environmental microbiology
environmental microbiology reports

Saccharomyces eubayanus and Saccharomyces arboricola reside in North Island native New Zealand forests

Journal:	*Environmental Microbiology and Environmental Microbiology Reports*
Manuscript ID	EMI-2015-1154.R1
Manuscript Type:	EMI - Research article
Journal:	Environmental Microbiology
Date Submitted by the Author:	n/a
Complete List of Authors:	Gayevskiy, Velimir; University of Auckland, School of Biological Sciences Goddard, Matthew; University of Auckland, School of Biological Sciences
Keywords:	environmental genomics, genomics/functional genomics/comparative genomics, microbial ecology, fungi, bioinformatics

Title

Saccharomyces eubayanus and *Saccharomyces arboricola* reside in North Island native New Zealand forests

Authors

Velimir Gayevskiy[1] and Matthew R Goddard[1,2]

[1]School of Biological Sciences, University of Auckland, Private Bag 90219, Auckland 1142, New Zealand

[2]School of Life Sciences, The University of Lincoln, LN6 7DL, United Kingdom

Corresponding Author

Matthew Goddard

mgoddard@lincoln.ac.uk

Running Title

Saccharomyces yeasts of native NZ forests

Summary

Saccharomyces is one of the best-studied microbial genera but our understanding of the global distributions and evolutionary histories of its members is relatively poor. Recent studies have altered our view of *Saccharomyces'* origin, but a lack of sampling from the vast majority of the world precludes a holistic perspective. We evaluate alternate Gondwanan and Far East Asian hypotheses concerning the origin of these yeasts. Being part of Gondwana, and only colonized by humans in the last ~1,000 years, New Zealand represents a unique environment for testing these ideas. Genotyping and ribosomal sequencing of samples from North Island native forest parks identified a widespread population of *Saccharomyces*. Whole genome sequencing identified the presence of *S. arboricola* and *S. eubayanus* in NZ, which is the first report of *S. arboricola* outside Far East Asia, and also expands *S. eubayanus'* known distribution to include the Oceanic region. Phylogenomic approaches place the *S. arboricola* population as significantly diverged from the only other sequenced Chinese isolate, but indicate that *S. eubayanus* might be a recent migrant from South America. These data tend to support the Far East Asian origin of the *Saccharomyces*, but the history of this group is still far from clear.

Introduction

A handful of microbial species are adept at alcoholic fermentation, which is characterized by their ability to display the Crabtree effect, and it appears this trait evolved independently in at least three separate lineages separated by 500 million years (Pfeiffer et al., 2001; Piskur et al., 2006; Goddard, 2008; Rozpedowska et al., 2011; Dashko et al., 2014). The Crabtree effect serves to rapidly utilize available sugars by inefficiently converting these to ethanol, which then, in combination with a rise in temperature, has the additional effect of poisoning the environment and sabotaging competing microbes. Together these aspects of Crabtree positive microbes provide them with a competitive advantage over non-fermenting species in sugar rich niches (Pfeiffer et al., 2001; Goddard, 2008; Dashko et al., 2014). The best-known group that utilizes this strategy are seven yeast species in the *Saccharomyces* genus: *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. arboricola, S. uvarum, and S. eubayanus* (Borneman and Pretorius, 2015). *S. cerevisiae* is best known largely due to its history of association with humans who unwittingly employed its fermentative abilities over at least the last 9000 years to make wine (McGovern et al., 2004). *S. cerevisiae* has also achieved a second distinction: it is one of the best understood genetic model organisms. *S. cerevisiae* was the first eukaryote to have its genome completely sequenced and is one of the best annotated, and this species is extremely tractable for genetic manipulations and analysis (Cherry et al., 2011). Genetic and ecological interest in the rest of the group is growing (Borneman and Pretorius, 2015). However, we understand relatively little of the ecology and life history of *S. cerevisiae* and its close relatives, which are among the world's most biotechnologically important microbes (Replansky et al., 2008; Goddard and Greig, 2015). The increasing interest in *Saccharomyces* yeast has meant an increase in efforts to survey for them, and we are correspondingly seeing wider global distributions and greater genetic diversities of these species, and the increasing contribution of

other *Saccharomyces* species to beverage fermentations (Almeida et al., 2014; Bing et al., 2014; Zhang et al., 2015).

While the biogeography and global population structure of some members of the *Saccharomyces* genus, particularly *S. cerevisiae, S. paradoxus* and S. *uvarum*, have been studied (Tsai et al., 2008; Liti et al., 2009; Almeida et al., 2014; Knight and Goddard, 2015), the origin and historic movements of the group as a whole is still unclear. Two competing hypotheses regarding the origin of this group have recently been put forward: one suggests a Far East Asian origin (Wang et al., 2012; Bing et al., 2014); the other suggests a Gondwanan one (Almeida et al., 2014; Peris et al., 2014). In support of the Far East Asia hypothesis is the fact that the greatest species diversity is found in this area given the sampling efforts so far: all species in the group have been uncovered here, with *S. mikatae* and *S. arboricola* not yet reported elsewhere (Naumov et al., 1997; 2000; Wang and Bai, 2008; Wang et al., 2012; Naumov et al., 2013; Almeida et al., 2014; Bing et al., 2014). In addition, the genetic diversity of *S. eubayanus*, *S. paradoxus*, *S. kudriavzevii* and *S. cerevisiae* are the same or higher in Far East Asian populations than elsewhere (Tsai et al., 2008; Hittinger et al., 2010; Bing et al., 2014). Lastly, for *S. cerevisiae*, genotypes that are inferred to comprise basal clades have been isolated from China (Wang et al., 2012). However, weighing in for the Gondwanan hypothesis is the fact that several diverse and abundant populations of the two basal *Saccharomyces* species - *S. uvarum* and *S. eubayanus* - have been extensively recovered in Patagonia (South America) (Almeida et al., 2014; Peris et al., 2014). *S. uvarum* in particular was found to be highly diverse within Patagonia and Australasia compared with northern hemisphere isolates from North America, Europe and Far East Asia (Almeida et al., 2014). The widespread nature of these basal species and large genetic divergence has been used as evidence to support the alternate idea that the *Saccharomyces* originated in Gondwana: the group is then hypothesized to have

subsequently moved to Southeast Asia where it further diverged and speciated (Almeida et al., 2014).

