Supporting Information

Libkind et al. 10.1073/pnas.1105430108

SI Materials and Methods

Reference-Based Genome Assemblies. Our reference-based genome assembly pipeline has been described previously and shown to detect SNPs for population and phylogenomics with a conservatively estimated error rate of less than 5×10^{-4} per base pair (bp) (1). We made several improvements and modifications for this study. Briefly, the original pipeline mapped single-pass 36-bp Illumina reads with a quality filter of five and up to three mismatches using RMAPQ v. 0.45 (2), but we allowed up to five mismatches in this study because of the higher (~7%) divergence between *Saccharomyces eubayanus* and *Saccharomyces uvarum*. Phred quality scores were used, rather than the now-obsolete Solexa quality scoring system. Because there is not a widely accepted model of indel evolution, we focused entirely on SNPs, which means that all assemblies are prealigned to the *S. uvarum* CBS 7001 reference genome (3). Note that the current reference genome lacks annotations for mitochondrial and 2-μm sequences.

All genomes were sequenced on a single lane of the Illumina GAII or GAIIx platform, as described previously (1), but this study spanned dramatic increases in read yield and quality because of technological improvements. To make all assemblies directly comparable and minimize the impact of low-level contamination (~1% among some libraries prepared using the same agarose gel), we jack-knifed all datasets to exactly 4,549,184 mapped 36-bp reads to build the reference-based final assemblies (>13× coverage). To control for any artifacts because of alignment or assembly errors, we shredded the *Saccharomyces pastorianus* genome (4) into 9,098,368 (because its hybrid genome is roughly twice as large) simulated error-free 36-bp Illumina reads and mapped them to a combined reference genome that included both *S. uvarum* CBS 7001 (3) and *Saccharomyces cerevisiae* S288c (www.yeastgenome.org, accessed 12 October 2009). This process provided a reliable and prealigned assembly of the mappable *S. eubayanus* portions of the *S. pastorianus* genome that had been treated in the same manner as the strains we sequenced directly.

Because we suspected some strains might contain both *S. eubayanus* and *S. uvarum* alleles at some loci, in addition to calling unambiguous homozygous bases as described previously (1), we also called base positions with strong evidence of heterozygosity. Heterozygous base calls were recorded when the combined error probability (p_c) was estimated to be less than 10^{-5} for two possible base calls and both these p_c values were at least 10^{-5} lower than any other third possible base call at that position. This procedure proved reliable in most nonrepetitive portions of the genome and showed that only CBS 380^T had evidence for substantial heterozygosity. To further remove spurious heterozygous calls, the final heterozygous calls only included positions where the two inferred alleles matched *S. uvarum* CBS 7001 and *S. eubayanus* FM1318 (leaving 83% or 107,211 of 128,800 heterozygous base pairs for CBS 380^T). Bases lacking unambiguous homozygous or heterozygous calls were recorded as "N." Complete assembly statistics are shown in Dataset S2.

Divergence and Heterozygosity Plots. All genome-wide estimates of divergence and heterozygosity consider only base pairs that were unambiguously called in all seven genomes in the study, passed a series of quality filters, and could be placed onto chromosomes. Divergence is not corrected for multiple substitutions. When calculating divergence, heterozygous base pairs were counted as one-half of a mismatch to each reference base represented in the heterozygous pair. Because highly divergent regions are more challenging or impossible to assemble to a reference genome, all divergence estimates should be regarded as a minimum estimates.

To filter and interpret the genome assemblies, we broke all contigs into 1-kbp (window and step) sliding windows and retained those windows with at least the same 500 bps unambiguously called in all seven genomes. For Fig. 1, each datapoint is the average of itself and five filtered windows on either side, meaning that there is a step of 1 kbp, with a window of 11 kbp in most regions but sparser data in some regions. We manually ordered the largest contigs of the Cliften et al. (3) *S. uvarum* assembly (those numbered contig 429 and higher) using the gene annotations and synteny with *S. cerevisiae* and allowing for translocations where applicable (e.g., ref. 5). Contigs that could not be reliably placed or were numbered contig 428 and lower are recorded at the end of Dataset S1 but were not included in Fig. 1 or in genome-wide divergence and heterozygosity estimates. This filtering left 8,083,962 filtered base pairs covering 9,514 1-kbp windows on all 16 chromosomes (~80% of the genome), which contain 89,347 heterozygous bps (1.11% heterozygosity) for CBS 380^T and <0.04% heterozygosity for all other sequenced strains.

