Supplementary Note S2

Phylogenetic relationship among *S. castellii*, *S. cerevisiae* and *C. glabrata*.

Phylogenetic trees reconstructed by a variety of methods from either single-copy orthologous loci (Class 4 in Fig. 2) or double-copy loci (Class 0) consistently show *C. glabrata* branching off outside a clade containing *S. cerevisiae* and S*. castellii*. However, the large numbers of ancestral loci in Class 2B (relative to 2D and 2F) and in Class 3A (relative to 3B and 3C), suggest an alternative topology where *S. castellii* is an outgroup to *C. glabrata* and *S. cerevisiae.* The number of ancestral loci is not homogeneously distributed among Classes 2B, 2D and 2F, nor among Classes 3A, 3B and 3C (χ^2 tests; $P < 0.05$).

In order to determine which of these topologies is correct, we searched for genomic rearrangements that could be phylogenetically informative about the relationship among *S. castellii*, *S. cerevisiae* and *C. glabrata*. The YGOB engine was used to search for all instances of a chromosomal inversion that is present in one track in any two post-WGD species, but absent from the same track in the third post-WGD species and absent from the pre-WGD species. The resulting list of about 200 candidate sites was examined manually. We searched specifically for chromosomal inversions, but in examining the candidate regions we also noticed some other rearrangements (interchromosomal translocations) that are phylogenetically informative. In total, five rearrangements shared by *S. cerevisiae* and *C. glabrata* to the exclusion of *S. castellii* were found, as described in Figures S2.1 and S2.2. No rearrangements supporting the alternative topology were identified, strongly suggesting that the novel phylogeny we propose is the correct one.

In addition, for loci in Class 3 (Fig. 2 in main text) the topology of the gene tree does not depend on the species phylogeny and can be inferred directly from synteny information. We exploited this to examine the ability of various tree reconstruction methods to recover correct topologies and show that all the methods we employed tend to return a topology where *C. glabrata* diverges from the *S. cerevisiae* lineage before *S. castellii*, even when synteny information clearly shows an alternative topology to be correct. This suggests that a systematic bias[1] may be affecting tree reconstruction and that the trees placing *C. glabrata* as an outgroup to *S. cerevisiae* and *S. castellii* are unreliable.

References

1. Phillips, M. J., Delsuc, F. & Penny, D. Genome-scale phylogeny and the detection of systematic biases. Mol Biol Evol 21, 1455-8 (2004).

Figure Legends

Figure S2.1:

Rearrangement #1: An inversion (boxed in red) is shared by Track B of *S. cerevisiae* and *C. glabrata* but not *S. castellii*. The gene order in *S. castellii* Track B is the same as in *K. waltii*, a representative pre-WGD species.

Rearrangement #2: An interchromosomal translocation (boxed in green) is shared by Track B of *S. cerevisiae* and *C. glabrata*, while Track B of *S. castellii* is colinear with the ancestral gene order.

Figure S2.2:
The large arrows in this Figure show the route by which the current genomic organizations are inferred to have arisen.

Rearrangement #3: A reciprocal translocation occurred between Track B of the region shown in (a), and Track B of the region shown in (b), in the common ancestor of *S. cerevisiae* and *C. glabrata*, after it had diverged from *S. castellii*. In region (a) the gene order in *S. castellii* is colinear with the ancestral order, represented by *K. waltii*. In region (b) there is a species-specific rearrangement in *S. castellii* Track B in the interval between genes 611.0d and 657.10, and there are also breaks in all three post-WGD species in Track A. However, the gene order seen in *K. waltii* through region (b) can be deduced to be ancestral because homologs of 611.0d and 657.10 are also linked in a *Z. rouxii* plasmid clone (data not shown).

Rearrangements #4 and #5: Two local chromosomal inversions that subsequently occurred in the green region are shown in (e). (A tandem duplication that produced the genes YOR172W and YOR162C, boxed in purple, may have been involved in these events).