Being geographically isolated but originally part of Gondwana makes New Zealand a prime location to evaluate these competing ideas concerning the origins of this *Saccharomyces* group, and may also help shed light on patterns of their global dispersal. New Zealand has only recently been populated by humans in the last 1,000 years, (Hurles et al., 2003) and Europeans only arrived with grape vines around 200 years ago. Previous work in New Zealand has revealed a widespread population of *S. cerevisiae* in both winemaking and natural environments (Goddard et al., 2010; Gayevskiy and Goddard, 2012; Knight and Goddard, 2015), but these have been shown to be recent and recurring arrivals, most likely vectored by humans (Goddard et al., 2010). In addition, *S. paradoxus* has been found in NZ on exotic oak trees, which themselves came from Europe, and the NZ strains very closely resemble those from European populations, again strongly suggesting introduction via human activity (Zhang et al., 2010). Lastly, *S. uvarum* has also been reported in NZ associated with winemaking, as well as in native forests (Almeida et al., 2014; Knight and Goddard, 2015). To date, none of the other *Saccharomyces* have been reported in NZ. The first and only suggestion for an ancient *Saccharomyces* population in the Pacific area comes from evidence supporting the Gondwanan hypothesis showing that New Zealand and Australia harbour a genetically-distinct population of *S. uvarum* (Almeida et al., 2014). The Gondwanan origin hypothesis predicts that NZ should harbour a diversity of endemic genetically distinct populations of *Saccharomyces* species, and the Far East Asian origin hypothesis predicts that NZ should not, and if species are present, that they represent exotic populations that have moved to NZ from elsewhere. Here we aim to evaluate these competing hypotheses by surveying NZ native forests for *Saccharomyces* and in doing so also hope to add to the body of knowledge on the global distribution and evolutionary history of the academically and commercially important *Saccharomyces* yeasts.

Results

Molecular Species Identification and Distributions

A total of 442 fruit, soil and bark samples were collected from remote native New Zealand forest conservation reserves across the North Island in 2012, spanning 100 by 160 kms, approximately centred around -39º S, 176º E (Fig S1). After placing in media designed to select for Crabtree positive microbes, 150 (34%) samples showed signs of fermentation and were plated for colony selection: 97 (22%) of these yielded colonies with 80 from soil, 15 from fruit, and 2 from bark. One to eight colonies were selected for molecular analyses from each of the 97 samples, totalling 731 isolates evaluated. RFLP analysis of the ribosomal ITS region was employed to classify isolates into putative species groups for fungi. Some isolates failed to amplify with fungal primers and we were curious to ascertain what these were, and so amplified the bacterial 16S ribosomal area for these. This simple method identified four fungal and three bacterial groups, and these were present in 57, 21, 1, 3 and 15, 14 and 2 samples respectively (due to clonal expansion during enrichment we simply report presence of each group in samples); see Figure S2. Sanger sequencing of the D1/D2 region of the 26S rDNA for specific fungal isolates (see Table S1 for accession numbers) showed ten isolates from YG1 had identical sequences that matched 100% to the sequence from the type strain of *Saccharomyces arboricola* (AS 2.3317; accession: EF580918); one YG1 isolate was identical to the sequence from the type strain of *S. paradoxus* (NRRL Y-17217; accession: AY048155); and one YG1 isolate matched 100% to the neotype strain of *Saccharomyces pastorianus* (NRRL Y-27171; accession: U68547) and *S. bayanus* (NRRL Y-12624, U94931), because the two hybrid species share 100% identical D1/D2 sequences. Seven isolates from YG2 fungal group yielded identical sequences that matched 100% to *Lachancea cidri* (originally *Zygosaccharomyces cidri;*

(Kurtzman, 2003). Two isolates from YG3 and three from YG4 matched with between 99.3%-100% pairwise identities to various species, most of them *Saccharomyces*: *Saccharomyces arboricola* (YG3), *Saccharomyces paradoxus* (2 strains; YG3 and YG4), *Kazachstania servazzii* (YG4) and *Saccharomyces cerevisiae* (YG4). Three strains from the bacterial group BG1 matched 99.7% or more to *Leuconostoc mesenteroides* at the 16S rDNA locus. 16S sequence provided no agreement for the five isolates assigned to groups BG2 and BG3, as they variously matched *Lactobacillus sakei*, *Bacillus cereus* and *Bacillus weihenstephanensi, Viridibacillus arvi* and to an unclassified *Bacillus* species.

This suite of analyses allows us to tentatively ascribe species names to these RFLP groups and more accurately identify the 31 isolates for which we directly sequenced ribosomal areas. YG1 most likely represents members of an *S. arboricola* population, but may also include *S. paradoxus*; YG2 more reliably represents members from a *Lachancea cidri* population; groups YG3 and YG4 most likely represent a mix of *Saccharomyces* species. Members from YG1, YG3 and YG4 were found in 80% of sites where yeast colonies were recovered, meaning that 17% of the total samples collected harboured some *Saccharomyces*, and that *Saccharomyces* was present in every niche we examined and in all but one forest park we sampled (Fig S2). Given this, it is clear that we may conclude that *Saccharomyces* species are widespread, but not abundant, in these North Island native New Zealand Forests, especially in light of the fact that *Saccharomyces* are typically found at very low incidences (Sniegowski et al., 2002; Wang et al., 2012; Naumov et al., 2013). Further, since 10 of the 12 isolates analysed in the most abundant *Saccharomyces* group (YG1) were identified as *S. arboricola*, and members of this RFLP group were found in 59% of samples which spanned all three niches and six of the nine forests surveyed, this allows us to tentatively suggest that *S. arboricola* is the most widespread *Saccharomyces* species in native North Island forests of New Zealand.

Identification by Genome Sequencing and Mapping

Here, we carry out whole genome sequencing on ten *Saccharomyces* species isolates that have not been previously reported in the native forests of New Zealand. Reflecting their recovery abundances, we analysed: eight *S. arboricola* isolates from maximally divergent sampling locations and niches; one *S. paradoxus;* and one *S. pastorianus* isolate. *S. arboricola* has not been isolated outside of Far East Asia and was the most abundant species in our samples. *S. paradoxus* has been isolated in northern hemisphere forests (Naumov et al., 1997; Sniegowski et al., 2002; Johnson et al., 2004) and shows strong signals of population structure by geographic origin (Liti et al., 2009). *S. paradoxus* has been isolated from exotic oak trees in NZ (Zhang et al., 2010), but it has not been isolated from native tree species in the southern hemisphere previously. *S. pastorianus* is a hybrid of *S. cerevisiae* and *S. eubayanus* and is commonly used in commercial lager fermentations (Libkind et al., 2011). It has not been previously isolated from natural environments and is not well adapted to them (Hittinger, 2013), suggesting this initial identification may be incorrect.