Quantifying the Donors to Each Strain. To estimate the relative contributions of *S. uvarum* and *S. eubayanus* to the genomes of each strain, we assigned all 9,514 filtered 1-kbp windows to the most likely donor species by determining whether the divergence of the strain in question is lower in comparison with the *S. uvarum* CBS 7001 reference or the *S. eubayanus* FM1318 reference. Windows with heterozygosity >1% were instead considered heterozygous. These data, reported in the main text and the legend of Fig. 1, show that FM1317 and CBS 395^T contain at least 99.9% *S. uvarum* alleles, the non-*S. cerevisiae* portions of the *S. pastorianus* genome contain at least 99.9% *S. eubayanus* alleles, and NBRC 1948 and CBS 380^T contain variable contributions from both *S. uvarum* and *S. eubayanus* (and *S. cerevisiae*) (Dataset S3).

Copy-Number Variant Analysis. Because of our inability to recover viable spores from CBS 380^T and CBS 395^T, we suspected that these strains might be aneuploid. We considered the same filtered 1-kbp windows shown in the divergence and heterozygosity plots, calculated the sum coverage for each window (sum of the number of times a mapped read included each position in the window), divided each by the average coverage/base pair for all filtered windows, and divided each by the length of the window (1,000 bps, except for the few windows at the end of contigs that still had 500 or more base pairs called in all strains). Windows with a value of approximately one have a single copy per haploid genome or two copies per diploid genome. All filtered data are presented in Dataset S1, but we wish to explicitly note the whole chromosomal aneuploidies present in CBS 380^T (chromosome I) and CBS 395^T [chromosomes VIII and XV, which interestingly have a documented reciprocal translocation relative to *S. cerevisiae* (5)]. All three cases involved values of ~1.5 across the length of chromosomes, indicating triploidy. This analysis also suggested FM1309 had a possible endoreduplication defect, as noted for one previously sequenced *Saccharomyces kudriavzevii* strain (1); the cause remains unknown, but it did not affect any other strain in this study.

Screening for Introgression, Hybridization, or Horizontal Gene Transfer. We deployed a previously developed pipeline (1) to screen our 36-bp sequencing reads for evidence of any foreign

genes from other *Saccharomyces* sensu stricto species by mapping reads with up to five mismatches to the best placement among the annotated coding sequences of all five reference genomes (*S. cerevisiae*, *S. paradoxus*, *Saccharomyces mikatae*, *S. kudriavzevii*, and *S. uvarum*) for which the original authors had identified *S. cerevisiae* orthologs [Cliften et al. (6) for *S. kudriavzevii*; Kellis et al. (7) for the other species]. We then considered all 2,805 genes that had orthologs annotated in all five reference genomes and manually examined the evidence for all genes where the majority of reads did not map to *S. uvarum* (closely related enough to *S. eubayanus* for this purpose). As in the previous study, most candidates were clearly artifacts (usually because of annotation errors), but we did uncover a number of bona fide transfers from *S. cerevisiae* into CBS 380^T and NBRC 1948 (Dataset S3). Because *S. pastorianus* is already well-established as being a hybrid with *S. cerevisiae*, we only searched for additional contributors and found none. We found no evidence of any non-*S. uvarum*/*eubayanus*-like contributors to either Patagonian strain or to CBS 395^T.

We suspected additional contributions from *S. cerevisiae* in some strains, so we also considered all *S. cerevisiae* genes for which there were not orthologs annotated in all five species. For each strain except *S. pastorianus*, we normalized the number of reads mapping to each nonconserved *S. cerevisiae* gene by multiplying by the total number of reads mapped for the strain with the fewest mapped reads in the 2,805-gene analysis above (1,325,804 reads from FM1318) and dividing by the total number of reads mapped in the analysis above for the strain being tested. We then manually examined the evidence for all genes that had at least 100 normalized reads (effective coverage of 3.6× for a 1-kbp coding region; the expected effective coverage was 14.7×) mapping to nonconserved *S. cerevisiae* genes. Again, other than experimental or annotation artifacts, only CBS 380^T and NBRC 1948 displayed additional evidence of foreign genes being contributed by *S. cerevisiae*. Although genomic position cannot be conclusively determined solely from single-read 36-mers, all transferred genes are subtelomeric in the *S. cerevisiae* and/or *S. uvarum* reference genomes (Dataset S3).

