

Supporting Information

Gordon et al. 10.1073/pnas.1112808108

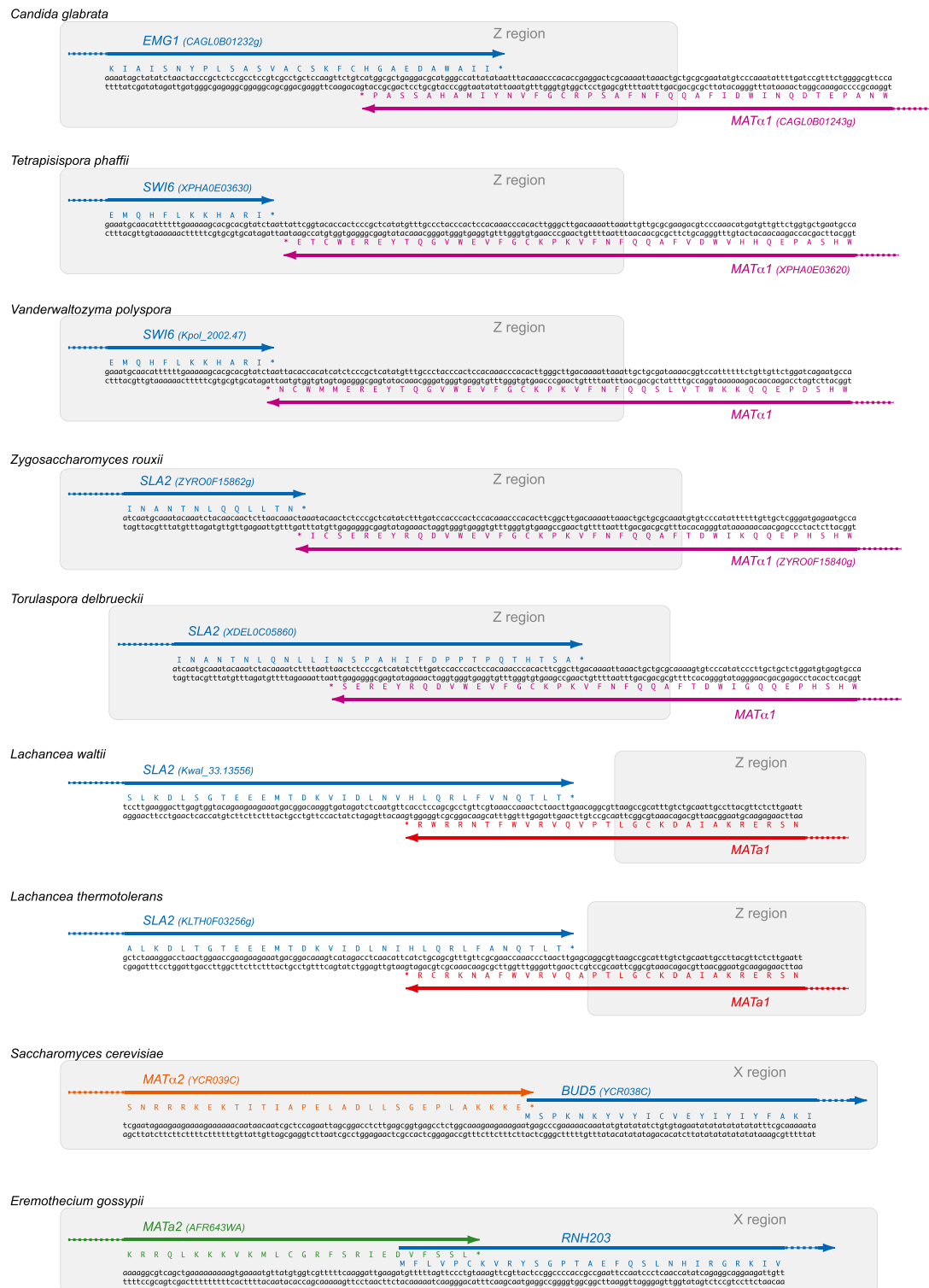
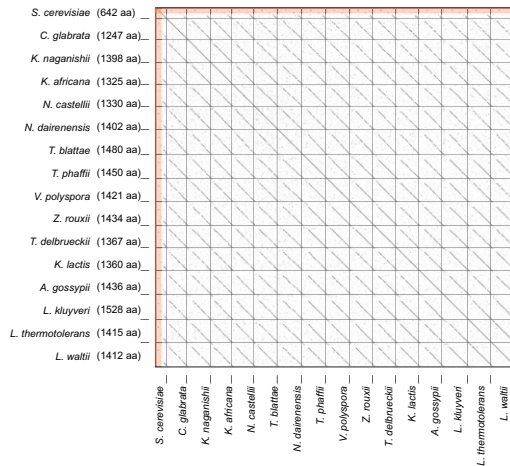
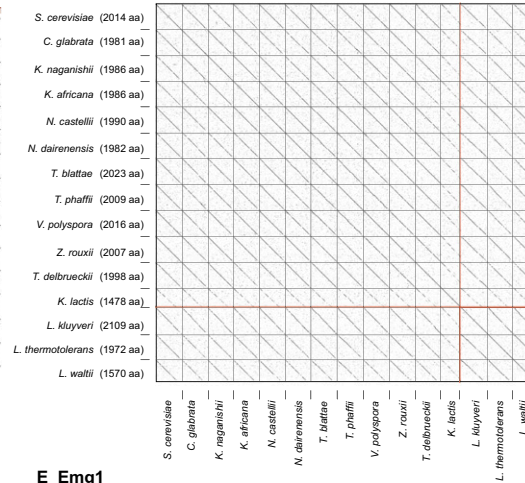


