Supplementary Notes

Note 1. Newly generated datasets, main reference genomes, and other previously published datasets used in this study

Note 1.1 Datasets generated in this study

The PacBio sequencing data for the 12 strains
Deposited location: EBI ENA database
Accession number: PRJEB7245
URL: http://www.ebi.ac.uk/ena/data/view/PRJEB7245

The Illumina sequencing data for the 12 strains
Deposited location: NCBI SRA database
Accession number: PRJNA340312
URL: http://www.ncbi.nlm.nih.gov/bioproject/340312

The final assemblies, annotations, extracted CDSs and protein sequences for the 12 strains
Deposited location: Our dedicated project website
URL: https://yjx1217.github.io/Yeast_PacBio_2016/data/
The assemblies were also deposited in Genbank under project PRJEB7245.
URL: https://www.ncbi.nlm.nih.gov/bioproject/PRJEB7245/

The re-annotation for *Saccharomyces arboricolus* H6
Deposited location: Our dedicated project website
URL: https://yjx1217.github.io/Yeast_PacBio_2016/data/

Note 1.2 Main reference genomes used in this study

***Saccharomyces cerevisiae* reference nuclear genome**
Strain: S288C
Version: R64-1-1
URL: http://downloads.yeastgenome.org/sequence/S288C_reference/genome_releases/

***Saccharomyces cerevisiae* reference mitochondrial genome**
Strain: S288C
Version: R64-1-1
URL: http://downloads.yeastgenome.org/sequence/S288C_reference/genome_releases/

***Saccharomyces paradoxus* reference nuclear genome**
Strain: CBS432
Version: para2 (inside the misc.tgz file)
URL: ftp://ftp.sanger.ac.uk/pub/users/dmc/yeast/latest/

***Saccharomyces paradoxus* reference mitochondrial genome**

Strain: CBS432
Accession: NCBI Genbank JQ862335
URL: https://www.ncbi.nlm.nih.gov/nuccore/JQ862335

Note 1.3 Other previously published datasets used in this study

Study: Population genomics of domestic and wild yeasts (2009)
Authors: Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O'Kelly MJT, van Oudenaarden A, Barton DBH, Bailes E, Nguyen AN, Jones M, Quail MA, Goodhead I, Sims S, Smith F, Blomberg A, Durbin R and Louis EJ
Reference: NCBI PubMed PMID: 19212322
URL: ftp://ftp.sanger.ac.uk/pub/users/dmc/yeast/latest/

Study: The awesome power of yeast evolutionary genetics: New genome sequences and strain resources for the *Saccharomyces sensu stricto* genus (2011)
Authors: Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, Rine J, Johnston M and Hittinger CT
Reference: NCBI PubMed PMID: 22384314
URL: http://www.saccharomycessensustricto.org/cgi-bin/s3.cgi

Study: High quality *de novo* sequencing and assembly of the *Saccharomyces arboricolus* genome (2013)
Authors: Liti G, Nguyen Ba AN, Blythe M, Müller CA, Bergström A, Cubillos FA, Dafhnis-Calas F, Khoshraftar S, Malla S, Mehta N, Siow CC, Warringer J, Moses AM, Louis EJ and Nieduszynski CA
Reference: NCBI PubMed PMID: 23368932
URL: http://www.moseslab.csb.utoronto.ca/sarb/

Study: A high-definition view of functional genetic variation from natural yeast genomes (2014)
Authors: Bergström A, Simpson JT, Salinas F, Barré B, Parts L, Zia A, Nguyen Ba AN, Moses AM, Louis EJ, Mustonen V, Warringer J, Durbin R, and Gianni Liti G
Reference: NCBI PubMed PMID: 24425782
URL: ftp://ftp.sanger.ac.uk/pub/users/dmc/yeast/SGRP2/input/strains/

Study: Long-read, whole-genome shotgun sequence data for five model organisms (2014)
Authors: Kim KE, Peluso P, Babayan P, Yeadon PJ, Yu C, Fisher WW, Chin CS, Rapicavoli NA, Rank DR, Li J, Catcheside DE, Celniker SE, Phillippy AM, Bergman CM, and Landolin JM
Reference: NCBI PubMed PMID: 25977796
URL: http://www.ncbi.nlm.nih.gov/assembly/GCA_000773925.1