Fig S2.1

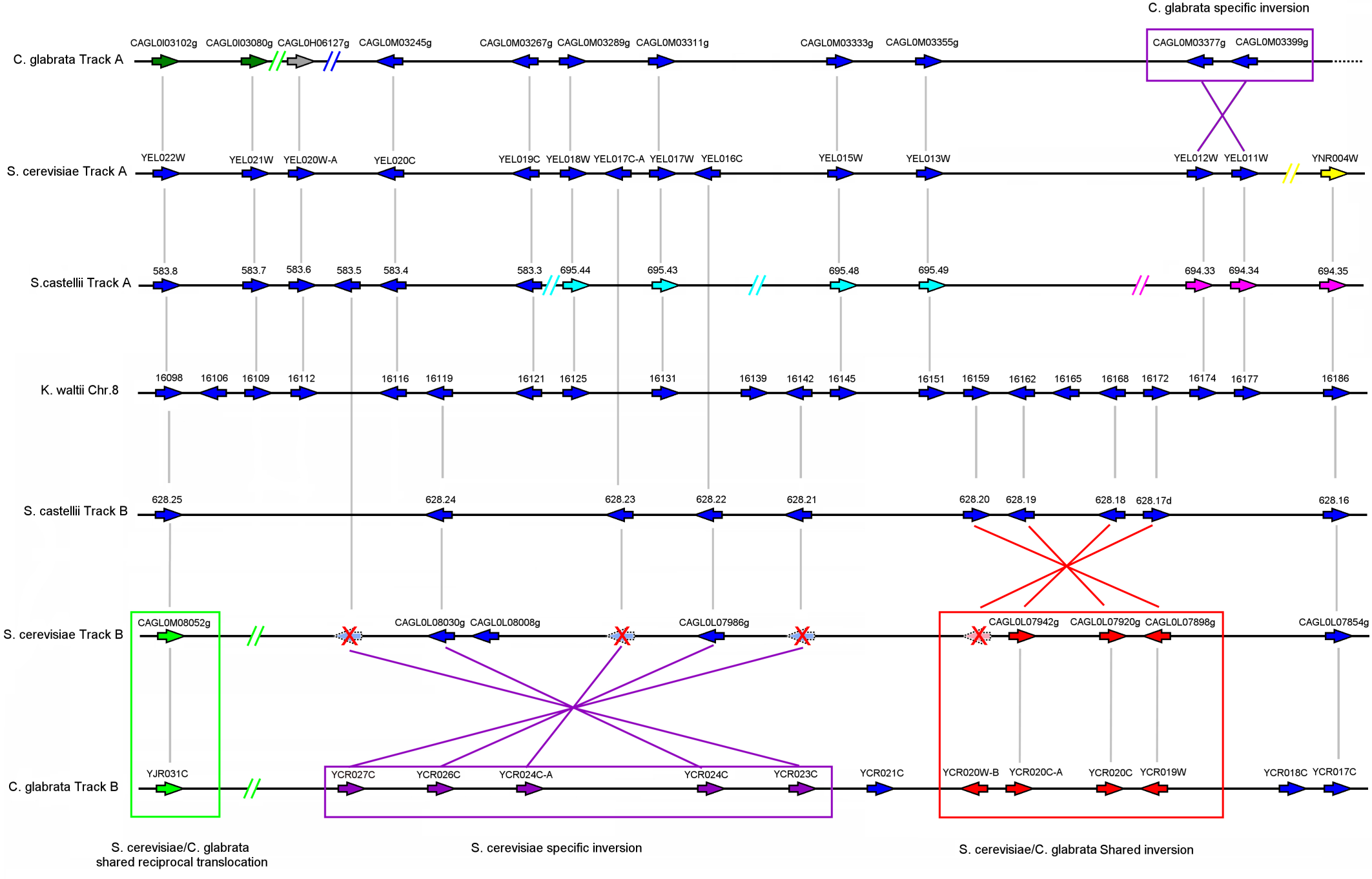
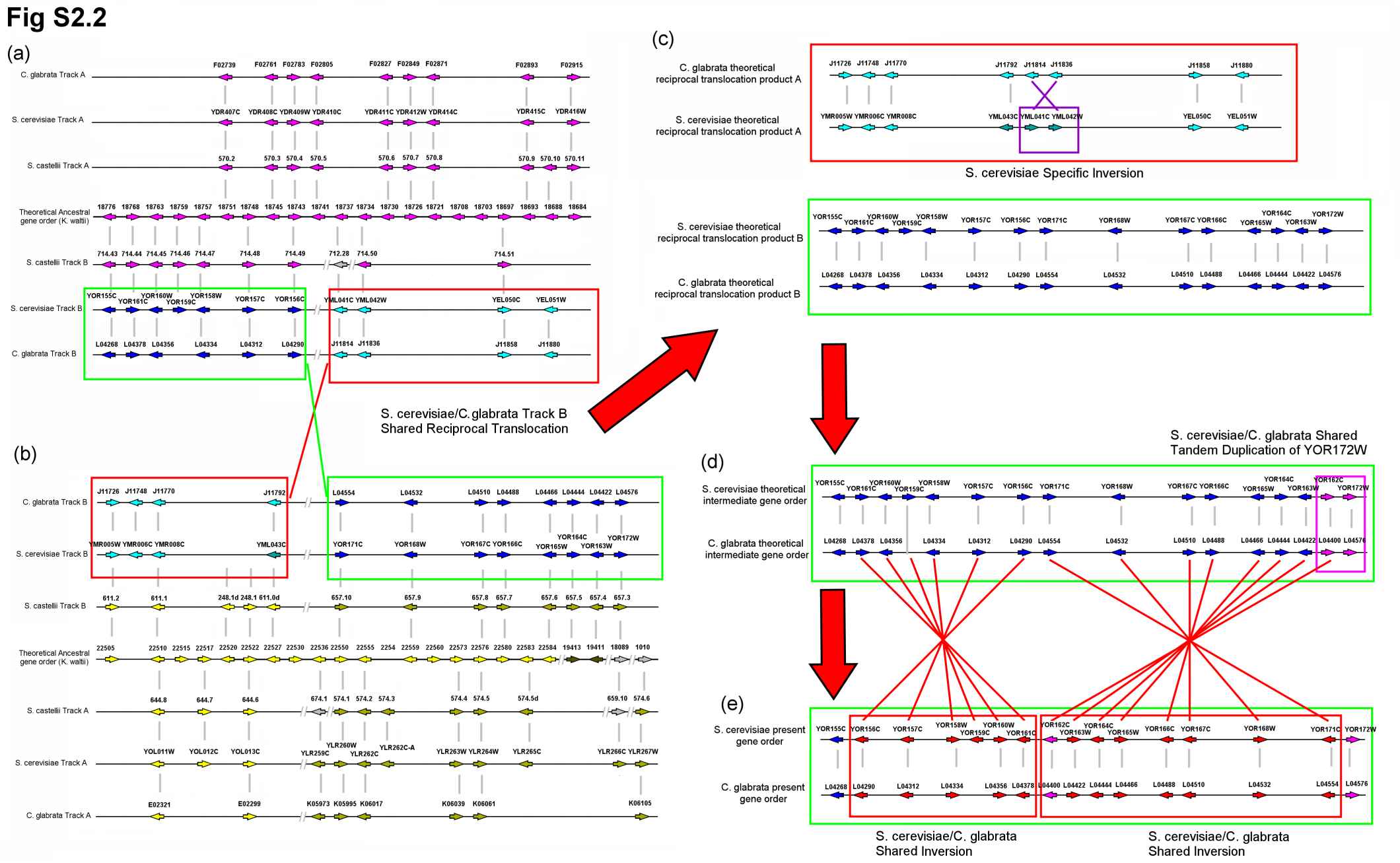