Fig. S1. Details of the gene overlaps indicated in Fig. 2.

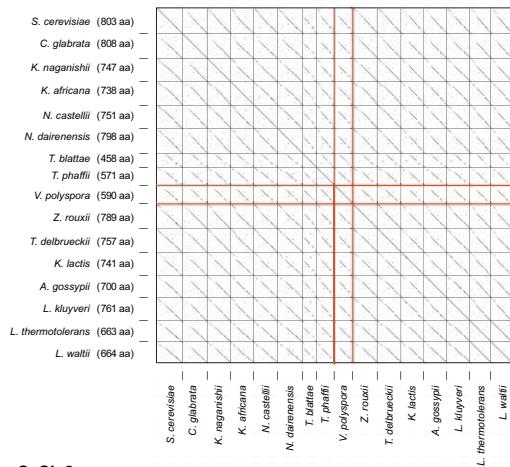
A Bud5



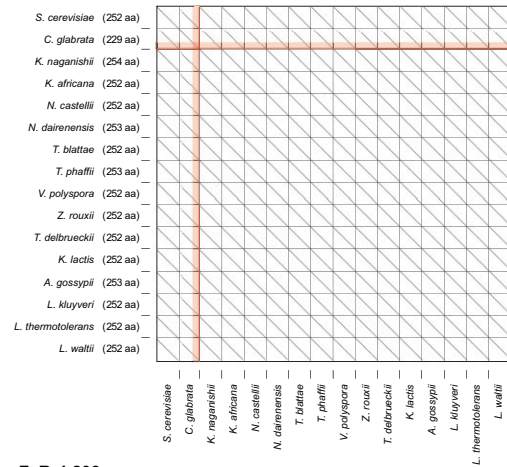
D Laa1



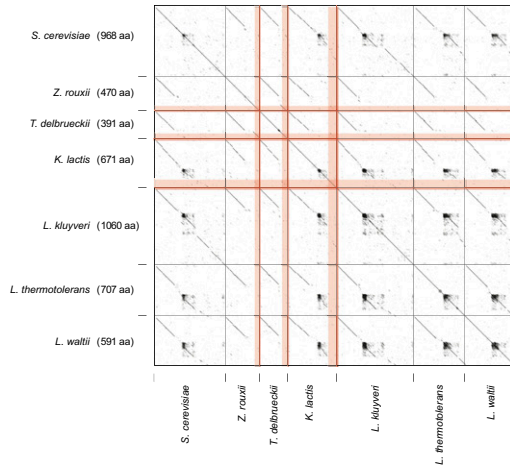
B Swi6



E Emg1



C Sla2



F Rnh203

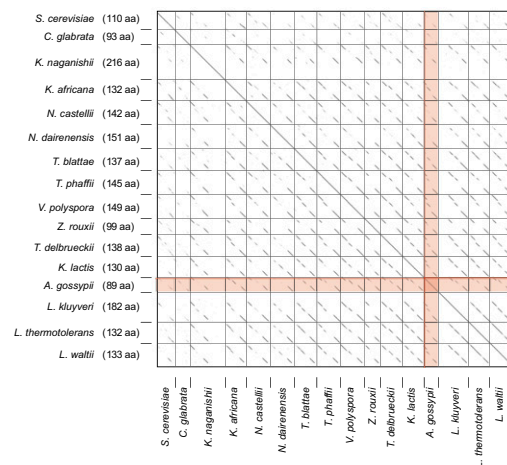


Fig. S2. Truncation of genes flanking the mating-type (*MAT*) locus. Shown are dot-matrix plots of concatenated protein sequences from all species for (A) Bud5, (B) Swi6, (C) Sla2, (D) Laa1, (E) Emg1, and (F) Rnh203. The number of amino acid residues in each protein is indicated. Pink highlighting denotes regions whose coding sequence lies inside the Z or X region. Each cell in the grids is a comparison between two species. If the proteins are alignable along their whole lengths, the diagonal signal runs into the corners of the cell. If one of the proteins is truncated, the signal does not go into a corner. For example, in C, the *Zygosaccharomyces rouxii* Sla2 protein has