Study: Oxford Nanopore sequencing, hybrid error correction, and *de novo* assembly of a eukaryotic genome (2015)
Authors: Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC and McCombie WR
Reference: NCBI PubMed PMID: 26447147
URL: http://schatzlab.cshl.edu/data/nanocorr/

Study: The genome sequence of *Saccharomyces eubayanus* and the domestication of Lager-brewing yeasts (2015)
Authors: Baker E, Wang B, Bellora N, Peris D, Hulfachor AB, Koshalek JA, Adams M, Libkind D,and Hittinger CT
Reference: NCBI PubMed PMID: 26269586
URL: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA243390/

Note 2. Detailed configuration for the Maker pipeline

For repeatmasking, RepeatMasker (v4.0.5) (http://www.repeatmasker.org/) was used with the configuration options: "model_org=fungi" and "softmask=1". *Ab initio* gene prediction was performed by SNAP (release 2013-11-29)[1] and AUGUSTUS (v3.1.0)[2] with pre-trained gene prediction parameters. For SNAP, the pre-trained HMM parameter file was downloaded from GitHub (https://github.com/hyphaltip/fungi-gene-prediction-params/blob/master/params/SNAP/saccharomyces_cerevisiae_S288C.hmm). For AUGUSTUS, we used the parameter file for *Saccharomyces* that shipped with AUGUSTUS. In complementary to *ab initio* gene prediction, the Maker pipeline also performs EST/protein alignment to further assess the automatically predicted gene models. For the EST data, we retrieved it from FungiDB (http://fungidb.org/common/downloads/release-3.2/Scerevisiae_/fasta/). For protein data, we combined the proteomes from several sources: the *S. cerevisiae* proteome from the *Saccharomyces* Genome Database (SGD) (http://downloads.yeastgenome.org/sequence/S288C_reference/orf_protein/), the protein sequences of characterized non-reference *S. cerevisiae* open reading frames (ORFs) documented in SGD, the protein sequences of non-reference *S. cerevisiae* ORFs identified in previous studies[3–5], the proteomes of *S. paradoxus* (strain CBS432), *S. mikatae* (strain IFO1815), *S. kudriavzevii* (strain IFO1802), *S. kudriavzevii* (strain ZP591), and *S. bayanus var. uvarum* (strain CBS7001) based on Scannell et al.[6], the proteome of *S. arboricolus* (strain H6) based on Liti et al.[7], and the proteome of *S. eubayanus* (strain FM1318) based on Baker et al.[8]. These EST and protein sequences were aligned with the PacBio genome assemblies using blastn and blastx respectively (both from the NCBI-BLAST+ package (v2.2.30+)[9]) and further polished by exonerate (v2.2.0)[10]. Other custom settings that we used for the Maker pipeline include:

"min_contig=10000, min_protein=30, split_hit=1500, single_exon=1, single_length=250 and correct_est_fusion=1".

Note 3. Centromeres annotation

Saccharomyces yeasts have point centromeres that encapsulate three Centromere DNA Elements (CDEI, CDEII and CDEIII). CDEI and CDEIII are evolutionary highly conserved whereas CDEII is much less conserved[11]. For *S. cerevisiae*, centromere annotation of the reference genome is available from SGD and the corresponding sequences were retrieved as the queries. For *S. paradoxus*, the query sequences were collected from three different studies[11–13], which covered 15 centromeres. All these *S. cerevisiae and S. paradoxus* centromere queries were searched against our PacBio assemblies by Exonerate (v2.2.0)[10] for centromere annotation. The *S. paradoxus* chrVIII centromere has not been previously described, for which we ran the *de novo* annotation using a script ("CENannotate.pl") provided by Dr. Bensasson[11] (can be retrieved from https://github.com/bensassonlab/scripts). All annotated centromeres were further verified based on their flanking genes, which are conserved across all the 12 strains. For each chromosome, the current chromosome identity was determined by the centromere that this chromosome encompasses.