Sequence Analyses of *IMA1* and *SUL1*. We analyzed modified Nei-Gojobori (8) divergence at synonymous sites (*dS*) near the *IMA1* breakpoint fusing an *S. eubayanus* to an *S. cerevisiae* chromosome (Fig. 2*A*) with a window of 100 sites, a step of 1, and an arbitrary intergenic spacer of 200 sites using DNASP v. 4.90.1 (9). Because our reference-based genome assemblies do not consider indels, we examined all *SUL1* coding sequences (active and inactivated) using the reference genomes for *S. uvarum* and *S. pastorianus* and de novo assemblies for all of the strains sequenced in this study. These assemblies were built using the default settings of VELVET 0.7.55 (10) to assemble all reads with a *k*-mer of 21. A local reciprocal-best-BLAST (11) criterion allowing for identical partially overlapping *k*-mers (12) was then applied against the coding sequence of *SUL1* to retrieve, order, merge, and orient contigs relative to *S. uvarum* reference genome. In one case, a single base-pair disagreement between the (less reliable) edges of two contigs was resolved using our reference-based genome assembly. The reference-based assemblies were also used to fill in small regions that VELVET failed to recover for some strains.

1. Hittinger CT, et al. (2010) Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* 464:54–58.
2. Smith AD, Xuan Z, Zhang MQ (2008) Using quality scores and longer reads improves accuracy of Solexa read mapping. *BMC Bioinformatics* 9:128.
3. Cliften PF, Fulton RS, Wilson RK, Johnston M (2006) After the duplication: Gene loss and adaptation in *Saccharomyces* genomes. *Genetics* 172:863–872.
4. Nakao Y, et al. (2009) Genome sequence of the lager brewing yeast, an interspecies hybrid. *DNA Res* 16:115–129.
5. Fischer G, James SA, Roberts IN, Oliver SG, Louis EJ (2000) Chromosomal evolution in *Saccharomyces*. *Nature* 405:451–454.
6. Cliften P, et al. (2003) Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* 301:71–76.
7. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423:241–254.
8. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426.
9. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
10. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829.
11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
12. Hittinger CT, Johnston M, Tossberg JT, Rokas A (2010) Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *Proc Natl Acad Sci USA* 107:1476–1481.

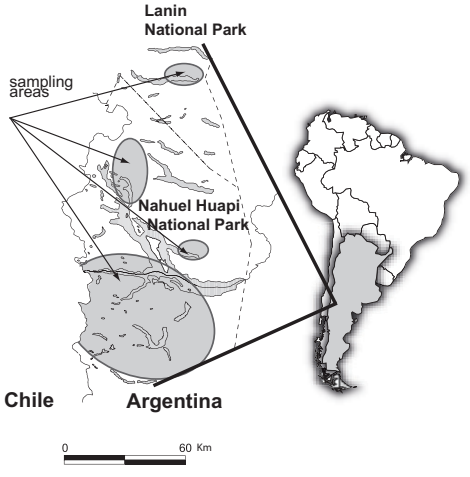


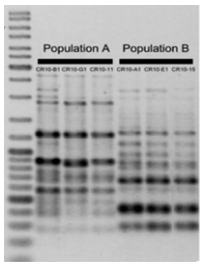
Fig. S1. *Saccharomyces* spp. were isolated from samples collected in different areas within and near Lanin and Nahuel Huapi National Parks in Northwestern Patagonia, Argentina.