no region homologous to the C terminus of Sla2 in many other species including *Saccharomyces cerevisiae*; and in A, the *S. cerevisiae* Bud5 protein has no region homologous to the N terminus of Bud5 in many other species including *Candida glabrata*. Bud5, Swi6, Sla2, and Laa1 are severely truncated in some species. Emg1 and Rnh203 are slightly shorter in the species in which they overlap the Z or X region than in other species. In the comparison between the truncated Sla2 proteins of *Torulaspota delbrueckii* and *Z. rouxii*, the Z-overlapping region is seen to be conserved between the two species and the *T. delbrueckii* protein shows a deletion just upstream of the Z region. Plots were made using the program DOTTER (1).

1. Sonnhammer EL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167(1-2):GC1–GC10.

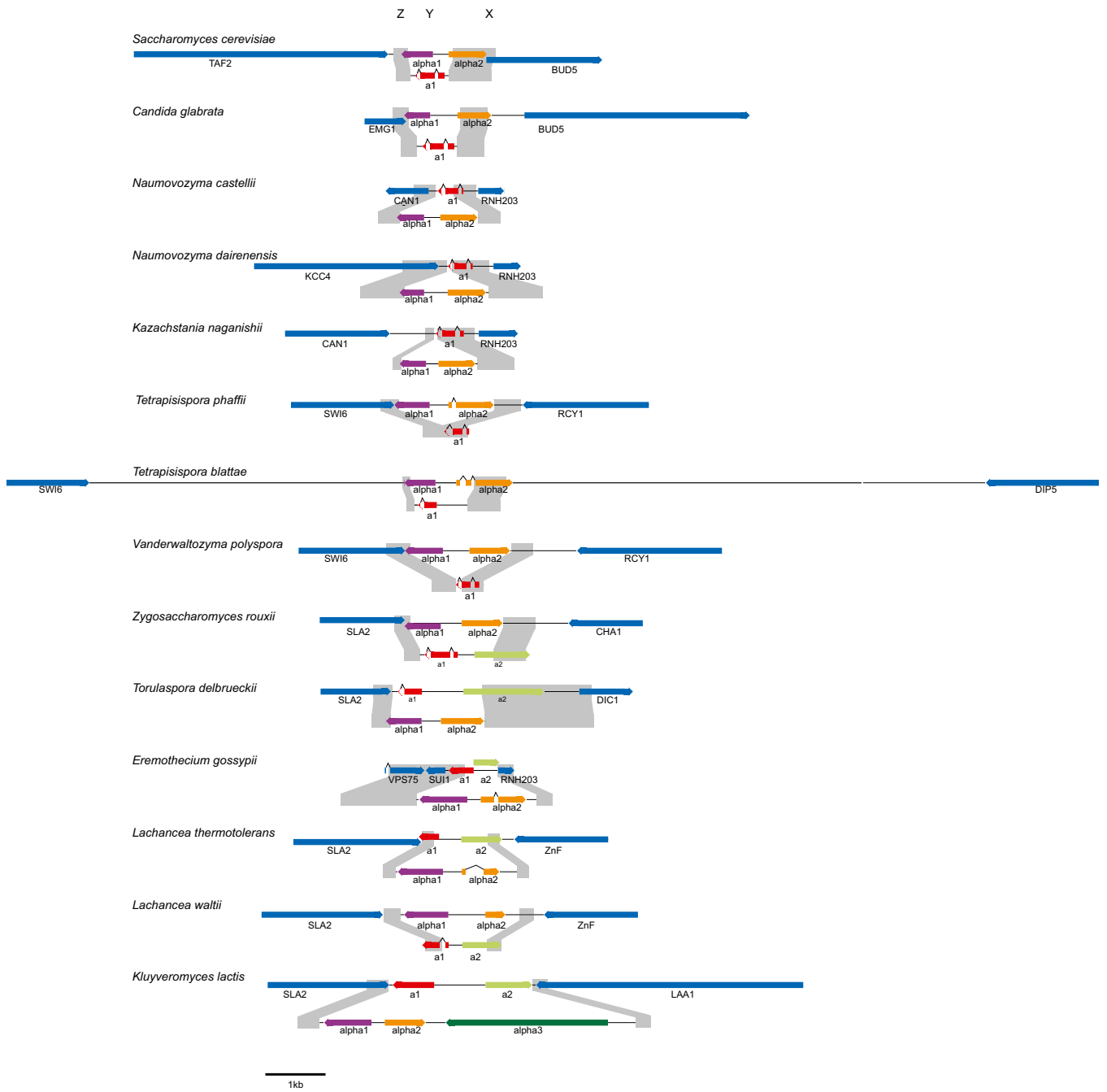


Fig. S3. Scale representation of *MAT* loci. Gray polygons show the extents of the Z and X regions.