Note 4. Annotation of the Ty retrotransposable elements

Transposable elements (TEs) are prevalent in eukaryotic genomes and their roles in promoting structural genome evolution has been long appreciated. The *S. cerevisiae* genome harbors five classes of LTR retrotransposable elements: Ty1 to Ty5. Here, we employed RepeatMasker (v4.0.5) with a previously described custom library (containing *S. cerevisiae* Ty1-5 and *S. paradoxus* Ty3)[14] to systematically annotate Ty retrotransposable elements for all the 12 strains. The resulting output was further processed by REannotate[15] (v17.03.2015) (options: -g -k <clustalw> -f <fuzzy_file> -d 10000 -t) for Ty defragmentation. Here, we treated the Ty1- and Ty2-LTR equivalently as suggested by previous study[14]. The identified full-length Ty1s and Ty2s were manually curated based on their internal sequences. To avoid false negative due to sequence divergence between the two species, we performed another round of search, in which the internal sequences and LTRs of the representative full-length *S. paradoxus* Tys annotated in the first round were further added into our custom Ty library for this round. All the truncated Tys and solo-LTRs were further curated based on the blastn search result against our Ty library (cutoffs:

identity >= 70%, aln_length >= 100 bp).

Note 5. Annotation of the telomere associated core X-elements

We retrieved core X-element sequences from the *S. cerevisiae* reference genome according to the SGD annotation and aligned them using MUSCLE (v3.8.31)[16]. Based on the alignment, we built an HMM profile (available at our dedicated project website: https://yjx1217.github.io/Yeast_PacBio_2016/welcome/) for the core X-element using the hmmbuild program (option: --dna) from the HMMER package (v3.1b2)[17]. This HMM profile was searched against our PacBio assemblies by nhmer (also from HMMER) (options: -E 1e-3 --tblout) to identify the core X-element.

Note 6. Annotation of the telomere associated Y'-elements

We retrieved the Y'-element sequences of the *S. cerevisiae* reference genome based on the SGD annotation. There are two major classes of Y'-element for *S. cerevisiae*, the short version and the long version, differed by several large indels[18]. We selected a long Y'-element (at chrIX-L) as the representative query and performed the homology search using BLAT[19] (option: -maxIntron=1000) with subsequent filtering by pslCDnaFilter (options: -minId=0.9 -minAlnSize=1000 -bestOverlap -filterWeirdOverlapped).

Note 7. The *CUP1* locus and copper tolerance

The *CUP1* gene encodes a copper scavenging short metallothionein that keeps the intracellular level of free copper extremely low and thus mediates copper tolerance. Across our 12 strains, this gene was tandemly amplified into four, seven, and 11 copies in three *S. cerevisiae* strains (DBVPG6765, Y12 and S288C respectively) while maintaining the ancestral single copy configuration in all the other strains. The duplication segment in Y12 is different from that of DBVPG6765 and S288C, which likely reflects a scenario of convergent evolution driven by selection for copper tolerance in independently domesticated beverage producing lineages, as previously suggested[20]. Consistent with previous observations[20], we found higher copy number of *CUP1* translates into faster growth (i.e. shorter generation time) in copper ($CuCl_2$: 0.38 mM), although Y12 with seven copies appears to grow slightly faster than S288C with 11 copies (including one pseudogene copy).

Note 8. The *ARR* cluster and arsenic tolerance

The *ARR* cluster contains three consecutive subtelomeric genes (*ARR1*, *ARR2* and *ARR3*) that function collectively to provide arsenic resistance. Despite their tricky genomic locations (only a few kb from the core-X element), we successfully characterized the exact genomic arrangement of the *ARR* cluster in all the 12 strains. Consistent with our previous estimates based on read mapping coverage[4], the *ARR* cluster was duplicated in the European *S. paradoxus* CBS432 while completely lost in two *S. cerevisiae* (SK1 and UWOPS03-461.4) and two *S. paradoxus* (N44 and UWOPS91-917.1) strains. Our growth rate assay confirmed the link between *ARR* cluster loss and extreme susceptibility to arsenic (3 mM arsenite, As[III]). Despite having two copies of the *ARR* cluster with no strongly deleterious mutation in either gene/copy, CBS432 grew poorly in arsenic. The As[III] sensitivity of the South American *S. paradoxus* UFRJ50816 could potentially be explained by the pseudogenization of its *ARR2*, although it is only known to protect against pentavalent arsenic, As[V][21]. The genes immediately proximal to the *ARR* cluster are located at the chr16-R subtelomere in all the 12 strains, implying that this subtelomere should be the ancestral location for the *ARR* cluster.