Whole genome sequencing of these 10 *Saccharomyces* yeasts yielded an average of 1.6 million 250bp paired-end reads per strain for a total of 4Gb of high quality data. To accurately identify these isolates, each of the 10 genomes was mapped against the six non-hybrid *Saccharomyces* species with genome assemblies available (strains in brackets): *S. cerevisiae* (S288C; Goffeau et al., 1996; EBI: GCA_000146045.2), *S. paradoxus* (CBS432; Liti et al., 2009), *S. mikatae* (IFO1815; Scannell et al., 2011), *S. kudriavzevii* (IFO1802; Scannell et al., 2011), *S. arboricola* (H-6; Liti et al., 2013) and *S. uvarum* (CBS7001; Scannell et al., 2011). *S. eubayanus* does not have an assembled genome available for mapping and was thus not initially included. An average mapping rate of 88.8% (SE=0.6%) and mapping quality (Phred score) of 56.5 (SE=2.5) was obtained for the best match per genome (see Table S2). The eight isolates classified as *S.*

arboricola given their ribosomal sequences best mapped to *S. arboricola*. However, the isolate initially identified as *S. paradoxus* by ribosomal sequence (P8F5; accession KP979610 see Table S1) only mapped 69% to *S. paradoxus,* but mapped best to *S. arboricola* (90%; see Table S2). We retrieved the D1/D2 26s sequence from this isolate (P8F5) from the Illumina data we obtained, and this was identical to the *S. arboricola* reference sequence EF580918.1. The conclusion is that the original Sanger sequence matching *S. paradoxus* was a contaminant or error. The isolate identified as *S. pastorianus* from ribosomal sequence (P1C1) mapped best to *S. uvarum* but with a slightly lower rate (86%).

To corroborate these initial identifications, and more accurately place the New Zealand derived genomes within the broader *Saccharomyces* genus, we extracted a well-known set of 106 orthologous loci (Rokas et al., 2003) dispersed through the genomes of all *Saccharomyces*, and included these loci from the unassembled *S. eubayanus* genome (Libkind et al., 2011). We then used these loci, comprising 194,384 bps, to construct a phylogeny using a Maximum likelihood approach (Guindon and Gascuel, 2003); Figure 1. The nine genomes that mapped best to *S. arboricola* all form a tight clade along with the *S. arboricola* reference, which falls as an outgroup within this clade. This positioning unequivocally classifies these NZ isolates as *S. arboricola*. The remaining New Zealand genome initially identified as *S. pastorianus* from the ribosomal Sanger sequence, that mapped best to *S. uvarum,* clusters very tightly with the only available *S. eubayanus* sequences with a short branch length and thus classifies it as *S. eubayanus*. Thus, these genomic sequence data and phylogenetic analyses confirm the presence of both *S. arboricola* and *S. eubayanus* in New Zealand.

Are the NZ *Saccharomyces* endemic or introduced?

The presence of *Saccharomyces* in NZ (or elsewhere for that matter) does not mean these represent members of ancient endemic populations. While there is support for an ancient *S. uvarum* population across Australasia, there is good evidence showing that at least some *S. cerevisiae* and *S. paradoxus* have been very recently introduced into NZ, likely by humans (Zhang et al., 2010; Knight and Goddard, 2015). There is only one other *S. arboricola* genome available, and so we cannot attempt to infer the origin of the nine *S. arboricola* genomes from NZ. However, the average pairwise number of whole-genome SNPs separating the New Zealand *S. arboricola* from one another is 10,676±566 (0.09% of the genome) compared with an average of 307,676±167 (2.6% of the genome) separating them from the Chinese reference, allowing us to simply state that the NZ isolates significantly deviate from the only other available genome from China. The recent debate on the origins of the *Saccharomyces* genus has yielded geographically dispersed sequence data for *S. eubayanus*, but we fall into a similar problem as we only have one NZ *S. eubayanus* genome. However, we reconstructed a novel phylogeny using the 56 available *S. eubayanus* strains, overlaying the six loci they have in common (FSY1,1210bp; FUN14, 429bp; GDH1+intergenic region before the gene, 679bp; HIS3, 534bp; MET2, 469bp; and RIP1, 509bp; totaling 3,830 bases). These strains come from West China, Patagonia, North America, and the Tibetan *S. uvarum* as an outgroup (Figure 2). The phylogeny clearly shows the previously published lineages of *S. eubayanus* from West China, Patagonia (A & B) and Tibet, along with the admixed population in North America (Bing et al., 2014; Peris et al., 2014). The New Zealand isolate is placed within the Patagonia B group and displays only a small level of divergence from the rest of this group.

Discussion

Fermentative Species of Native New Zealand Forests

To our knowledge, there has been no other widespread sampling effort for *Saccharomyces* yeasts in Pacific region native forests. The present work has identified a diversity of microbial species capable of withstanding the toxic effects of ethanol. Samples principally contained yeasts (19% of all samples) instead of bacteria (7% of all samples), likely due to the use of enrichment media. The isolation and selection protocol was designed to select for *Saccharomyces* species so it is not surprising that bacteria were found to a lesser extent.

While not abundant components of microbial communities, one striking result of our sampling effort is the widespread range of *Saccharomyces* yeasts in the North Island native forests of New Zealand. *Saccharomyces* sp. were present in 13% of samples overall, present in all three niches, and in four of the five forest parks sampled. Whole genome sequencing identified two *Saccharomyces* species previously not known to occur in New Zealand, and has expanded the known global ranges of *S. arboricola* and *S. eubayanus*. *S. arboricola* has not been identified outside of Far East Asia while *S. eubayanus* has only previously been reported in the Americas (Peris et al., 2014; Rodríguez et al., 2014) and in Far East Asia (Bing et al., 2014).

The practical constraints of the sample processing meant we enriched at a room temperature of around 23ºC. It is well described that *Saccharomyces* species have differential thermal growth profiles, at least for members of other populations that have been studied (Salvado et al., 2011). Thus, 23ºC may prejudice in favour of certain species over others, and here we may only comment on the incidence of the *Saccharomyces* species we recovered, and not necessarily draw any inference concerning those species we did not recover, or the differential frequencies of species recovery. Not withstanding this, *S. arboricola* appears the most abundant and widespread of the *Saccharomyces* species in NZ native forests, and this widespread and relatively common distribution is in contrast to isolation frequency in Far East Asia, where just

four strains were recovered from hundreds of soil, mushroom and plant samples in China and Taiwan (Wang and Bai, 2008; Naumov et al., 2013). Given the similarity of our sampling protocol and sampled niches with the studies reporting *S. arboricola* in Far East Asia, it is tempting to conclude this species is more common in New Zealand than in Far East Asia: further work will reveal if this is actually the case. Despite being described in 2008 (Wang and Bai, 2008), no other reports of *S. arboricola* outside of Far East Asia have been published to our knowledge. Further sampling is clearly required to characterize the global distribution of this species. We recovered only a single isolate of *S. eubayanus*, resulting in a low estimation of its prevalence in NZ. Other than *S. cerevisiae*, the only other *Saccharomyces* species identified in New Zealand's native forests to date is *S. uvarum,* which is *S. eubayanus*' sister species (Almeida et al., 2014). This makes the presence of *S. eubayanus* in NZ not altogether surprising but does expand its known distribution to include the Oceanic region and therefore match that of *S. uvarum* (Almeida et al., 2014). *Saccharomyces* species were 5-fold (Figure S2b) more abundant in soil than fruit, showing these species are capable of inhabiting niches other than fruit regardless of the Crabtree effect, or perhaps that the Crabtree effect is not an adaption to fruits, but a trait conferring adaptation to a range of niches more generally, or possibly not an adaptation at all (Goddard and Greig, 2015).