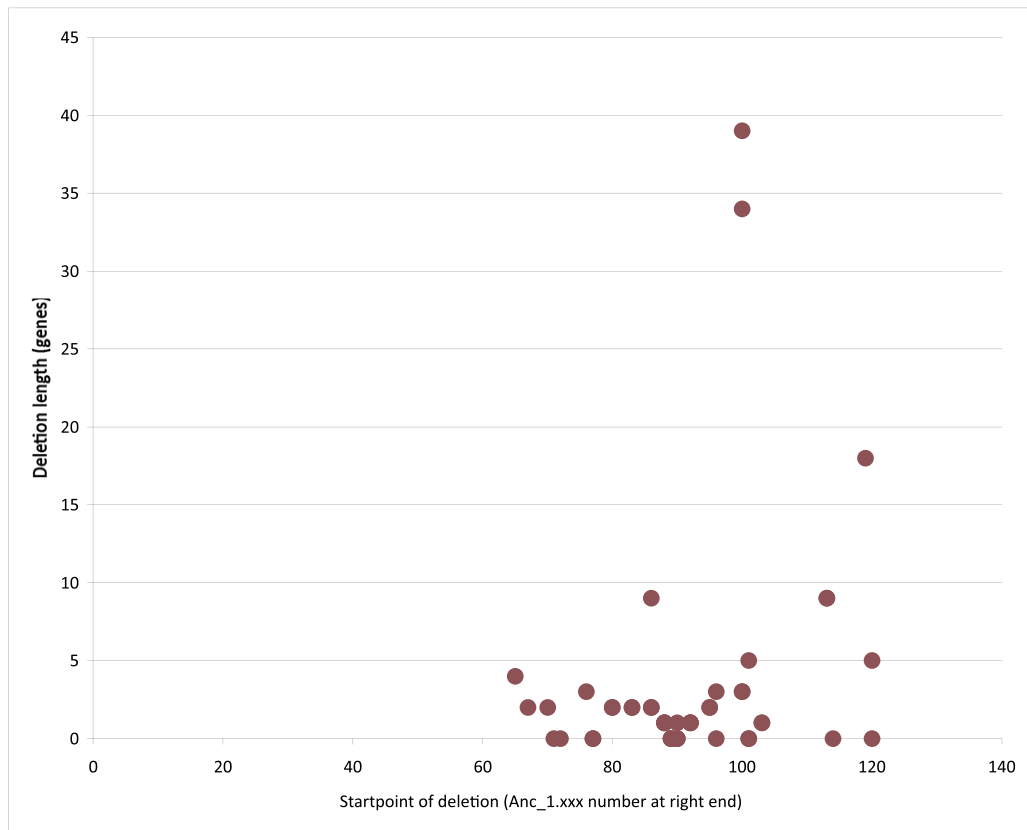


Fig. S4. Relationship between deletion length and deletion position on the Z side of the *MAT* locus. For each deletion event indicated in Fig. 3, its length (number of Ancestral genes removed) is compared with its startpoint (the gene at its end closest to *MAT*). Earlier deletion events (those starting closer to *MAT* at position Anc_1.120) tend to be larger than those occurring later ($r = 0.28$), indicating that the deletion process is slowing down.