References for Supplementary Notes

1. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5,** 59 (2004).
2. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32,** W309–W312 (2004).
3. Novo, M. *et al.* Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. *Proc. Natl. Acad. Sci.* **106,** 16333–16338 (2009).
4. Bergström, A. *et al.* A high-definition view of functional genetic variation from natural yeast genomes. *Mol. Biol. Evol.* **31,** 872–888 (2014).
5. Song, G. *et al.* AGAPE (Automated Genome Analysis PipelinE) for pan-genome analysis of Saccharomyces cerevisiae. *PLoS One* **10,** (2015).
6. Scannell, D. R. *et al.* The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. *G3 (Bethesda).* **1,** 11–25 (2011).
7. Liti, G. *et al.* High quality de novo sequencing and assembly of the Saccharomyces arboricolus genome. *BMC Genomics* **14,** 69 (2013).
8. Baker, E. *et al.* The genome sequence of saccharomyces eubayanus and the

domestication of lager-brewing yeasts. *Mol. Biol. Evol.* **32,** 2818–2831 (2015).
9. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).
10. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6,** 31 (2005).
11. Bensasson, D. Evidence for a high mutation rate at rapidly evolving yeast centromeres. *BMC Evol. Biol.* **11,** 211 (2011).
12. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423,** 241–54 (2003).
13. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458,** 337–41 (2009).
14. Carr, M., Bensasson, D. & Bergman, C. M. Evolutionary genomics of transposable elements in Saccharomyces cerevisiae. *PLoS One* **7,** e50978 (2012).
15. Pereira, V. Automated paleontology of repetitive DNA with REANNOTATE. *BMC Genomics* **9,** 614 (2008).
16. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
17. Eddy, S. Profile hidden Markov models. *Bioinformatics* **14,** 755–763 (1998).
18. Louis, E. J. & Haber, J. E. The structure and evolution of subtelomeric Y' repeats in Saccharomyces cerevisiae. *Genetics* **131,** 559–574 (1992).
19. Kent, W. J. BLAT - The BLAST-like alignment tool. *Genome Res.* **12,** 656–664 (2002).
20. Warringer, J. *et al.* Trait variation in yeast is defined by population history. *PLoS Genet.* **7,** e1002111 (2011).
21. Mukhopadhyay, R., Shi, J. & Rosen, B. P. Purification and characterization of Acr2p, the Saccharomyces cerevisiae arsenate reductase. *J. Biol. Chem.* **275,** 21149–21157 (2000).

Supplementary Tables

Supplementary Table 1. Strain sampling for this study.

Species	Strain	Geographic origin	Source	Genetic modification	Liti lab internal collection ID	Note
S. cerevisiae	S288C	USA	Rotting fig	*Wild type*	OS 96	Lab strain
S. cerevisiae	DBVPG6044	West Africa	Bili wine	*Mat α, ura3::KanMX, lys2::URA3, ho::HygMX*	CC406	
S. cerevisiae	DBVPG6765	Europe	Unknown	*Mat α, ura3::KanMX, lys2::URA3, ho::HygMX*	CC405	
S. cerevisiae	SK1	USA	Soil	*Wild type*	OS981	Lab strain
S. cerevisiae	Y12	Japan	Sake	*Mat α, ura3::KanMX, lys2::URA3, ho::HygMX*	CC408	
S. cerevisiae	YPS128	USA	*Quercus alba*	*Mat a, ura3::KanMX, ho::HygMX*	CC403	
S. cerevisiae	UWOPS03-461.4	Malaysia	Betram palm	*Wild type*	OS1005	
S.paradoxus	CBS432	Russia	*Quercus spp.*	*Wild type*	OS1045	Neotype strain of *S. paradoxus*
S.paradoxus	N44	Russia	*Quercus mongolica*	*Wild type*	OS1018	
S.paradoxus	YPS138	USA	*Quercus velutina*	*Wild type*	OS1022	
S.paradoxus	UFRJ50816	Brazil	*Drosophila spp.*	*Wild type*	OS1017	Previously regarded as *S. cariocanus*
S.paradoxus	UWOPS91-917.1	Hawaii	*Myoporum sandwicense*	*Mat α, ho::KanMX*	CC293	

Supplementary Table 2. PacBio and Illumina sequencing depth.

