of 171 migrants per generation, 1.4-fold more than the average inward migration rate (Supplementary Table S4). In line with the low F_{ST} estimates, a high degree of individuals with shared ancestry from InStruct, and a large proportion of admixed individuals (Figure 1), Marlborough and the Hawke's Bay are the most connected regions, and experience an average of twofold more migration between these regions than the average overall migration rate. The extent of migration between regions does not correlate with geographic location ($P > 0.21$) showing the difference in the extent of gene flow is not simply a function of distance.

Discussion

We have very few models attempting to generally describe the population biology of microbes. Accurate quantification of short-term population level processes is necessary to understand the likely longer-term evolutionary trajectories of populations (Smadja and Butlin, 2011; Gray and Goddard, 2012), as well as how microbes may interact with other members of the community (Ruxton *et al.*, 2014). We have attempted to make a significant step forward: rather than simply describing this *S. cerevisiae* population as either structured or not, here we paint a more biologically realistic picture by quantifying the role that geography has in defining structure, and go on to provide quantitative estimates of gene flow between populations residing in different regions.

S. cerevisiae has clearly been isolated many times from managed vineyard ecosystems and ferments of fruit (Lopandic *et al.*, 2008; Liti *et al.*, 2009; Schacherer *et al.*, 2009; Goddard *et al.*, 2010; Gayevskiy and Goddard, 2012; Bokulich *et al.*, 2014). This species is also well reported from native niches in the northern hemisphere (Sniegowski *et al.*, 2002; Wang *et al.*, 2012; Hyma and Fay, 2013), and in the Southern hemisphere has been isolated from exotic *Quercus* species in NZ, and from *Nothofagus* in Patagonia (Zhang *et al.*, 2010; Libkind *et al.*, 2011). Here we provide the first report of *S. cerevisiae* from multiple native tree species in the South Pacific region. Overall, this NZ *S. cerevisiae* metapopulation displays large genetic variance, compounding evidence that NZ harbours a large and diverse population of this species (Goddard *et al.*, 2010; Gayevskiy and Goddard, 2012; Cromie *et al.*, 2013). Within regions, which typically encompass a radius of under 100 km, there is no compelling evidence for genetic differentiation between niches within managed ecosystems nor more strikingly between managed and native ecosystems. The lack of genetic differentiation between managed and native ecosystems seen here does not permit us to determine whether vineyards or native forests are the sources or sinks of these populations, just that they are connected. Thus, there appears to be a free flow of individuals between these various niches at subregional scales, supporting previous reports from NZ and the United States of America (Goddard *et al.*, 2010; Hyma and Fay, 2013). The inference of little differentiation between niches at regional scales is in contrast to previous reports showing differentiation between isolates from various ecological niches at global scales (Fay and Benavides, 2005; Legras *et al.*, 2007). One explanation for this is the extent of sample effort within any one population. The studies, including this one, reporting a minor effect of niche examined a large number of individuals from specific more localised populations, and in some sense evaluate 'ecological scale' processes: it may be that some strains are less

well adapted to various niches and that selection will eventually result in their removal. Studies evaluating strains from different geographic and ecological sources only include a handful of strains from any one specific population and unfortunately tend to confound geographic location with niche, but conclude that niche has a stronger role; in some sense these studies might examine populations where selection has possibly had more time to operate. Perhaps, the drivers of population structure differ at different scales? Lastly, it might be that NZ has relatively recently been colonised by only one of the inferred lineages of *S. cerevisiae,* and this has radiated to all niches. This would also provide a signal for the lack of differentiation between niches. Estimates of the rates of global flux for *S. cerevisiae* would help disentangle these possibilities.

Whereas populations appear homogeneous within regions, analyses provide compelling evidence for various degrees of genetic differentiation between populations inhabiting major NZ regions. This differentiation is not absolute and there is also a degree of connectivity between regions. This is in line with a previous smaller-scale study with this species that reported both differentiation and connection between regions in the North Island of NZ (Gayevskiy and Goddard, 2012), and this is also echoed at global scales (Liti *et al.*, 2009; Wang *et al.*, 2012). Here all analyses, both classic and more sophisticated Bayesian approaches, converge on the same conclusion. F_{ST}, Bayesian, ancestry profile and migration analyses show that the populations residing in Nelson and Central Otago are the most distinct and experience the least inward migration. Conversely, Marlborough and Hawke's Bay have smaller pairwise F_{ST} values and Bayesian, ancestry profile and migration analyses show that these regions are the most mixed and connected. Marlborough and Hawke's Bay experience the most inward migration at approximately three times that into Nelson and Central Otago. This is consistent with the higher genetic diversity observed in these regions and implies that they accumulate genetic diversity from around the country.