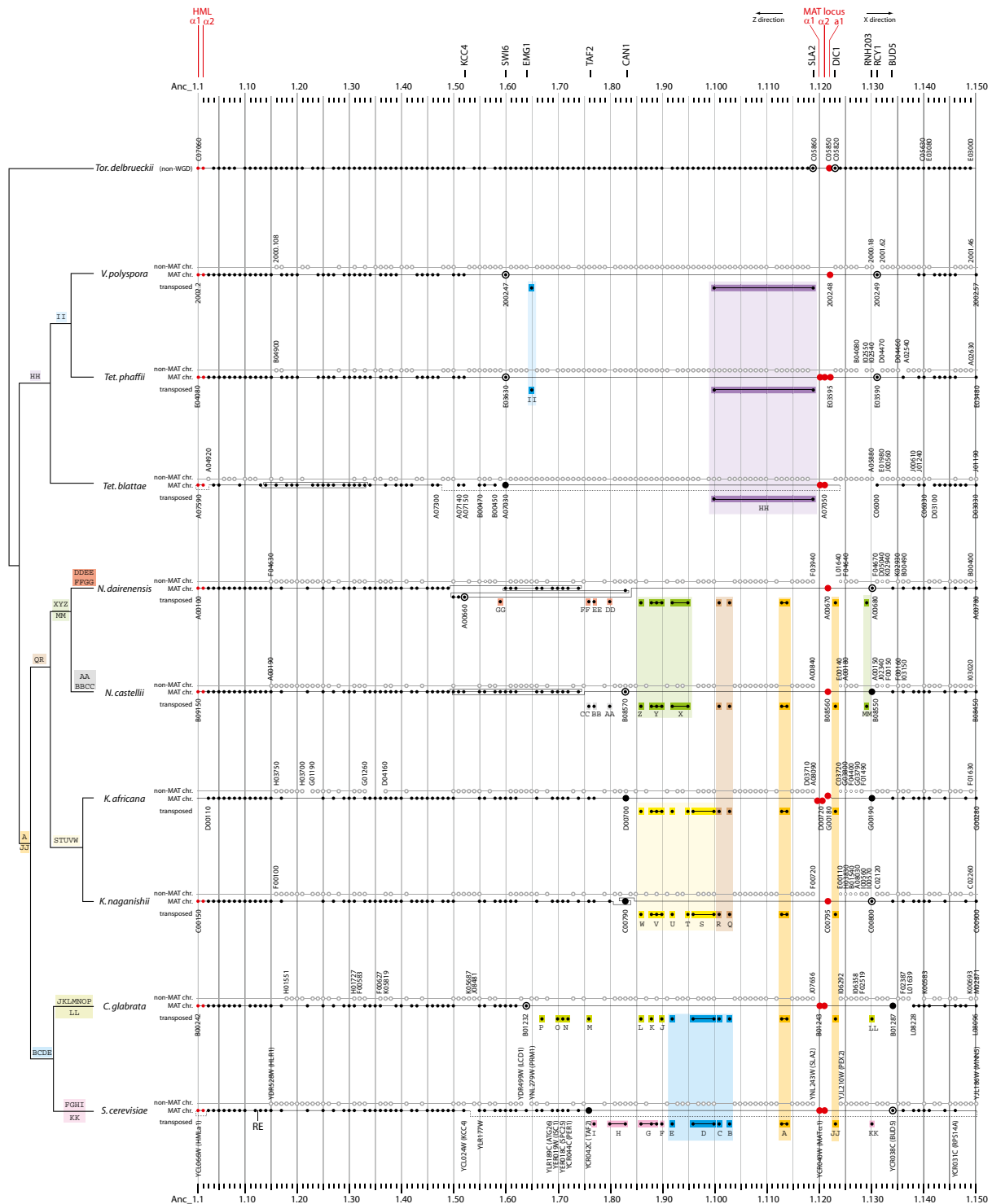


Fig. S5. Detailed version of Fig. 3. Gene names are indicated at the ends of each segment of the *MAT* and non-*MAT* chromosomes that remains intact in each species. The names of transposed genes are given in Table S1. Some small inversions, and genes not in the Ancestral genome, are not shown. Small open circles (e.g., Anc_1.124 in *N. dairenensis*) represent genes whose assignment to the *MAT* or non-*MAT* chromosome was uncertain; these are arbitrarily shown on the non-*MAT* chromosome. Dashed lines denote large inversions in *S. cerevisiae* and *Tetrapispora blattae* that span the *MAT* locus. In both of these species, the inversion of *MAT* has been compensated by a small inversion of the *HML* genes, maintaining the parallel orientation of *MAT* and *HML*. The topology of the phylogenetic tree is based on Hedtke et al. (1), except that we find that *Tetrapispora phaffii* is more closely related to *Vanderwaltozyma polyspora* than to *Tetrapispora blattae* [in PhyML (2) analysis of 30 protein-coding genes retained in duplicate in all post-WGD species (those that underwent whole-genome duplication), *Tetrapispora phaffii* grouped with *V. polyspora* 44 times, and with other species 16 times].

1. Hedtke SM, Townsend TM, Hillis DM (2006) Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Syst Biol* 55:522–529.
2. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.