Strain	PacBio sequencing depth	Illumina sequencing depth
S288C	96.88	195.64
DBVPG6044	193.65	208.09
DBVPG6765	109.63	212.99
SK1	197.21	193.71
Y12	217.32	217.76
YPS128	112.39	515.55
UWOPS03-461.4	301.16	239.80
CBS432	111.61	214.04
N44	124.78	205.35
YPS138	117.42	201.50
UFRJ50816	109.60	187.20
UWOPS91-917.1	105.07	314.91

Supplementary Table 3. Pilon correction for nuclear genome assemblies.

Strain	Pilon change	SNP/Indel called before Pilon correction*	SNP/Indel called after Pilon correction*
S288C	157		
DBVPG6044	1191	22/1092	6/50
DBVPG6765	292	20/222	15/27
SK1	569	19/510	11/47
Y12	525	25/442	9/36
YPS128	487	28/506	19/72
UWOPS03-461.4	410	114/363	102/100
CBS432	411		
N44	515	102/513	89/99
YPS138	335	103/349	92/93
UFRJ50816	133		
UWOPS91-917.1	99		

Footnote:

*: The SNP/Indel calling was based on independent Illumina sequencing data from what we used in Pilon correction. This independent data set was generated by Bergström et al. MBE (2014), which covered most strains sampled in our current study. The strain S288C, CBS432, and UWOPS91-917.3 were not sequenced. The strain UFRJ50816 was sequenced but the data quality is suboptimal.

Supplementary Table 4. Pilon correction for mitochondrial genome assemblies.

Strain	Pilon change	SNP/Indel called before Pilon correction*	SNP/Indel called after Pilon correction*
S288C	13		
DBVPG6044	71	5/68	2/1
DBVPG6765	11	3/7	0/0
SK1	9	0/9	0/0
Y12	14	2/12	0/0
YPS128	12	0/12	0/2
UWOPS03-461.4	46	4/46	2/2
CBS432	6		
N44	5	0/5	0/0
YPS138	8	0/9	0/1
UFRJ50816	7		
UWOPS91-917.1	5		

Footnote:

*: The SNP/Indel calling was based on independent Illumina sequencing data from what we used in Pilon correction. This independent data was generated by Bergström et al. MBE (2014), which covered most strains sampled in our current study. The strain S288C, CBS432, and UWOPS91-917.3 were not sequenced. The strain UFRJ50816 was sequenced but the data quality is suboptimal.

Supplementary Table 5. Total occurrences of different genomic features annotated in the nuclear genome.

Strain	Centromere	Protein-coding gene*	tRNA	Ty-related**	Core X-element	Y'-element
S288C	16	5,569	273	359	33	23
DBVPG6044	16	5,543	275	440	33	21
DBVPG6765	16	5,538	275	309	31	9
SK1	16	5,542	275	408	32	31
Y12	16	5,547	275	315	38	10
YPS128	16	5,537	275	295	37	10
UWOPS03-461.4	16	5,564	272	397	30	4
CBS432	16	5,530	275	439	36	15
N44	16	5,532	275	439	32	0
YPS138	16	5,512	276	390	31	8
UFRJ50816	16	5,522	275	553	29	9
UWOPS91-917.1	16	5,520	270	518	23	7

Footnote:

*: Pseudogenes are not included.

**: Ty-related features include full-length Tys, truncated Tys and Ty solo-LTRs.

Supplementary Table 6. Total occurrences of Ty-related features annotated in the nuclear genome.

	Full-length Ty					Truncated Ty					Ty solo-LTR			
Strain	Ty1	Ty2	Ty3	Ty4	Ty5	Ty1	Ty2	Ty3	Ty4	Ty5	Ty1/ Ty2*	Ty3	Ty4	Ty5
S288C	40	13	2	3	1	2	0	0	0	0	224	37	30	7
DBVPG6044	20	1	1	0	0	5	0	0	0	0	315	42	36	20
DBVPG6765	0	15	0	0	1	2	2	0	0	0	214	48	21	6
SK1	19	5	1	0	1	6	0	0	0	0	291	37	29	19
Y12	19	2	2	5	2	3	0	0	0	0	216	31	29	6
YPS128	2	11	7	4	0	3	0	0	0	0	205	22	27	14
UWOPS03-461.4	0	0	0	0	0	5	1	0	0	0	250	47	76	18
CBS432	9	0	2	0	7	6	0	0	2	2	281	60	50	20
N44	2	0	5	4	0	5	0	0	2	0	235	70	103	13
YPS138	0	0	0	1	0	3	0	0	0	0	240	42	73	31
UFRJ50816	22	0	0	23	0	4	0	0	2	0	312	38	125	27
UWOPS91-917.1	9	0	0	1	0	1	0	0	1	5	357	24	100	20

Footnote:
* The solo-LTR of Ty1 and Ty2 are highly similar with frequent recombination, therefore we treated them together without further differentiation.