Lachancea cidri was reasonably abundant (22% of samples with microbial growth), and has been found in the high-sugar New Zealand *Nothofagus* honeydew system (Serjeant et al., 2008). Unsurprisingly given its name, this species has also been isolated from cider (Coton et al., 2006), sherry (Esteve-Zarzoso et al., 2001) and wine (Ganga and Martinez, 2004) fermentations worldwide indicating it is widespread, but the phylogeography of this species is not described. The remaining yeast species identified was *Kazachstania servazzii* and has been found in South America on *Araucaria araucana* seeds (Rodríguez et al., 2014), and been

isolated from sourdough (Di Cagno et al., 2014) and kimchi (Kobayashi et al., 2013) fermentations, as well as on wine grapes in Denmark (Lederer et al., 2013). Unsurprisingly given the ethanol selection pressure in the isolation media, all three genera found are known to be Crabtree-positive (Hagman et al., 2013).

The origins of NZ *Saccharomyces*

In this study whole genome sequencing has confirmed the presence of *S. arboricola* and *S. eubayanus* in NZ, and the presence of *S. cerevisiae* in previous studies (Knight and Goddard, 2015); in addition, multi-locus sequencing has also confirmed the presence of *S. paradoxus* (Zhang et al., 2010) and *S. uvarum* (Almeida et al., 2014; Knight and Goddard, 2015; Zhang et al., 2015). There is some evidence to suggest at least some NZ *S. cerevisiae* derived from Europe (Goddard et al, 2010). In addition, the *S. paradoxus* isolates recovered in NZ derived from exotic oak trees, which were brought in from Europe by humans as acorns, and the sequence data deriving from the *S. paradoxus* on these now mature oaks suggests these yeast also came from Europe (Zhang et al., 2010).

Phylogenomic analyses of the nine New Zealand *S. arboricola* isolates showed they form a tight clade, and were more similar to one another than to the basal *S. arboricola* strain from China. Based on this very limited information, this level of divergence suggests restricted gene flow between New Zealand and Far East Asian *S. arboricola*. Of note, the level of genetic divergence between these geographically disparate *S. arboricola* closely mirrors that for *S. paradoxus*. A population of 12 *S. paradoxus* sampled from two parks in the United Kingdom had, on average, ~10,000 pairwise SNP differences (0.09% of the genome), but an average ~450,000 (3.75% of the genome) from a North American isolate (Bergstrom et al., 2014). *S. paradoxus* is known to

be geographically delimited (Tsai et al., 2008; Liti et al., 2009) and our results tentatively indicate this is could be the case for *S. arboricola* as well.

The single NZ *S. eubayanus* isolate we sequenced sits inside the poorly resolved Patagonia B group for this species (Peris et al., 2014), and the lack of significant genetic divergence indicates this isolate is closely related to, and thus recently originated from, this population. In this respect, this observation tentatively suggests that at least this NZ *S. eubayanus* isolate might share a similar history to the NZ *S. cerevisiae* and *S. paradoxus* populations generally, i.e. it is a recent migrant. Interestingly, our novel combination of data from these three locations places the industrial lager strain CBS1503 (*S. pastorianus*/*S. carlsbergensis*) as most related to the Tibetan *S. eubayanus* clade. This is in line with the assertion by (Bing et al., 2014) that commercial lager strains likely have their origin in Far East Asia rather than in South America, which was suggested by (Peris et al., 2014). The remaining structure in the global *S. eubayanus* population appears to be the product of multiple long-distance dispersal events and further data are required to identify these events, likely dates and directions.

Evaluating the Gondwanan/Far East Asian Origin Hypotheses

The Far East Asian origin hypothesis for the *Saccharomyces* genus is based on this region containing the greatest number of *Saccharomyces* species (Bing et al., 2014) and genetic diversity (Wang et al., 2012); the Gondwanan hypothesis is based on the observation that several diverse and abundant populations of the two basal *Saccharomyces* species - *S. uvarum* and *S. eubayanus* - have been extensively recovered in Patagonia (Libkind et al., 2011; Rodríguez et al., 2014; Almeida et al., 2014). The prediction arising from the Gondwanan origin hypothesis is that NZ should harbour a greater diversity of endemic genetically distinct basal *Saccharomyces* species. The Far East Asian origin hypothesis predicts that NZ should not

contain these, and the presence of any species will be due to more recent migration. Five of the seven *Saccharomyces* are present in NZ. On the face of it this apparent NZ species richness might tend to support the Gondwanan origin hypothesis. However, the key to evaluate this fairly is to determine whether these isolates represent ancient endemic populations, or are recent immigrants into NZ. The results of sampling in New Zealand, and the subsequent phylogenetic and phylogenomic analyses indicate that: *S. cerevisiae, S. paradoxus, S. uvarum, S. eubayanus* and *S. arboricola* are present in NZ; *S. cerevisiae, S. paradoxus* and possibly *S. eubayanus* have been recently introduced; that *S. uvarum* appears to be more ancient; and, *S. arboricola* is relatively widespread and different from the one Chinese isolate that has been sequenced. Thus, for at least three of the five *Saccharomyces* species present in NZ, we have evidence that the isolates discovered so far do not represent endemic populations. While the NZ *S. uvarum* population appears ancient (Almeida et al., 2014), the single isolate of *S. eubayanus* is very closely related to the Patagonian B isolates. Further, our analyses of *S. eubayanus* generally suggest that the Chinese group is basal for this species, and the South and North American isolates are more derived and intermingled with Tibetan ones, suggesting transfer from Asia to the Americas, not *vice versa*. Together, these observations tend to reject the Gondwanan origin hypothesis, and support the Far East Asian origin hypothesis for the *Saccharomyces*.

Clearly, further sampling in NZ and elsewhere might uncover different species diversities and populations with different histories, and then again it may not. Given the data we have, on balance we cannot convincingly support or reject either hypotheses, and it is clear that we do not yet have an accurate understanding of the global diversity and relatedness within the *Saccharomyces* genus. Indeed, it seems that every major sampling effort in forests distant from humans over the last 5 years has uncovered unexpected populations of *Saccharomyces* yeasts that do not fit previous hypotheses relating to the origin of the genus. The majority of sampling in North America and Europe has been from either human-associated environments or forests

nearby to these, and the species of focus have been *S. cerevisiae* and/or *S. paradoxus* (Naumov et al., 1998; Sniegowski et al., 2002; Johnson et al., 2004; Hyma and Fay, 2013). More recent studies in South America and Far East Asia have aimed to characterize the full diversity of *Saccharomyces* present and have used lower incubation temperatures and rapid genotyping to identify and compare large numbers of isolates (Wang et al., 2012; Peris et al., 2014; Bing et al., 2014; Almeida et al., 2014). *Saccharomyces* has been recently called the "premier model genus" (Hittinger, 2013) and it is certainly fulfils many of the requirements for this. However, we suggest greater knowledge of the ecology, distributions, past evolutionary history and migration within this genus is fundamental to it obtaining this status. The recent extensive sampling efforts in South America and now New Zealand have uncovered the presence of *Saccharomyces* species, with a confusing mix of apparently more ancient diverse groups of species intermingled with groups of species that have seemingly been recently moved, probably by humans. Together this information does not allow one to categorically accept or reject either the Far East Asian or Gondwanan origin hypotheses, or even allow one to sensibly suggest an alternative, indicating the evolutionary history and origin of the *Saccharomyces* is not yet resolved.