Table S1. Details of the gene transposition events shown in Fig. 3

Transposition event	Ancestral (old) location of transposed genes (name of <i>S. cerevisiae</i> ortholog in parentheses)	Species with transposition	Transposed gene(s)	Current (new) location*			
				Left neighbor at insertion site		Right neighbor at insertion site	
				Anc name	Gene	Anc name	Gene
A	Anc_1.113 (JJJ3), Anc_1.114 (YJR098C)	Scer	YJR097W (JJJ3), YJR098C	Anc_7.468	YJR096W	Anc_7.470	YJR099W (YUH1)
		Cgla	CAGL0C04565g, CAGL0C04587g	Anc_7.468	CAGL0C04543g	Anc_7.470	CAGL0C04609g
		Knag	KNAG0D01360, KNAG0D01350	Anc_7.468	KNAG0D01370	Anc_7.470	KNAG0D01330
		Kafr	KAFR0G00910, KAFR0G00900	Anc_7.468	KAFR0G00920	Anc_7.470	KAFR0G00890
		Ncas	NCAS0H00550, NCAS0H00540	Anc_7.468	NCAS0H00560	Anc_7.470	NCAS0H00530
		Ndai	NDAI0D00500, NDAI0D00490	Anc_7.467	NDAI0D00510	Anc_7.470	NDAI0D00480
B	Anc_1.103 (COQ6)	Scer	YGR255C (COQ6)	Anc_5.62	YGR254W (ENO1)	Anc_5.55	YGR256W (GND2)
C	Anc_1.101 (TGL2)	Cgla	CAGL0F08239g	Anc_5.62	CAGL0F08261g	Anc_5.76	CAGL0F08217g
		Scer	YDR058C (TGL2)	Anc_3.306	YDR057W (YOS9)	Anc_3.308	YDR059C (UBC5)
D	Anc_1.96 (LIN1), Anc_1.100 (YSP1)	Cgla	CAGL0E04708g	Anc_3.306	CAGL0E04686g	Anc_3.308	CAGL0E04752g
		Scer	YHR156C (LIN1), YHR155W (YSP1)	Anc_5.89	YHR157W (REC104)	Anc_5.94	YHR154W (RTT107)
E	Anc_1.92 (CWC23)	Cgla	CAGL0I02002g, CAGL0I01980g	Anc_5.89	CAGL0I02068g	Anc_5.94	CAGL0I01958g
		Scer	YGL128C (CWC23)	Anc_6.116	YGL127C (SOH1)	Anc_6.227	YGL129C (RSM23)
F	Anc_1.90 (YGL140C)	Cgla	CAGL0I09526g	Anc_1.336	CAGL0I09504g	Anc_6.227	CAGL0I09548g
		Scer	YGL140C	Anc_2.335	YGL141W (HUL5)	Anc_6.241	YGL139W (FLC3)
G	Anc_1.86 (TAM41), Anc_1.88 (TFC4), Anc_1.89 (UFD1)	Scer	YGR046W (TAM41), YGR047C (TFC4), YGR048W (UFD1)	Anc_4.189	YGR044C (RME1)	Anc_4.191	YGR049W (SCM4)
		Scer	YEL064C (AVT2), YEL063C (CAN1)	Anc_6.17	YEL062W (NPR2)	None (telomeric)	YEL065W (SIT1)
H	Anc_1.80 (AVT1), Anc_1.83 (CAN1)	Scer	YOR154W (SLP1)	Anc_5.500	YOR153C (PDR5)	Anc_5.501	YOR155C (ISN1)
		Cgla	CAGL0K08206g	Anc_1.194	CAGL0K08184g	Anc_1.220	CAGL0K08228g
I	Anc_1.88 (TFC4)	Cgla	CAGL0A03630g	Anc_2.398	CAGL0A03586g	Anc_3.122	CAGL0A03652g
		Cgla	CAGL0G03861g	Anc_4.32	CAGL0G03883g	Anc_4.35	CAGL0G03795g
J	Anc_1.86 (TAM41)	Cgla	CAGL0L00297g	Anc_1.176	CAGL0L00319g	None (telomeric)	CAGL0L00227g