C. glabrata specific inversion
C. glabrata Track A
CAGL0I03102g
CAGL0I03080g
CAGL0H06127g
CAGL0M03245g
CAGL0M03267g
CAGL0M03289g
CAGL0M03311g
CAGL0M03333g
CAGL0M03355g
CAGL0M03377g
CAGL0M03399g
S. cerevisiae Track A
YEL022W
YEL021W
YEL020W-A
YEL020C
YEL019C
YEL018W
YEL017C-A
YEL017W
YEL016C
YEL015W
YEL013W
YEL012W
YEL011W
YNR004W
S.castellii Track A
583.8
583.7
583.6
583.5
583.4
583.3
695.44
695.43
695.48
695.49
694.33
694.34
694.35
K. waltii Chr.8
16098
16106
16109
16112
16116
16119
16121
16125
16131
16139
16142
16145
16151
16159
16162
16165
16168
16172
16174
16177
16186
S. castellii Track B
628.25
628.24
628.23
628.22
628.21
628.20
628.19
628.18
628.17d
628.16
S. cerevisiae Track B
CAGL0M08052g
CAGL0L08030g
CAGL0L08008g
CAGL0L07986g
CAGL0L07942g
CAGL0L07920g
CAGL0L07898g
CAGL0L07854g
C. glabrata Track B
YJR031C
YCR027C
YCR026C
YCR024C-A
YCR024C
YCR023C
YCR021C
YCR020W-B
YCR020C-A
YCR020C
YCR019W
YCR018C
YCR017C
S. cerevisiae/C. glabrata shared reciprocal translocation
S. cerevisiae specific inversion
S. cerevisiae/C. glabrata Shared inversion