Experimental Procedures

Sample Sites and Processing

Sampling was conducted in five geographically disparate forest parks in the centre of North Island, New Zealand in May 2012: Kaimanawa Forest Park, Kaweka Forest Park, Ruahine Forest Park, Te Urewera National Park and Tongariro National Park (Figure S1). Sample sites were located as far as possible from roads and human settlements but were adjacent to public walking tracks. All sample sites were at least 50km away from the nearest winemaking region.

At each sample site, up to 20 native fruiting trees 2-10 meters from the track were separately sampled for ~10mL of fruit, soil and bark into 50mL sterile tubes using forceps, scalpels and scissors sterilized in 70% ethanol for at least 1 minute between samples. Tree genera sampled include (but are not limited to): *Pseudopanax, Coprosma, Rhopalostylis, Leptospermum, Kunzea, Leptosporum, Griselinea, Prumnopitys* and *Schefflera*. In cases where fruiting trees could not be readily located, a non-fruiting tree was substituted and the fruit sample was replaced with a second soil sample. Samples were collected over the course of 3 days and transported to the laboratory at ambient temperature for enrichment.

Samples were placed in 30mLs enrichment media: 1% yeast extract, 2% peptone, 10% glucose and 8% ethanol (added after autoclaving) at room temperature (~23˚C). A release of pressure was conducted after 10 days and thereafter based on signs of fermentation. After 4 weeks, those samples showing signs of fermentation (a layer of white on the sample, fizzing and/or cloudiness not caused by the sample) were diluted and serial plated onto YPD agar (1% yeast extract, 2% peptone, 2% glucose and 1.5% agar). Samples with visible colonies after a week

had 8 colonies selected to maximize phenotypic diversity and these were stored in 15% glycerol at -80°C for molecular analysis.

Species Identification

Initial molecular identification was carried out using RFLP analysis. DNA was extracted using Zymolyase and amplified for the internal transcribed spacer region (ITS1-5.8rRNA-ITS2) using the ITS1 and ITS4 primers (White et al., 1990). Isolates not yielding bands were microscopically classified as bacteria and the 16S ribosomal DNA region was amplified using the universal 1492R/616V primer pair (Polz and Cavanaugh, 1998; Spring et al., 1998). RFLP digests were carried out using HaeIII and HinfI for the ITS PCR products and HaeIII and HindIII for 16S products. Each isolate was classified into a tentative species group based on its RFLP patterns and ITS/16S fragment size.

Twenty-three isolates spanning these RFLP groups were selected for two-way 26S or 16S ribosomal sequencing in proportions corresponding to their abundances in samples. Sequences were aligned using MUSCLE (Edgar, 2004) within Geneious (Biomatters Ltd., 2012) and manually validated for errors. Blastn searches were performed on the resulting sequences against Genbank and the top ten hits with 98%+ identities were used for identification. These sequences have the accession numbers KP979610-KP979640.

Whole Genome Sequencing of *Saccharomyces spp*.

High molecular weight genomic DNA was extracted using the Qiagen Blood & Cell Culture DNA Kit. Libraries were constructed using the Illumina TruSeq Nano DNA Sample Prep Kit with 550bp insert size, and sequencing was carried out using a 2 x 250 bp paired-end Illumina

MiSeq run. We have made the raw sequences publicly available through the Sequence Read Archive with the accession SRP048568.

Genome Quality Control and Mapping

Quality control of the raw data was performed using FASTQC v0.11.2 (Andrews, 2012). Given conflicting species identifications of yeast groups using 26S Sanger sequencing, we mapped all 10 genomes against the six *Saccharomyces* species with genome assemblies available (strains in brackets): *S. cerevisiae* (S288C), *S. paradoxus* (CBS432), *S. mikatae* (IFO1815), *S. kudriavzevii* (IFO1802), *S. arboricola* (H-6) and *S. uvarum* (CBS7001). Mapping was carried out using BWA-MEM (v.7.8, Li, 2013). Following mapping, samtools (v0.1.18; Li et al., 2009) was used for alignment conversion, sorting and indexing. Mapping quality reports were generated using Qualimap (García-Alcalde et al., 2012) and further analyses of genomes were conducted on the alignment with the highest mapping rate. A variant call file was produced using the mpileup command within samtools with the "-Bu" parameters, which disregards the reference genome in making genotype calls. The variant call file was used to create a consensus genome using the vcf2fq Perl script within samtools. Genomic alignments were carried out using Mauve (Darling et al., 2004) within Geneious (Biomatters Ltd., 2012). We conservatively removed alignment columns containing a gap or unknown nucleotide in at least one of the genomes to reduce sequencing and mapping biases. The number of SNPs within the remaining positions was quantified within Geneious.

Saccharomyces Phylogenomics

We compiled a set of genomes from the seven *Saccharomyces* species. No more than five genomes were collected per species and these were as geographically and/or genetically diverse as possible. Some species did not have five whole genomes available and in these

cases we used every available genome resulting in a total of 33 genomes. Table S1 provides a list of genomes and their geographic origins. We extracted a well-known set of 106 orthologous loci dispersed through the genomes of all *Saccharomyces* and at least back to *Candida albicans* from each genome (Rokas et al., 2003). The *S. cerevisiae* S288C sequence for each of these loci was used as the query in a Discontinuous MegaBLAST search against each of the genomes. Phylogenomic analyses was conducted using PhyML (v2.2; Guindon and Gascuel, 2003) within Geneious (V8.1.2; Biomatters Ltd., 2012). The final phylogenetic tree visualized in FigTree (Rambaut, 2012).

One isolate was putatively identified as *S. eubayanus* and we infer its ancestry by comparing with two recently published datasets including 21 South American and North American strains (Peris et al., 2014); and 37 Far East Asian strains (Bing et al., 2014). We obtained the multilocus sequences from these publications and determined six loci shared between them (FSY1, FUN14, GDH1, HIS3, MET2 and RIP1). After generating consensus sequences for each of these loci, we performed a BLAST search against the putative NZ *S. eubayanus* genome to retrieve sequences. A multiple alignment was conducted using MUSCLE (Edgar, 2004) within Geneious (V8.1.2; Biomatters Ltd., 2012) and any alignment columns with more than 5% missing data were removed due to unequal loci lengths between the studies. PhyML was again employed to conduct phylogenetic analysis. Alignments and tree files are publically available at http://goddardlab.auckland.ac.nz/

Acknowledgements

We thank Michelle Kelly for help with sampling and initial sample processing, and two anonymous reviewers for helpful comments. We also thank the New Zealand Department of Conservation for managing New Zealand's natural assets and granting us permission to conduct sampling in national/forest parks (Permits: TW-33762-FLO, BP-33788-RES and WE-31992-FLO).