A - sequence analysis

Nomenclature	Strain	Strain	Strain	Strain	Strain information	FSY	ITS	FUN14	RIP1	MET2	GDH1	HIS3
Patagonia - group A (*S. eubayanus*)	CRUB 1568	CBS 12357	2a	2-2R10	Fruiting body of Cyttaria harioti, Patagonia, Argentina, isolated in 2006	JF786690	JF786673	JF786614	JF786665	JF786640		JF786623
Saccharomyces bayanus	CBS 380^{T}	PYCC 4456	NRRL Y-12624	CLIB 181	Turbid beer; Type of S. bayanus	JF786654	AJ229058	JF786615	JF786661	AJ627635	AJ627639	JF786624
Saccharomyces bayanus	NBRC 1948	NCYC 114			Brewing contaminant (Rainieri et al. 2006) A.C. Chapman, isolated in 1925	JF786691	JF786674	JF786616	JF786666	JF786641		
Saccharomyces bayanus	CBS 424	PYCC 4568			Pear juice, Switzerland; isolated in 1899; Type of S. globosus	JF786692	Z95947	JF786617	JF786667		*(2)	JF786625
Saccharomyces heterogenicus	CBS 425	PYCC 4569	CLIB 255		Fermented apple juice, isolated in 1924; Type of S. heterogenicus	JF786693	EU145762	JF786618	JF786668	AJ251011	*(2)	JF786626
Saccharomyces bayanus	CBS 1505	CLIB 254			Juice of Fendant grapes, isolated in 1904; Type of S. intermedius var. valdensis	JF786694	JF786675	JF786656	JF786662	AJ251010	*(2)	JF786627
Saccharomyces inusitatus	CBS 1546	CLIB 252			Beer, Netherlands, isolated in 1965; Type of S. inusitatus	JF786655	EU145763	JF786657	JF786663	JF786642		
Saccharomyces bayanus	CBS 378	PYCC 4565			Unknown, A. Klöcker	JF786695	JF786676	JF786658	JF786664	JF786659		
Saccharomyces bayanus	DBVPG 6347	PYCC 4652	NCYC 374-1		Hazy ale, Scottish Brewery, isolated in 1953	JF786696	JF786677	JF786619	JF786669	JF786660		
Saccharomyces bayanus	CBS 3008				Must of soft fruit, A. von Szilvinyi	JF786697	JF786678	JF786620	JF786670	JF786643		JF786628
Saccharomyces bayanus	NBRC 0539				Unknown,Y. Kobayashi	JF786698	JF786679	JF786621	JF786671	JF786644		JF786629
Patagonia - group B (*S. uvarum*)	CRUB 1595	4-1	4-4R10		Fruiting body of Cyttaria harioti, Patagonia, Argentina, isolated in 2006	JF786699	JF786680	JF786622	JF786672	JF786645		JF786630
Saccharomyces uvarum	CBS 395^{T}	CLIB 251			Juice of Ribes nigrum (black currant), Netherlands; Type of S. uvarum	JF786700	Z95946			AJ627638	AJ627640	JF786631
Saccharomyces uvarum	CBS 7001				Mesophylax adopersus, Avila, Spain, J. Santa Maria	PORF3043		PORF325	PORF6152	PORF19897	PORF24096	PORF23664
Saccharomyces uvarum	CBS 426	NCYC 2589			Honey, O. Verona	JF786701				JF786646	*(2)	JF786632
Saccharomyces uvarum	UWO 99-807-1-1				Nothofagus sp. Patagonia, Argentina, isolated by M.-A. Lachance	JF786702	JF786681			JF786647		JF786633
Saccharomyces uvarum	UWO 99-808-3				Nothofagus sp. Patagonia, Argentina, isolated by M.-A. Lachance	JF786703	JF786682					
Saccharomyces uvarum	NCAIM 00677				Fermented drink, Hungary (originally identified as S. bayanus)	JF786704	JF786683					
Saccharomyces uvarum	NCAIM 00789				Exudate of Carpinus betulus (hornbeam) Hungary (originally identified as S bayanus (3))	JF786705	JF786684			JF786648		JF786634
Saccharomyces uvarum	NCAIM 00868				Slimy material on a stump, Hungary (originally identified as S. bayanus)	JF786706	JF786685			JF786649		JF786635
Saccharomyces uvarum	NCAIM 01107				Canned sour cherry, Hungary (originally identified as S. bayanus)	JF786707	JF786686			JF786650		JF786636
Saccharomyces uvarum	NCAIM 01116				Grape berries, Russia (originally identified as S. bayanus (4))	JF786708	JF786687			JF786651		JF786637
Saccharomyces uvarum	UCD 61-137	CBS 8696			Drosophila pseudoobscura, Berryessa Hills, California, USA	JF786709	JF786688			JF786652		JF786638
Saccharomyces uvarum	UCD 51-206	CBS 8697			Drosophila persimiles, Yosemite Natl. Park, California, USA	JF786710	JF786689			JF786653		JF786639