S. cerevisiae cells and spores are sessile; however, there are a variety of possible vectors that may move this unicellular eukaryote around. *S. cerevisiae* has been shown to be associated with both wasps and bees and has long been known to be associated with fruit flies (Reuter *et al.*, 2007; Goddard *et al.*, 2010; Stefanini *et al.*, 2012). Recent work provides evidence that certain volatiles released by *S. cerevisiae* attract *Drosophila*, and this enhances the likelihood of movement, and potentially facilitates a mutualism between these species (Buser *et al.*, 2014). These insect species easily move over regional scales, and so presumably have some part in the homogenisation of *S. cerevisiae* within regions. Insects less likely move *S. cerevisiae* over hundreds of kilometres between regions, although

S. cerevisiae may also be associated with birds that can easily cover these distances (Francesca *et al.*, 2012). Humans are also obvious vectors. Indeed, the patterns of separation, and rates of migration in and out the various regions shown here are nicely in line with the flow of fruit and equipment because of the actions of the NZ wine industry. Marlborough and the Hawke's Bay are the two largest viticultural and winemaking regions in the country, and fruit from other regions is often transferred to them, mirroring the inferred migration of *S. cerevisiae* into these regions. This national 'ecological' scale picture complements and mirrors the global 'evolutionary' scale picture revealed for this species: that this is a genetically diverse species that shows some degree of structure and connectivity, and these patterns are consistent with human-influenced dispersal (Fay and Benavides, 2005; Legras *et al.*, 2007).

Although the above interpretation fits nicely with the population patterns observed here, it is important to consider alternate explanations. The connections between populations could instead be indicative of recent divergence events. The NZ wine industry is very young in evolutionary terms and it is possible that *S. cerevisiae* was introduced to these regions via the introduction of vines and winery equipment such as barrels (Goddard *et al.*, 2010). The patterns observed in this analysis could be explained by the large wine-producing regions of Hawke's Bay and Marlborough being the source of variation and the outlying regions resulting from founder events with subsequent population expansion and divergence (Hartl and Clark, 1997). The method of migration analysis employed here assumes that population divergence has not occurred, and only invokes migration to explain any similarity in genetic diversity between populations (Beerli, 2009). One issue with a divergence (as opposed to migration) explanation is that source populations must exist before the populations they are proposed to have founded. Whereas Hawke's Bay is one of the oldest wine-producing regions in NZ, Marlborough is one of the youngest having only been established around 1970. Thus, the divergence hypothesis fits less well, given the vast diversity and admixture observed in the recently established Marlborough region. In addition, whilst it appears that the NZ *S. cerevisiae* population is reasonably internationally distinct, these patterns of differentiation may also be explained by the inward migration of genotypes from offshore. These explanations are not mutually exclusive and it is likely that population divergence from founding populations is occurring alongside inevitable national and international migration of strains because of the vast movement of fruit, equipment and people by the wine industry.

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 geographic origin, and this is

encapsulated in the concept of *terroir* (Bokulich *et al.*, 2014). Classically, this was thought to largely result from the interaction between specific *Vitis vinifera* varieties and the local soils, geography and climate; however, there is limited but increasing evidence showing that the microbes that influence vine growth, fermentation and wine style (as *S. cerevisiae* does) also exhibit regional differentiation (Gayevskiy and Goddard, 2012; Bokulich *et al.*, 2014; Taylor *et al.*, 2014), as we again demonstrate here. Thus, these data further support the concept that there could be a microbial aspect to *terroir*. Metabolic profiling of regionally defined genotypes is necessary to determine whether the genetic differentiation demonstrated here translates to phenotypes that are relevant to wine, and thus whether microbes contribute to *terroir* in a predictable and consistent way.

Here we provide a more advanced insight into the population biology of a well-established model microbial eukaryote that has also been biotechnologically harnessed by humans since the dawn of civilisation. We take a significant step towards quantifying these processes by providing the first estimates for metapopulation separation and similarity. We reveal *S. cerevisiae* population differentiation in NZ at scales over 100 km, with the most signal provided by the more remote regions, but no differentiation within regions, even between populations inhabiting native forests and vineyards. We also show differential migration of this species between regions, and postulate that this may be due, at least in part, to human influence. By quantifying the magnitude of these forces in microbes we begin to provide one crucial aspect of an inclusive framework attempting to more fully integrate ecological and evolutionary processes.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

References

Aa E, Townsend JP, Adams RI, Nielsen KM, Taylor JW. (2006). Population structure and gene evolution in *Saccharomyces cerevisiae*. *FEMS Yeast Res* **6**: 702–715.