References

Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarède, I., et al. (2014) A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat Commun* **5**: 4044.

Andrews, S. (2012) FastQC: A quality control tool for high throughput sequence data.

Bergstrom, A., Simpson, J.T., Salinas, F., Barre, B., Parts, L., Zia, A., et al. (2014) A high-definition view of functional genetic variation from natural yeast genomes. *Molecular Biology and Evolution* **31**: 872–888.

Bing, J., Han, P.-J., Liu, W.-Q., Wang, Q.-M., and Bai, F.-Y. (2014) Evidence for a Far East Asian origin of lager beer yeast. *Curr. Biol.* **24**: R380–R381.

Biomatters Ltd. (2012) Geneious.

Borneman, A.R. and Pretorius, I.S. (2015) Genomic Insights into the Saccharomyces sensu stricto Complex. *Genetics* **199**: 281–291.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., et al. (2011) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Research* **40**: D700–D705.

474 Coton, E., Coton, M., Levert, D., Casaregola, S., and Sohier, D. (2006) Yeast ecology in French
475 cider and black olive natural fermentations. *International Journal of Food Microbiology* **108**:
476 130–135.
477 Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T. (2004) Mauve: multiple alignment of
478 conserved genomic sequence with rearrangements. *Genome Research* **14**: 1394–1403.
479 Dashko, S., Zhou, N., Compagno, C., and Piskur, J. (2014) Why, when, and how did yeast
480 evolve alcoholic fermentation? *Fems Yeast Res* **14**: 826–832.
481 Di Cagno, R., Pontonio, E., Buchin, S., De Angelis, M., Lattanzi, A., Valerio, F., et al. (2014)
482 Diversity of the lactic acid bacterium and yeast microbiota in the switch from firm- to liquid-
483 sourdough fermentation. *Applied and Environmental Microbiology* **80**: 3161–3172.
484 Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high
485 throughput. *Nucleic Acids Research* **32**: 1792–1797.
486 Esteve-Zarzoso, B., Peris-Toran, M.J., Garcia-Maiquez, E., Uruburu, F., and Querol, A. (2001)
487 Yeast population dynamics during the fermentation and biological aging of sherry wines.
488 *Applied and Environmental Microbiology* **67**: 2056–2061.
489 Ganga, M.A. and Martinez, C. (2004) Effect of wine yeast monoculture practice on the
490 biodiversity of non-*Saccharomyces* yeasts. *J Appl Microbiol* **96**: 76–83.
491 García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., et al.
492 (2012) Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**:
493 2678–2679.
494 Gayevskiy, V. and Goddard, M.R. (2012) Geographic delineations of yeast communities and
495 populations associated with vines and wines in New Zealand. **6**: 1281–1290.
496 Goddard, M.R. (2008) Quantifying the complexities of *Saccharomyces cerevisiae*'s ecosystem
497 engineering via fermentation. *Ecology* **89**: 2077–2082.
498 Goddard, M.R. and Greig, D. (2015) *Saccharomyces cerevisiae*: a nomadic yeast with no
499 niche? *Fems Yeast Res*. 15, fov009

500 Goddard, M.R., Anfang, N., Tang, R., Gardner, R.C., and Jun, C. (2010) A distinct population of
501 *Saccharomyces cerevisiae* in New Zealand: evidence for local dispersal by insects and
502 human-aided global dispersal in oak barrels. *Environmental Microbiology* **12**: 63–73.
503 Goffeau, A, B G Barrell, H Bussey, R W Davis, B Dujon, H Feldmann, F Galibert, et al. (1996)
504 Life with 6000 Genes. *Science* **274**: 546–67.
505 Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large
506 phylogenies by maximum likelihood. *Systematic Biology* **52**: 696–704.
507 Hagman, A., Säll, T., Compagno, C., and Piskur, J. (2013) Yeast “make-accumulate-consume”
508 life strategy evolved as a multi-step process that predates the whole genome duplication.
509 *PLoS ONE* **8**: e68734.
510 Hittinger, C.T. (2013) *Saccharomyces* diversity and evolution: a budding model genus. *Trends*
511 *Genet.* **29**: 309–317.
512 Hittinger, C.T., Gonçalves, P., Sampaio, J.P., Dover, J., Johnston, M., and Rokas, A. (2010)
513 Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* **464**:
514 54–58.
515 Hurles, M.E., Matisoo-Smith, E., and Gray, R.D. (2003) Untangling Oceanic settlement: the
516 edge of the knowable. *Trends in Ecology & Evolution* **18**: 531–540.
517 Hyma, K.E. and Fay, J.C. (2013) Mixing of vineyard and oak-tree ecotypes of *Saccharomyces*
518 *cerevisiae* in North American vineyards. *Molecular Ecology* **22**: 2917–2930.
519 Johnson, L.J., Koufopanou, V., Goddard, M.R., Hetherington, R., Schäfer, S.M., and Burt, A.
520 (2004) Population genetics of the wild yeast *Saccharomyces paradoxus*. *Genetics* **166**: 43–
521 52.
522 Knight, S. and Goddard, M.R. (2015) Quantifying separation and similarity in a Saccharomyces
523 cerevisiae metapopulation. *ISME J* **9**: 361–370.
524 Kobayashi, M., Kawamura, M., Kobayashi, A., Yamazaki, A., and Fukuda, M. (2013) Isolation of
525 yeasts from kimchi and construction of a hydrostatic-pressure sensitive mutant. *The Review*

526 *of High Pressure Science and Technology* **23**: 59–67.
527 Kurtzman, C.P. (2003) Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and
528 other members of the *Saccharomycetaceae*, and the proposal of the new genera
529 *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygotorulaspora*. *Fems*
530 *Yeast Res* **4**: 233–245.
531 Lederer, M.A., Nielsen, D.S., Toldam-Ansersen, T.B., J, H., V, and Arneborg, N. (2013) Yeast
532 species associated with different wine grape varieties in Denmark. *Acta Agriculturae*
533 *Scandinavica, Section B - Soil & Plant Science* **61**: 89–96.
534 Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
535 1–3.
536 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009) The Sequence
537 Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
538 Libkind, D., Hittinger, C.T., Valério, E., Gonçalves, C., Dover, J., Johnston, M., et al. (2011)
539 Microbe domestication and the identification of the wild genetic stock of lager-brewing
540 yeast. *PNAS* **108**: 14539–14544.
541 Liti, G., Ba, A.N.N., Blythe, M., ller, C.A.M., m, A.B., Cubillos, F.A., et al. (2013) High quality de
542 novo sequencing and assembly of the *Saccharomyces arboricolus* genome. *BMC*
543 *Genomics* **14**: 1–14.
544 Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., et al. (2009) Population
545 genomics of domestic and wild yeasts. *Nature* **458**: 337–341.
546 McGovern, P.E., Zhang, J., Tang, J., Zhang, Z., Hall, G.R., Moreau, R.A., et al. (2004)
547 Fermented beverages of pre-and proto-historic China. *PNAS* **101**: 17593–17598.
548 Naumov, G.I., James, S.A., Naumova, E.S., Louis, E.J., and Roberts, I.N. (2000) Three new
549 species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*,
550 *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *International Journal of*
551 *Systematic and Evolutionary Microbiology* **50**: 1931–1942.