Supplementary Table 7. Total occurrences of different genomic features annotated in the mitochondrial genome.

Strain	Protein-coding gene	tRNA	rRNA	*rnpB*
S288C	8	27	2	1
DBVPG6044	8	27	2	1
DBVPG6765	8	28	2	1
SK1	8	27	2	1
Y12	8	25	2	1
YPS128	8	27	2	1
UWOPS03-461.4	8	26	2	1
CBS432	8	24	2	1
N44	8	25	2	1
YPS138	8	24	2	1
UFRJ50816	8	24	2	1
UWOPS91-917.1	8	24	2	1

Supplementary Table 8. Cumulative lengths of annotated features in the nuclear genome.

Strain	Total length* (bp)	Centromere (bp)	CDS** (bp)	tRNA (bp)	Ty-related*** (bp)	Core X-element (bp)	Y'-element (bp)
S288C	12,157,149	1,891	8,321,070	21,795	425,100	14,316	129,664
DBVPG6044	11,958,041	1,892	8,327,484	21,941	240,470	12,666	89,478
DBVPG6765	11,813,188	1,895	8,305,296	21,941	173,995	13,068	48,304
SK1	12,063,285	1,891	8,335,593	21,941	254,537	13,008	147,944
Y12	11,881,147	1,896	8,317,002	21,941	245,945	15,298	55,199
YPS128	11,910,474	1,893	8,351,790	21,940	209,294	15,234	66,055
UWOPS03-461.4	11,747,346	1,890	8,341,971	21,673	114,234	12,284	21,892
CBS432	12,021,328	1,895	8,357,142	21,955	222,044	14,845	86,862
N44	11,811,689	1,900	8,349,201	21,950	186,199	13,475	0
YPS138	11,811,722	1,889	8,329,080	22,023	107,930	13,749	57,494
UFRJ50816	12,137,283	1,894	8,323,545	21,950	390,671	12,755	61,568
UWOPS91-917.1	11,856,618	1,890	8,242,776	21,570	189,009	9,700	50,470

Footnote:

*: This total assembly length includes a 17,357 bp rDNA gap.

**: The CDSs from pseudogenes are not included.

***: Ty-related features include full-length Tys, truncated Tys and Ty solo-LTRs.

Supplementary Table 9. Cumulative lengths of annotated features in the mitochondrial genome.

Strain	Total length (bp)	CDS (bp)	tRNA (bp)	*rnpB* (bp)	*rns* (bp)	*rnl* (bp)	*COX1* introns (bp)	*COB* introns (bp)
S288C	85,793	6,684	2,041	448	1,649	4,439	11,278	5,951
DBVPG6044	81,093	6,648	2,042	359	1,650	3,287	6,322	5,745
DBVPG6765	81,722	6,651	2,118	458	1,645	3,303	5,905	6,197
SK1	84,638	6,648	2,042	359	1,650	4,425	8,867	5,475
Y12	82,868	6,660	1,855	382	1,646	3,288	10,085	3,748
YPS128	77,479	6,651	2,032	342	1,648	3,286	6,354	5,978
UWOPS03-461.4	74,179	6,654	1,940	393	1,646	3,202	5,037	3,752
CBS432	71,482	6,603	1,785	301	1,616	3,462	5,635	5,125
N44	69,948	6,612	1,865	296	1,614	3,531	5,685	7,398
YPS138	71,396	6,651	1,786	367	1,608	3,583	3,937	3,190
UFRJ50816	77,386	6,660	1,786	376	1,617	3,650	4,638	4,626
UWOPS91-917.1	73,171	6,660	1,786	318	1,625	3,634	4,549	3,190

Supplementary Table 10. Distribution of group I and group II introns in the mitochondrial genome.

(See separate file)

Supplementary Table 11. The chromosomal core boundaries defined based on synteny conservation across the 12 strains.