		Cgla	CAGL0M11858g, CAGL0M11880g	Anc_7.55	CAGL0M11902g	Anc_7.59	CAGL0M11836g
K	Anc_1.76 (TAF2)	Cgla	CAGL0E06556g	Anc_5.247	CAGL0E06534g	None (telomeric)	CAGL0E06600g
		Cgla	CAGL0E00385g	Anc_6.377	CAGL0E00407g	None (telomeric)	CAGL0E00363g
L	Anc_1.67 (MDL1)	Knag	KNAG0B00490	Anc_3.530	KNAG0B00500	Anc_3.531	KNAG0B00480
		Kafr	KAFR0F04240	Anc_3.530	KAFR0F04230	Anc_6.359	KAFR0F04250
		Ncas	NCAS0E00580	Anc_3.530	NCAS0E00570	Anc_3.524	NCAS0E00590
		Ndai	NDAI0I02990	Anc_3.530	NDAI0I02980	Anc_3.531	NDAI0I03000
		Ndai	NDAI0C00700	Anc_2.570	NDAI0C00680	Anc_2.585	NDAI0C00710
M	Anc_1.71 (SPC25), Anc_1.72 (PER1)	Knag	KNAG0K02480, KNAG0K02490	Anc_3.31	KNAG0K02500	Anc_3.35	KNAG0K02470
		Kafr	KAFR0A08480, KAFR0A08490	Anc_3.30	KAFR0A08500	Anc_3.35	KAFR0A08470
N	Anc_1.70 (ISC1)	Cgla	CAGL0E06556g	Anc_5.247	CAGL0E06534g	None (telomeric)	CAGL0E06600g
		Cgla	CAGL0E00385g	Anc_6.377	CAGL0E00407g	None (telomeric)	CAGL0E00363g
O	Anc_1.103 (COQ6)	Knag	KNAG0B00490	Anc_3.530	KNAG0B00500	Anc_3.531	KNAG0B00480
		Kafr	KAFR0F04240	Anc_3.530	KAFR0F04230	Anc_6.359	KAFR0F04250
		Ncas	NCAS0E00580	Anc_3.530	NCAS0E00570	Anc_3.524	NCAS0E00590
		Ndai	NDAI0I02990	Anc_3.530	NDAI0I02980	Anc_3.531	NDAI0I03000
		Ndai	NDAI0C00700	Anc_2.570	NDAI0C00680	Anc_2.585	NDAI0C00710
P	Anc_1.103 (COQ6)	Knag	KNAG0A07210	Anc_2.572	KNAG0A07200	Anc_2.577	KNAG0A07220
		Kafr	KAFR0I01710	Anc_2.572	KAFR0I01720	Anc_2.577	KAFR0I01690
Q	Anc_1.101 (TGL2)	Ncas	NCAS0H02970	Anc_2.572	NCAS0H02980	Anc_2.585	NCAS0H02960
		Ndai	NDAI0C00700	Anc_2.570	NDAI0C00680	Anc_2.585	NDAI0C00710
		Knag	KNAG0K02480, KNAG0K02490	Anc_3.31	KNAG0K02500	Anc_3.35	KNAG0K02470
R	Anc_1.96 (LIN1), Anc_1.100 (YSP1)	Knag	KNAG0D03580	Anc_4.148	KNAG0D03570	Anc_4.150	KNAG0D03590
		Kafr	KAFR0F03460	Anc_4.148	KAFR0F03450	Anc_4.150	KAFR0F03470
S	Anc_1.95 (ORC5 ¹)	Knag	KNAG0B01890	Anc_1.282	KNAG0B01880	Anc_1.284	KNAG0B01910
		Kafr	KAFR0A01530	Anc_1.282	KAFR0A01520	Anc_1.284	KAFR0A01540
T	Anc_1.92 (CWC23)	Knag	KNAG0C00550, KNAG0C00540, KNAG0C00530	Anc_1.46	KNAG0C00520	Anc_1.48	KNAG0C00560
		Kafr	KAFR0D00470, KAFR0D00460, KAFR0D00450	Anc_1.47	KAFR0D00440	Anc_1.48	KAFR0D00480
U	Anc_1.88 (TFC4), Anc_1.89 (UFD1), Anc_1.90 (YGL140C)	Knag	KNAG0H02120	Anc_2.270	KNAG0H02130	Anc_2.272	KNAG0H02110
V	Anc_1.86 (TAM41)	Knag	KNAG0H02120	Anc_2.270	KNAG0H02130	Anc_2.272	KNAG0H02110