552 Naumov, G.I., Lee, C.-F., and Naumova, E.S. (2013) Molecular genetic diversity of the
553 *Saccharomyces* yeasts in Taiwan: *Saccharomyces arboricola*, *Saccharomyces cerevisiae*
554 and *Saccharomyces kudriavzevii*. *Antonie van Leeuwenhoek* **103**: 217–228.
555 Naumov, G.I., Naumova, E.S., and Sniegowski, P.D. (1997) Differentiation of European and Far
556 East Asian populations of *Saccharomyces paradoxus* by allozyme analysis. *Int. J. Syst.*
557 *Bacteriol.* **47**: 341–344.
558 Naumov, G.I., Naumova, E.S., and Sniegowski, P.D. (1998) *Saccharomyces paradoxus* and
559 *Saccharomyces cerevisiae* are associated with exudates of North American oaks. *Can. J.*
560 *Microbiol.* **44**: 1045–1050.
561 Peris, D., Sylvester, K., Libkind, D., Gonçalves, P., Sampaio, J.P., Alexander, W.G., and
562 Hittinger, C.T. (2014) Population structure and reticulate evolution of *Saccharomyces*
563 *eubayanus* and its lager-brewing hybrids. *Molecular Ecology* **23**: 2031–2045.
564 Pfeiffer, T., Schuster, S., and Bonhoeffer, S. (2001) Cooperation and competition in the
565 evolution of ATP-producing pathways. *Science* **292**: 504–507.
566 Piskur, J., Rozpedowska, E., Polakova, S., Merico, A., and Compagno, C. (2006) How did
567 *Saccharomyces* evolve to become a good brewer? *Trends Genet.* **22**: 183–186.
568 Polz, M.F. and Cavanaugh, C.M. (1998) Bias in template-to-product ratios in multitemplate
569 PCR. *Applied and Environmental Microbiology* **64**: 3724–3730.
570 Rambaut, A. (2012) FigTree.
571 Replansky, T., Koufopanou, V., Greig, D., and Bell, G. (2008) *Saccharomyces sensu stricto* as a
572 model system for evolution and ecology. *Trends in Ecology & Evolution* **23**: 494–501.
573 Rodríguez, M.E., Pérez-Través, L., Sangorrín, M.P., Barrio, E., and Lopes, C.A. (2014)
574 *Saccharomyces eubayanus* and *Saccharomyces uvarum* associated with the fermentation
575 of *Araucaria araucana* seeds in Patagonia. *Fems Yeast Res* **14**: 948–965.
576 Rokas, A., Williams, B.L., King, N., and Carroll, S.B. (2003) Genome-scale approaches to
577 resolving incongruence in molecular phylogenies. *Nature* **425**: 798–804.

578 Rozpedowska, E., Hellborg, L., Ishchuk, O.P., Orhan, F., Galafassi, S., Merico, A., et al. (2011)
579 Parallel evolution of the make-accumulate-consume strategy in *Saccharomyces* and
580 *Dekkera* yeasts. *Nat Commun* **2**: 302.
581 Salvado, Z., Arroyo-Lopez, F.N., Guillamon, J.M., Salazar, G., Querol, A., and Barrio, E. (2011)
582 Temperature adaptation markedly determines evolution within the genus *Saccharomyces*.
583 *Applied and Environmental Microbiology* **77**: 2292–2302.
584 Scannell, D.R., Zill, O.A., Rokas, A., Payen, C., Dunham, M.J., Eisen, M.B., et al. (2011) The
585 awesome power of yeast evolutionary genetics: new genome sequences and strain
586 resources for the *Saccharomyces* sensu stricto genus. *G3: Genes, Genomes, Genetics* **1**:
587 11–25.
588 Serjeant, K., Tang, R., Anfang, N., Beggs, J.R., and Goddard, M.R. (2008) Yeasts associated
589 with the New Zealand *Nothofagus* honeydew system. *New Zeal J Ecol* **32**: 209–213.
590 Sniegowski, P.D., Dombrowski, P.G., and Fingerman, E. (2002) *Saccharomyces cerevisiae* and
591 *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display
592 different levels of reproductive isolation from European conspecifics. *Fems Yeast Res* **1**:
593 299–306.
594 Spring, S., Lins, U., Amann, R., Schleifer, K.H., Ferreira, L., Esquivel, D., and Farina, M. (1998)
595 Phylogenetic affiliation and ultrastructure of uncultured magnetic bacteria with unusually
596 large magnetosomes. *Arch. Microbiol.* **169**: 136–147.
597 Tsai, I.J., Bensasson, D., Burt, A., and Koufopanou, V. (2008) Population genomics of the wild
598 yeast *Saccharomyces paradoxus*: Quantifying the life cycle. *Proceedings of the National*
599 *Academy of Sciences* **105**: 4957–4962.
600 Wang, Q.-M., Liu, W.-Q., Liti, G., Wang, S.-A., and Bai, F.-Y. (2012) Surprisingly diverged
601 populations of *Saccharomyces cerevisiae* in natural environments remote from human
602 activity. *Molecular Ecology* **21**: 5404–5417.
603 Wang, S.A. and Bai, F.Y. (2008) *Saccharomyces arboricolus* sp. nov., a yeast species from tree

bark. *International Journal of Systematic and Evolutionary Microbiology* **58**: 510–514.

White, T.J., Bruns, T., Lee, S., Taylor, J.W., others (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications* **18**: 315–322.

Zhang, H., Richards, K.D., Wilson, S., Lee, S.A., Sheehan, H., Roncoroni, M., and Gardner, R.C. (2015) Genetic characterization of strains of *Saccharomyces uvarum* from New Zealand wineries. *Food Microbiology* **46**: 92–99.

Zhang, H., Skelton, A., Gardner, R.C., and Goddard, M.R. (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.

Figure Legends

Figure 1: 50% Majority rule consensus tree produced by maximum likelihood analysis showing the positions of the ten New Zealand derived genomes (in red) sequenced in this study within the broader *Saccharomyces* genus constructed using 106 orthologous loci from whole genome sequences. Bootstrap values for labelled clades are indicated near branches.