Chromosome*	Left boundary gene	Right boundary gene
chr01	YAL062W	YAR042W
chr02	YBL107C	YBR297W
chr03	YCL067C	YCR095C
chr04	YDL240W	YDR541C
chr05	YEL066W	YER185W
chr06	YFL050C	YFR055W
chr07	YGL258W	YGR287C
chr08	YHL040C	YHR210C
chr09	YIL166C	YIR038C
chr10	YJL213W	YJR154W
chr11	YKL222C	YKR101W
chr12	YLL060C	YLR460C
chr13	YML131W	YMR319C
chr14	YNL330C	YNR071C
chr15	YOL154W	YOR387C
chr16	YPL272C	YPR193C

Footnote:

* The chromosome number listed here represents the ancestral chromosome identity, which takes into account the large-scale interchromosomal rearrangements in UWOPS03-461.4, UFRJ50816 and UWOPS91-917.1.

Supplementary Table 12. Subtelomeres involved in large-scale chromosomal rearrangements based on their ancestral locations.

Strain	Current location	Ancestral location*
UWOPS03-461.4	chrVII-L	chr10-L
	chrVII-R	chr08-R
	chrVIII-R	chr07-R
	chrX-L	chr13-R
	chrXI-R	chr07-L
	chrXIII-R	chr11-R
UFRJ50816	chrII-R	chr16-R
	chrIV-L	chr11-L
	chrIX-L	chr15-L
	chrXI-L	chr04-L
	chrXII-L	chr14-L
	chrXIII-L	chr14-R
	chrXIV-L	chr12-L
	chrXIV-R	chr13-L
	chrXV-L	chr09-L
	chrXVI-R	chr02-R
UWOPS91-917.1	chrV-R	chr13-R
	chrXIII-R	chr05-R

Footnote:
*: The ancestral location of the subtelomere was determined based on the gene content of its flanking chromosomal core.

Supplementary Table 13. Chromosome-end structure characterized for the 12 strains.

(See separate file)

Supplementary Table 14. Enriched Gene Ontology (GO) terms of genes involved in unbalanced rearrangements.

GO ID	GO Term	Category	FDR	*P*-value
GO:0008554	Sodium-exporting ATPase activity, phosphorylative mechanism	F	1.01E-03	2.73E-07
GO:0008556	Potassium-transporting ATPase activity	F	1.01E-03	2.73E-07
GO:0035725	Sodium ion transmembrane transport	P	2.68E-03	1.09E-06
GO:0006813	Potassium ion transport	P	2.16E-02	4.34E-05
GO:0046870	Cadmium ion binding	F	2.16E-02	4.37E-05
GO:0010273	Detoxification of copper ion	P	2.16E-02	4.37E-05
GO:0071585	Detoxification of cadmium ion	P	2.16E-02	4.37E-05

Footnote:

1. For the category column, “C” stands for cellular component, “F” for biological function, and “P” biological process. The statistical significance was assessed by Fisher’s exact test with FDR correction.
2. For the FDR column, FDR stands for false discovery rate.

Supplementary Table 15. Pairs of subtelomeric duplication blocks within each strain.

Strain	Total	Found in both species	Species-specific*	Strain-specific
S288C	12	10	2	0
DBVPG6044	11	5	6	2
DBVPG6765	11	5	6	4
SK1	8	4	4	2
Y12	14	10	4	2
YPS128	10	8	2	0
UWOPS03-416.4	17	11	6	2
CBS432	21	9	12	9
N44	13	7	6	4
YPS138	14	8	6	3
UFRJ50816	16	10	6	3
UWOPS91-917.1	26	10	16	13
All strains	173	97	76	44

Footnote:

*: The species-specific cases here include those cases that are strain-specific.

Supplementary Data Set 1. One-to-one nuclear orthologous gene groups across the 12 strains and the six outgroups.

(See separate file)

Supplementary Data Set 2. Genomic coordinates for subtelomeres identified in the 12 strains.

(See separate file)

Supplementary Data Set 3. Genomic features enclosed in subtelomeres and chromosome-ends.

(See separate file)

Supplementary Data Set 4. Balanced rearrangement events identified in the 12 strains.

(See separate file)

Supplementary Data Set 5. Unbalanced rearrangement events identified in the 12 strains.

(See separate file)

Supplementary Data Set 6. Subtelomeric duplication blocks identified in the 12 strains.

(See separate file)

Supplementary Data Set 7. Strain-sharing patterns of duplicated subtelomere pairs.

(See separate file)