* GenBank accession numbers not available

B - PCR fingerprinting



C - PCR-RFLP analysis

Population	Strain	Gene *FUN14* *Ban*II	Gene *RIP1* *Pst* I	Gene *HIS3* *Rsa*I
A	CRUB 1569			
A	CRUB 1575			
A	CRUB 1576			
A	CRUB 1581			
A	CRUB 1582			
B	CRUB 1566			
B	CRUB 1572			
B	CRUB 1586			
B	CRUB 1590			
B	CRUB 1595			

References

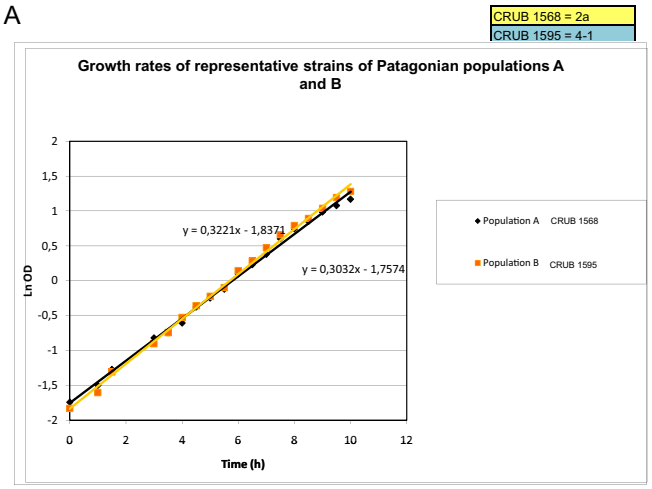
1. Rainieri S et al. (2006) Pure and mixed genetic lines of Saccharomyces bayanus and Saccharomyces pastorianus and their contribution to the lager brewing strain genome. Appl Environm Microbiol 72:3968-3974.
2. Nguyen H-V, Gaillardin C (2005) Evolutionary relationships between the former species Saccharomyces uvarum and the hybrids Sacharomyces bayanus and Saccharomyces pastorianus; reinstatement of Saccharomyces uvarum (Beijerinck) as a distinct species. FEMS Yeast Res 5:471-483
3. Naumov G, Gazdiev DO, Naumova ES (2003) The finding of the yeast species Saccharomyces bayanus in Far East Asia. Microbiology 72: 738-743 (translated from Mikrobiologiya 72: 834-839).
4. Naumov GI et al. (1995) A new genetically isolated population of the Saccharomyces sensu stricto complex from Brazil. Antonie van Leeuwenhoek 67: 351-355.

Accronyms of culture collections

CBS: Centraalbureau voor Schimmelculture (Netherlands)
CLIB: Collection de Levures de Intérêt Biotechnologique (France)
CRUB: Centro Regional Universitario Bariloche Culture Collection (Argentina)
DBVPG: Industrial Yeasts Collection (Italy)
NBRC: NITE Biological Resource Center (Japan)
NCAIM: National Collection of Agricultural and Industrial Microorganisms (Hungary)
NCYC: National Collection of Yeast Cultures (UK)
PYCC: Portuguese Yeast Culture Collection (Portugal)
UCD: University of California, Davis - Phaff Yeast Culture Collection (USA)
UWO: University of Western Ontario, Dr M.-A. Lachance (Canada)

Fig. S2. Preliminary characterization of Patagonian isolates of groups A and B and generalized heterogeneity of European isolates of *S. bayanus*. (*A*) sequence analyses using multiple gene sequences and comparison with reference strains of *S. bayanus* and *S. uvarum*. Yellow boxes denote *S. eubayanus* alleles, blue boxes represent S. uvarum alleles and orange boxes represent heterozygous (*S. eubayanus* x *S.uvarum*) sequences. The results for the Patagonian isolates are shown for two exemplary strains (highlighted in light and dark gray shades), although a larger set of isolates was investigated with identical results. Note that reference strains previously identified as *S. bayanus* are either not pure lineages (*Upper*) or belong to *S. uvarum* (*Lower*); for the list of primers used see Dataset S6. (*B*) PCR-fingerprints of representative strains of the two populations obtained with the microsatellite primer (GTG)5. (*C*) Results of PCR-RFLP for representative strains of the two populations: FUN14 was digested with BanII, RIP1 with Pst I and HIS3 with RsaI (1); color codes are the same as in *A*.