Table S1. Cont.

Transposition event	Ancestral (old) location of transposed genes (name of <i>S. cerevisiae</i> ortholog in parentheses)	Species with transposition	Transposed gene(s)	Current (new) location*			
				Left neighbor at insertion site		Right neighbor at insertion site	
				Anc name	Gene	Anc name	Gene
X	Anc_1.92 (CWC23), Anc_1.95 (ORC5 [†])	Kafr	KAFR0H03410	Anc_2.270	KAFR0H03420	Anc_2.272	KAFR0H03400
		Ncas	NCAS0H03480, NCAS0H03490	Anc_2.661	NCAS0H03470	Anc_2.662	NCAS0H03520
Y	Anc_1.88 (TFC4), Anc_1.89 (UFD1), Anc_1.90 (YGL140C)	Ndai	NDAI0C00200, NDAI0C00190	Anc_2.661	NDAI0C00210	Anc_2.662	NDAI0C00180
		Ncas	NCAS0B08010, NCAS0B08020, NCAS0B08030	Anc_1.231	NCAS0B08000	Anc_1.235	NCAS0B08040
Z	Anc_1.86 (TAM41)	Ndai	NDAI0B05320, NDAI0B05310, NDAI0B05300	Anc_1.231	NDAI0B05330	Anc_1.235	NDAI0B05290
		Ncas	NCAS0B01420	Anc_8.649	NCAS0B01410	Anc_8.651	NCAS0B01430
AA	Anc_1.80 (AVT2)	Ndai	NDAI0E01340	Anc_8.649	NDAI0E01350	Anc_8.651	NDAI0E01330
BB	Anc_1.77 (SLP1)	Ncas	NCAS0J01840	Anc_4.305	NCAS0J01830	Anc_4.308	NCAS0J01850
CC	Anc_1.76 (TAF2)	Ncas	NCAS0H02780	Anc_5.567	NCAS0H02770	Anc_5.569	NCAS0H02790
DD	Anc_1.80 (AVT2)	Ncas	NCAS0B07500	Anc_1.311	NCAS0B07510	Anc_1.312	NCAS0B07490
EE	Anc_1.77 (SLP1)	Ndai	NDAI0I00130	Anc_6.350	NDAI0I00140	<u>None (telomeric)</u>	NDAI0I00120
FF	Anc_1.77 (SLP1)	Ndai	NDAI0F02310	Anc_6.190	NDAI0F02300	Anc_6.192	NDAI0F02320
GG	Anc_1.76 (TAF2)	Ndai	NDAI0B04850	Anc_4.29	NDAI0D04860	Anc_4.30	NDAI0D04840
HH	Anc_1.59 (VTA1), Anc_1.100 (YSP1), Anc_1.119 (SLA2)	Ndai	NDAI0J02120	Anc_4.227	NDAI0J02100	Anc_4.230	NDAI0J02130
		Vpol	Kpol_1028.50, Kpol_1028.49	Anc_6.328	Kpol_1028.51	Anc_6.330	Kpol_1028.48
II	Anc_1.65 (PRM1)	Tpha	TPHA0D03870, TPHA0D03880	Anc_6.328	TPHA0D03860	Anc_6.330	TPHA0D03890
		Tbla	TBLA0I02320, TBLA0I02330	Anc_6.328	TBLA0I02310	Anc_6.330	TBLA0I02340
JJ	Anc_1.123 (DIC1)	Vpol	Kpol_388.9	Anc_4.21	Kpol_388.8	Anc_4.25	Kpol_388.10
		Tpha	TPHA0F00270	Anc_4.24	TPHA0F00280	Anc_4.25	TPHA0F00260
KK	Anc_1.130 (RNH203)	Scer	YLR348C (DIC1)	Anc_4.180	YLR347C (KAP95)	Anc_4.182	YLR350W (ORM2)
		Cgla	CAGL0G01166g	Anc_4.180	CAGL0G01144g	Anc_4.182	CAGL0G01188g
		Knag	KNAG0B05730	Anc_4.180	KNAG0B05720	Anc_4.183	KNAG0B05740
		Kafr	KAFROA06340	Anc_4.180	KAFROA06350	Anc_4.182	KAFROA06330
		Ncas	NCAS0A03540	Anc_4.180	NCAS0A03530	Anc_4.185	NCAS0A03550
LL	Anc_1.130 (RNH203)	Ndai	NDAI0A03380	Anc_4.180	NDAI0A03370	Anc_4.185	NDAI0A03390
		Scer	YLR154C (RNH203)	Anc_8.367	YLR153C (ACS2)	rDNA, Anc_8.376	YLR163C (MAS1)
MM	Anc_1.129 (NCE101)	Cgla	CAGL0F00341g	Anc_1.526	CAGL0F00363g	Anc_1.529	CAGL0F00319
		Ncas	NCAS0D04660	<u>Anc_6.349</u>	Ncas0D04650	<u>Anc_6.377</u>	NCAS0D04670
		Ndai	NDAI0I00300	<u>Anc_6.349</u>	NDAI0I00310	<u>Anc_6.377</u>	NDAI0I00290

These events can be viewed in the Yeast Gene Order Browser database (1) by entering a gene name or Ancestral name.

*In most cases, the left and right neighbors at the insertion site are a few positions apart in the Ancestral genome (they are not perfectly consecutive due to gene loss after WGD). In a few other cases (underlined), the transposition is to a telomeric region, or coincides with a site of rearrangement relative to the Ancestral genome.

[†]The transposed gene is a paralog of *S. cerevisiae* *ORC5*.

1. Byrne KP, Wolfe KH (2005) The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res* 15:1456–1461.