Achtman M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* **62**: 53–70.

Anderson JB, Kohn LM. (1998). Genotyping, gene genealogies and genomics bring fungal population genetics above ground. *Trends Ecol Evol* **13**: 444–449.

Anderson JL, Shearer CA. (2011). Population genetics of the aquatic fungus *Tetracladium marchalianum* over space and time. *PLoS One* **6**: e15908.

Beerli P. (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**: 341–345.

Beerli P. (2009). How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? Bertorelle G, Bruford MW, Hauffe HC, Rizzoli A, Vernesi C (eds), *Population Genetics for Animal Conservation*. Cambridge University Press: New York, USA, pp 42–79.

Beerli P, Felsenstein J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach *Proc Natl Acad Sci USA* **98**: 4563–4568.

Beerli P, Palczewski M. (2010). Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **185**: 313–326.

Bokulich NA, Ohta M, Richardson PM, Mills DA. (2013). Monitoring seasonal changes in winery-resident microbiota. *PLoS One* **8**: e66437.

Bokulich NA, Thorngate JH, Richardson PM, Mills DA. (2014). Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *Proc Natl Acad Sci USA* **111**: E139–E148.

Buser CC, Newcomb RD, Gaskett AC, Goddard MR. (2014). Niche construction initiates the evolution of mutualistic interaction. *Ecol Lets* (in press).

Cavalieri D, McGovern PE, Hartl DL, Mortimer R, Polsinelli M. (2003). Evidence for *S. cerevisiae* fermentation in ancient wine. *J Mol Evol* **57**(Suppl 1): S226–S232.

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

Chao A. (1987). Estimating the population size for capture - recapture data with unequal catchability. *Biometrics* **43**: 783–791.

Colwell RK. (2006). EstimateS: Statistical estimation of species richness and shared species from samples. www.purl.oclc.ord/estimates.

Cromie GA, Hyma KE, Ludlow CL, Garmendia-Torres C, Gilbert TL, May P *et al.* (2013). Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3 (Bethesda)* **3**: 2163–2171.

de Melo Pereira GV, Ramos CL, Galvão C, Souza Dias E, Schwan RF. (2010). Use of specific PCR primers to identify three important industrial species of *Saccharomyces* genus: *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces pastorianus*. *Lett Appl Microbiol* **51**: 131–137.

Di Maio S, Polizzotto G, Di Gangi E, Foresta G, Genna G, Verzera A *et al.* (2012). Biodiversity of indigenous *Saccharomyces* populations from old wineries of south-eastern Sicily (Italy): preservation and economic potential. *PLoS One* **7**: e30428.

Dujon B. (2010). Yeast evolutionary genomics. *Nat Rev Genet* **11**: 512–524.

Fay JC, Benavides JA. (2005). Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet* **1**: 0066–0071.

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

Francesca N, Canale DE, Settanni L, Moschetti G. (2012). Dissemination of wine-related yeasts by migratory birds. *Environ Microbiol Rep* **4**: 105–112.

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.

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

Gayevskiy V, Klaere S, Knight S, Goddard MR. (2014). ObStruct: a method to objectively analyse factors driving population structure using bayesian ancestry profiles. *PLoS One* **9**: e85196.

Gelman A, Rubin DB. (1992). Inference from iterative simulation using multiple sequences. *Stat Sci* **7**: 457–472.

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.

Gray JC, Goddard MR. (2012). Gene-flow between niches facilitates local adaptation in sexual populations. *Ecol Lett* **15**: 955–962.

Green JL, Bohannan BJM, Whitaker RJ. (2008). Microbial biogeography: from taxonomy to traits. *Science* **320**: 1039–1043.

Haasl RJ, Payseur BA. (2010). The number of alleles at a microsatellite defines the allele frequency spectrum and facilitates fast accurate estimation of θ. *Mol Biol Evol* **27**: 2702–2715.

Halkett F, Simon J-C, Balloux F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends Ecol Evol* **20**: 194–201.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.

Härnström K, Ellegaard M, Andersen TJ, Godhe A. (2011). Hundred years of genetic structure in a sediment revived diatom population. *Proc Natl Acad Sci USA* **108**: 4252–4257.

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

Hittinger CT. (2013). *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet* **29**: 309–317.