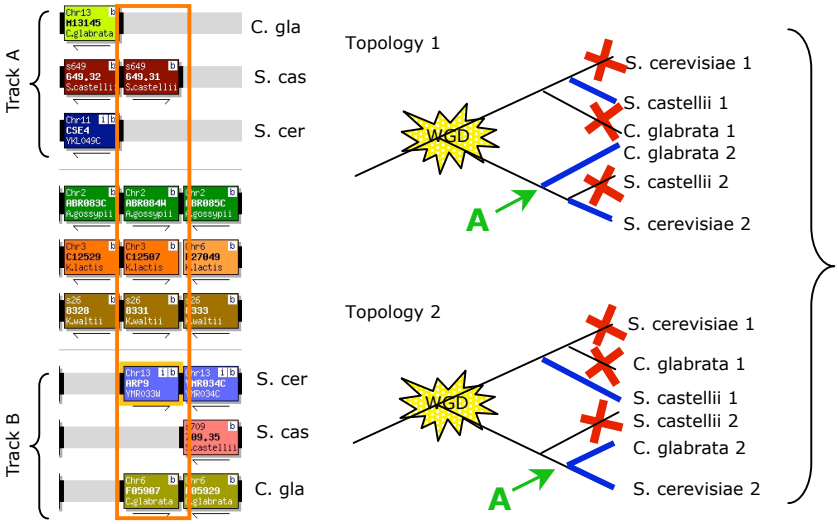
Fig S2.2


(a)
C. glabrata Track A
S. cerevisiae Track A
S. castellii Track A
Theoretical Ancestral gene order (K. waltii)
S. castellii Track B
S. cerevisiae Track B
C. glabrata Track B
S. cerevisiae/C. glabrata Track B
Shared Reciprocal Translocation
(b)
C. glabrata Track B
S. cerevisiae Track B
S. castellii Track B
Theoretical Ancestral gene order (K. waltii)
S. castellii Track A
S. cerevisiae Track A
C. glabrata Track A
(c)
C. glabrata theoretical reciprocal translocation product A
S. cerevisiae theoretical reciprocal translocation product A
S. cerevisiae Specific Inversion
S. cerevisiae theoretical reciprocal translocation product B
C. glabrata theoretical reciprocal translocation product B
(d)
S. cerevisiae/C. glabrata Shared Tandem Duplication of YOR172W
S. cerevisiae theoretical intermediate gene order
C. glabrata theoretical intermediate gene order
(e)
S. cerevisiae present gene order
C. glabrata present gene order
S. cerevisiae/C. glabrata Shared Inversion
S. cerevisiae/C. glabrata Shared Inversion

Phylogenetic trees reconstructed from loci in Classes 3A, 3B, 3C.

At loci in Class 3 we know the true topology of the gene tree, independent of the species tree, because it is shown by the high-quality synteny evidence. One of the remaining genes is a paralog of the other two, so it must be the outgroup. The only situation in which this assumption could be invalid is if one gene copy in a species over-writes the other by gene conversion, but we consider it unlikely that gene conversion could produce the systematically biased results we observe.