A



B

Population	Strain	Growth medium	10ºC		30ºC	
			Respiration (mmol $g^{-1} h^{-1}$)	Total flux (mmol $g^{-1} h^{-1}$)	Respiration (mmol $g^{-1} h^{-1}$)	Total flux (mmol $g^{-1} h^{-1}$)
Population B *(S. uvarum)*	4-1	YNB + 2% Glucose	0.26	1.32	0.56	7.96
			nd	nd	0.51	7.51
	CR30-23a		nd	nd	0.54	7.29
Population A *(S. eubayanus)*	2a	YNB + 2% Glucose	0.24	1.23	0.72	6.75
			nd	nd	0.51	6.08
	CR10-4		nd	nd	0.08	7.8

nd, not determined

1. Umbreit W, Burris R, Stauffer J (1964) Manometric Techniques. Minneapolis: Burgess Publishing Co

C

	CRUB 1568	3-5S30	CR10-4	CR10-11	CR30-37	CRUB 1595	CR30-23a	CR10-13
Glucose	+	+	+	+	+	+	+	+
Galactose	+	+	+	+	+	+	+	+
Maltose	+	+	+	+	+	+	+	+
Melibiose	+	+	+	+	+	+	+	+
Trehalose	d	d	-	-	-	d	d	d
D-Mannitol	d	d	d	d	d	d	d	-
Methyl-α-D-glucoside	+	+	+	+	+	d	+	+
Palatinose	+	+	+	+	+	d	+	+
Melezitose	+	+	+	+	+	d	d	d
Growth at 34ºC	+	+	+	+	+	+	+	+
Growth at 35ºC	+	+	+	+	+	+	+	+
Growth at 36ºC	-	-	-	-	-	-	-	-
sensitivity to possible killer activity of CRUB 1568		-	-	nd	nd	-	-	nd
sensitivity to possible killer activity of CRUB 1595	-	-	-	nd	nd		-	nd

+, positive; -, negative;d, delayed; nd, not determined.

Fig. S3. Phenotypic similarity between the two Patagonian populations. (*A*) Growth rates of two representatives of populations A and B in YNB medium with 2% glucose, at 25 °C. (*B*) Glycolytic and respiratory fluxes determined for representative strains of Patagonian populations A and B. Glucose consumption rates were measured in resting cells using a Warburg volumetric apparatus (B. Braun Melsungen), which allows the quantification of CO_2 produced and O_2 consumed (1). (*C*) Assimilation of carbon sources, growth temperatures and killer activity/sensitivity of representatives of populations A and B.

Table S1. *Saccharomyces* isolations in Patagonia

	n	30 °C Seub	30 °C Suva	30 °C Scer	10 °C Seub	10 °C Suva	10 °C Scer
Bark							
Nothofagus antarctica	23	10	—	—	12	2	—
Nothofagus dombeyi	27	—	2	—	2	17	—
Nothofagus pumilio	15	8	—	—	9	3	—
Soil							
N. antarctica	15	2	—	—	3	—	—
N. dombeyi	15	—	—	—	1	9	—
N. pumilio	15	7	1	—	12	—	—
Cyttaria hariotii							
N. antarctica	9	—	—	—	6	2	—
N. dombeyi	12	—	—	2	5	8	—
N. pumilio	2	—	—	—	1	1	—
Total	133	27	3	2	51	42	0

Each sample was separated in two portions and incubated at 30 °C and 10 °C (when the same sample yielded the same *Saccharomyces* species at the two temperatures only one isolate was retained; Seub, *S. eubayanus*; Scer, *S. cerevisiae*; Suva, *S. uvarum*).

Table S2. Genetic distance and hybrid spore viability among *Saccharomyces* species

Comparison	Genetic distance	Spore viability	Reference
S. cerevisiae – *S. paradoxus*	13%	0.6%	1
S. eubayanus – *S. uvarum*	6.9%	7.3%	Present study
S. paradoxus FE – *S. paradoxus* NA	4.7%	32%	1
S. paradoxus EUR – *S. paradoxus* NA	4.7%	40%	1
S. paradoxus FE – *S. paradoxus* EUR	1.2%	75%	1
S. kudriavzevii FE – *S. kudriavzevii* EUR	1.1%	82%	2

EUR, Europe; FE, Far East; NA, North America.

1. Liti G, Barton DBH, Louis EJ (2006) Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* 174:839–850.
2. Hittinger CT, et al. (2010) Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* 464:54–58.

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLS)
Dataset S3 (XLS)
Dataset S4 (XLS)
Dataset S5 (XLS)
Dataset S6 (XLSX)