Figure 2: 50% Majority rule consensus tree produced by maximum likelihood analysis showing the position of the *S. eubayanus* isolate in this study amongst the global *S. eubayanus* population from Far East Asia, North America and South America constructed using six loci totalling 3,830. *S. uvarum* isolates are included for outgroup purposes. Bootstrap values are only shown for major labelled clades to avoid clutter and are show on or near branches.

Supplementary Information

Table S1: List of species and strains with whole genomes available used for phylogenomic analyses. Strain names in bold were sequenced in this study, and are placed in the species they match best using whole genome sequence data. The initial identification of these isolates by 26S D1/D2 sequence is indicated, along with their homology to the type strain of these species at this ribosomal region. The accession number of these new ribosomal sequence are also reported.

Table S2: Alignment proportions for the 10 *Saccharomyces* genomes sequenced in this study against the six *Saccharomyces* species with reference genomes available (strains in brackets).

Figure S1: Location of Forest Parks sampled for Saccharomyces; from Google images ©

Figure S2: (a) Number of samples containing each of the species groups found per Forest Park (FP) and National Park (NP) sample site from RFLP analysis of isolated colonies broken down by geographic location (refer to Figure S1). The putative identification from sanger sequencing select isolates from each group is indicated, and the numbers in parentheses indicate the number of samples each group was present in: note, due to clonal expansion during enrichment we simply report presence of each group in samples, and not the number of colonies identified. (b) The presence of each of the yeast and bacterial groups among samples broken down by niche.

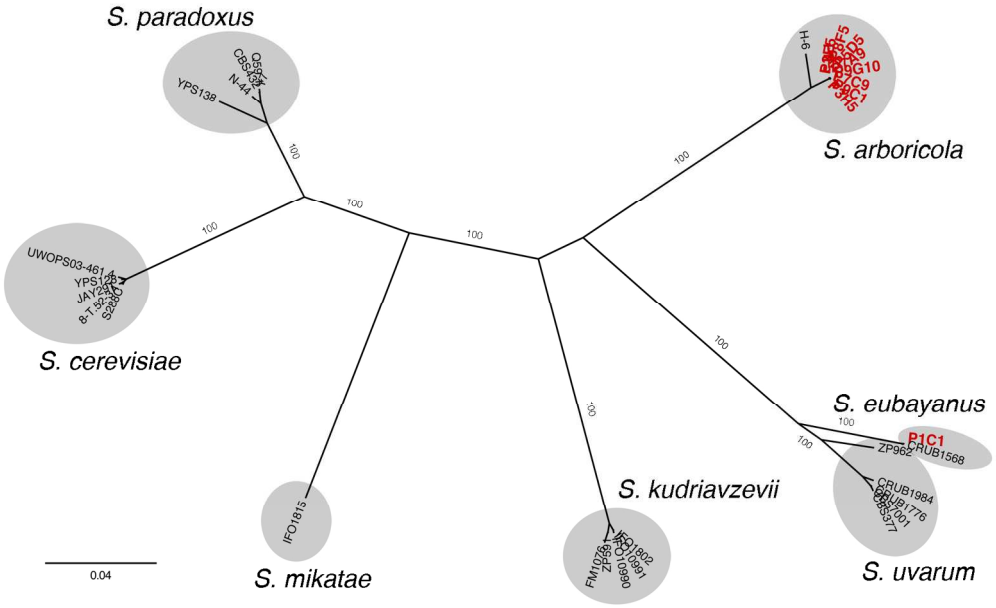

Figure 1: 50% Majority rule consensus tree produced by maximum likelihood analysis showing the positions of the ten New Zealand derived genomes (in red) sequenced in this study within the broader Saccharomyces genus constructed using 106 orthologous loci from whole genome sequences. Bootstrap values for labelled clades are indicated near branches.
169x103mm (300 x 300 DPI)

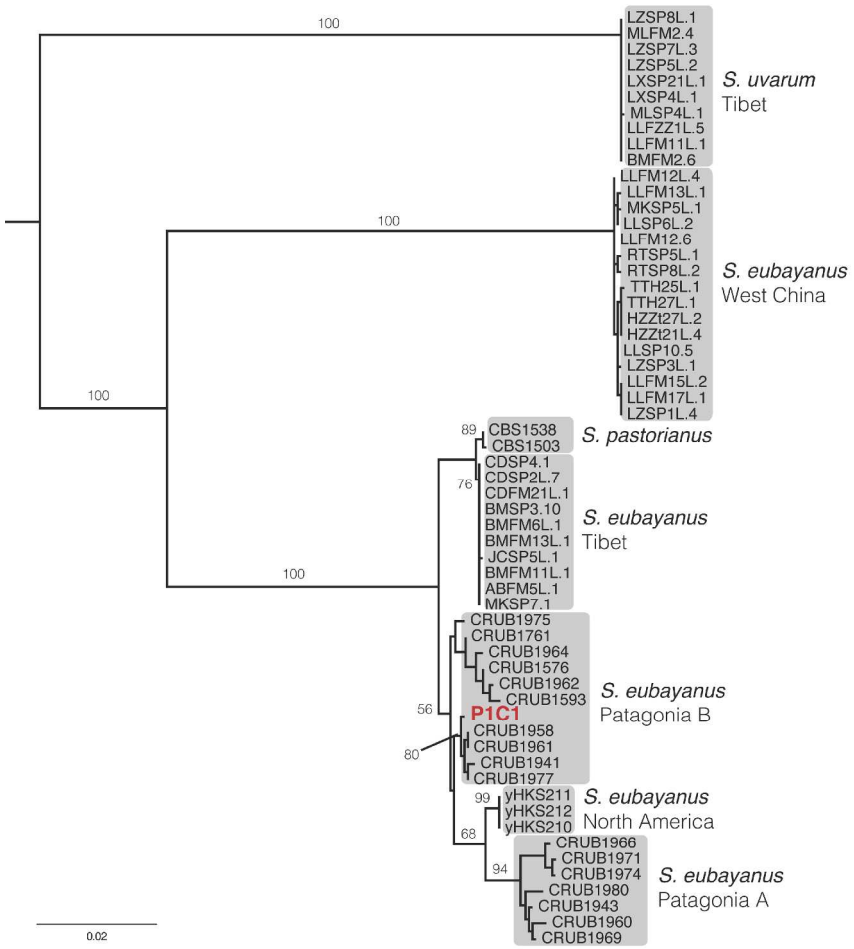

Figure 2: 50% Majority rule consensus tree produced by maximum likelihood analysis showing the position of the S. eubayanus isolate in this study amongst the global S. eubayanus population from Far East Asia, North America and South America constructed using six loci totalling 3,830. S. uvarum isolates are included for outgroup purposes. Bootstrap values are only shown for major labelled clades to avoid clutter and are show on or near branches.
209x297mm (300 x 300 DPI)