Example: *S. cerevisiae ARP9*



S. cerevisiae and *C. glabrata* have retained the same copy of the ancestrally duplicated *ARP9* gene. Under either species topology they have a common ancestor at A while their most recent common ancestor with the *S. castellii* copy is at the WGD. Therefore, irrespective of the species topology, the *S. cerevisiae* and *C. glabrata* genes should be more similar to each other than either is to the *S. castellii* homolog.

Trees were drawn for all Class 3 loci (3A, 3B and 3C) using *K. lactis* as an outgroup.

Phylogenetic Methods

NJ	Phylip [1]	http://evolution.genetics.washington.edu/phylip.html
Parsimony	Phylip [1]	http://evolution.genetics.washington.edu/phylip.html
Quartet Puzzling	Tree Puzzle 2	http://www.tree-puzzle.de/
Maximum Likelihood	Phylip [1]	http://evolution.genetics.washington.edu/phylip.html

ML trees were drawn using JTT + G(4) + I + F

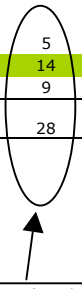
References

1. Felsenstein, J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5:** 164-166.
2. Schmidt, H.A., K. Strimmer, M. Vingron, and A. von Haeseler (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 18:502-504.

For each locus we compared the topology indicated by Synteny information (i.e., the tree we know to be the true tree because reliable synteny data shows one sequence to be a paralog of the other two) to the topology obtained from the Sequence data using a variety of phylogenetic analysis methods as indicated.

Numbers in cells are the number of loci showing a particular combination of Synteny and Sequence topologies.
Green highlighting indicates agreement between the Synteny and Sequence trees.

Phylogenetic method used to generate Sequence tree	Outgroup species as indicated by Sequence tree	Outgroup species as indicated by Synteny information: S. castellii (Class 3A loci)	S. cerevisiae (Class 3B loci)	C. glabrata (Class 3C loci)	Total number of Sequence trees	Number of conflicts between Sequence trees and Synteny trees	Chi-square test for non-homogeneity
NJ	S. castellii	42	6	9	57	15	
	S. cerevisiae	22	12	10	44	32	
	C. glabrata	67	13	32	112	80	
	Total in each Class	131	31	51	213	127	1.77E-12
NJ + 70% Bootstrap	S. castellii	23	2	7	32	9	
	S. cerevisiae	8	8	4	20	12	
	C. glabrata	48	6	23	77	54	
	Total in each Class	79	16	34	129	75	1.01E-11
Parsimony	S. castellii	44	8	11	63	19	
	S. cerevisiae	16	10	4	30	20	
	C. glabrata	64	11	33	108	75	
	Total in each Class	124	29	48	201	114	1.83E-12
Quartet Puzzling	S. castellii	38	5	12	55	17	
	S. cerevisiae	19	12	7	38	26	
	C. glabrata	48	11	26	85	59	
	Total in each Class	105	28	45	178	102	5.67E-07
Maximum Likelihood	S. castellii	49	5	11	65	16	
	S. cerevisiae	19	14	11	44	30	
	C. glabrata	61	9	29	99	70	
	Total in each Class	129	28	51	208	116	1.04E-09



EXAMPLE: Using the Maximum Likelihood method, out of 28 loci in Class 3B (where the *S. cerevisiae* gene is a paralog of the *S. castellii* and *C. glabrata genes* , and so should appear as the outgroup), only 14 of the ML trees correctly recovered *S. cerevisiae* as the outgroup; 9 incorrectly identified *C. glabrata* as outgroup, and 5 incorrectly identified *S. castelli* as outgroup. Overall, for the ML method, 116 of the 208 gene trees were incorrect according to the synteny data, and in 70 of the 116 cases, the incorrectly proposed outgroup was *C. glabrata.*

In the great majority of loci where a conflict is seen between the Synteny and Sequence trees, C.glabrata is the outgroup proposed by the Sequence tree, regardless of phylogenetic method used. This suggests that a systematic bias is causing C. glabratato branch too deeply in phylogenetic trees inferred from sequence data.

The statistical test for non-homogeneity is a chi-square test of the hypothesis that the conflicting trees are uniformly distributed across the three types of Sequence tree.