Hyma KE, Fay JC. (2013). Mixing of vineyard and oak-tree ecotypes of *Saccharomyces cerevisiae* in North American vineyards. *Mol Ecol* **22**: 2917–2930.

Johnson LJ, Koufopanou V, Goddard MR, Hetherington R, Schäfer 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.

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

Libkind D, Hittinger CT, Valeŕio E, Gonçalves C, Dover J, Johnston M *et al.* (2011). Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci USA* **108**: 14539–14544.

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.

Lopandic K, Tiefenbrunner W, Gangl H, Mandl K, Berger S, Leitner G *et al.* (2008). Molecular profiling of yeasts isolated during spontaneous fermentations of Austrian wines. *FEMS Yeast Res* **8**: 1063–1075.

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, Voigt MM, Glusker DL, Exner LJ. (1996). Neolithic resinated wine. *Nature* **381**: 480–481.

Mercado L, Sturm ME, Rojo MC, Ciklic I, Martínez C, Combina M. (2011). Biodiversity of *Saccharomyces cerevisiae* populations in Malbec vineyards from the 'Zona Alta del Río Mendoza' region in Argentina. *Int J Food Microbiol* **151**: 319–326.

Mortimer R, Polsinelli M. (1999). On the origins of wine yeast. *Res Microbiol* **150**: 199–204.

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

Peakall R, Smouse P. (2012). GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research – an update. *Bioinformatics* **28**: 2537–2539.

Pompanon F, Bonin A, Bellemain E, Taberlet P. (2005). Genotyping errors: causes, consequences and solutions. *Nat Rev Genet* **6**: 847–859.

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

Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP *et al.* (2007). The role of ecological theory in microbial ecology. *Nat Rev Microbiol* **5**: 384–392.

Reuter M, Bell G, Greig D. (2007). Increased outbreeding in yeast in response to dispersal by an insect vector. *Curr Biol* **17**: R81–R83.

Richards KD, Goddard MR, Gardner RC. (2009). A database of microsatellite genotypes for

Saccharomyces cerevisiae. Antonie van Leeuwenhoek **96**: 355–359.

Rosenberg NA. (2004). Distruct: a program for the graphical display of population structure. *Mol Ecol Notes* **4**: 137–138.

Rueckert A, Morgan HW. (2007). Removal of contaminating DNA from polymerase chain reaction using ethidium monoazide. *J Microbiol Methods* **68**: 596–600.

Ruxton GD, Wilkinson DM, Schaefer HM, Sherratt TN. (2014). Why fruit rots: theoretical support for Janzen's theory of microbe–macrobe competition. *Proc R Soc B Biol Sci* **281**: 20133320.

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.

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

Smadja CM, Butlin RK. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Mol Ecol* **20**: 5123–5140.

Sniegowski PD, Dombrowski PG, Fingerman E. (2002). *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Res* **1**: 299–306.

Stefanini I, Dapporto L, Legras J-L, Calabretta A, Di Paola M, De Filippo C *et al.* (2012). Role of social wasps in *Saccharomyces cerevisiae* ecology and evolution. *Proc Natl Acad Sci USA* **109**: 13398–13403.

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

Taylor MW, Tsai P, Anfang N, Ross HA, Goddard MR. (2014). Pyrosequencing reveals regional differences in fruit-associated fungal communities. *Environ Microbiol*; doi:10.1111/1462-2920.12456.

Tsai IJ, Bensasson D, Burt A, Koufopanou V. (2008). Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. *Proc Natl Acad Sci USA* **105**: 4957–4962.

Van Der Heijden MGA, Bardgett RD, Van Straalen NM. (2008). The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol Lett* **11**: 296–310.

Wang Q-M, Liu W-Q, Liti G, Wang S-A, Bai F-Y. (2012). Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Mol Ecol* **21**: 5404–5417.

Whitaker RJ, Banfield JF. (2006). Population genomics in natural microbial communities. *Trends Ecol Evol* **21**: 508–516.

Wright S (Ed. (1978). *Variability within and among Natural Populations. Evolution and the Genetics of Populations.* University of Chicago Press: Chicago.

Xufre A, Albergaria H, Inácio J, Spencer-Martins I, Gírio F. (2006). Application of fluorescence *in situ* hybridisation (FISH) to the analysis of yeast population dynamics in winery and laboratory grape must fermentations. *Int J Food Microbiol* **108**: 376–384.

Zhang H, Skelton A, Gardner RC, Goddard MR. (2010). *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 this paper on The ISME Journal website (http://www.nature.com/ismej)