

Independence of gene losses in two yeast species descended from the same polyploidization event

Kevin P Byrne, Devin R Scannell, A Carolin Frank, Gavin C Conant & Kenneth H Wolfe
Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland
kevin.byrne@tcd.ie



1. Abstract

The genomes of many eukaryotes are paleopolyploids but the factors governing the fate of duplicated genes formed by polyploidy are poorly understood. The relative importance of chance events versus natural selection is unknown. In yeasts, several major clades of species are descended from a polyploid common ancestor.

We find that the process of gene loss in one clade, represented by a draft sequence of the *Kluyveromyces polysporus* genome, has proceeded almost completely independently of gene loss in the *Saccharomyces cerevisiae* clade. The two clades diverged very soon after the whole-genome duplication (WGD). Similar types of genes were retained in duplicate in the two genomes, which confirms the role of natural selection in this process. However within these functional categories that were preferentially retained in duplicate in both genomes, only a minority of the actual genes retained in duplicate overlap, showing the independence of gene loss after polyploidization.

Furthermore, at loci where each species retained only one copy we see an almost random choice of which copy to retain. Thus, the *K. polysporus* genome contains pairs of duplicated chromosomal regions that are superficially similar to those in *S. cerevisiae*, but whose gene composition is very different. Almost half of the genes that are single-copy in both species are paralogs, not orthologs.

2. Motivation & Questions

The WGD [1] in the hemiascomycete yeasts (Figure 1) has had a major impact on the evolution of post-WGD yeasts, notably contributing to their rapid speciation [2].

S. cerevisiae (5500 genes), with 11% retained duplicates [3] and 89% of loci returned to single copy after WGD, is just one observation of the outcome of the yeast WGD.

How much of this post-WGD sorting-out was an inevitable evolutionary outcome and how much was due to stochastic processes?

To answer this question we ideally want to compare genomes that diverged very soon after the WGD.

Genome survey sequencing of *K. phaffii* (Clade 5 in Figure 1) and *K. polysporus* (Clade 6) showed both to be post-WGD genomes, fixing the timing of the WGD.

K. polysporus was selected for further sequencing.

In particular we wish to ask:

1. Are the retained duplicates a special subset of the genome, more amenable to retention?
2. Is the same percentage of genes retained in duplicate in both genomes?
3. Are the same genes retained in duplicate, or is there little overlap?
4. Are the same functional categories retained in duplicate and/or returned to single-copy?
5. For single copy genes, will the same orthologous copy be retained in both genomes?

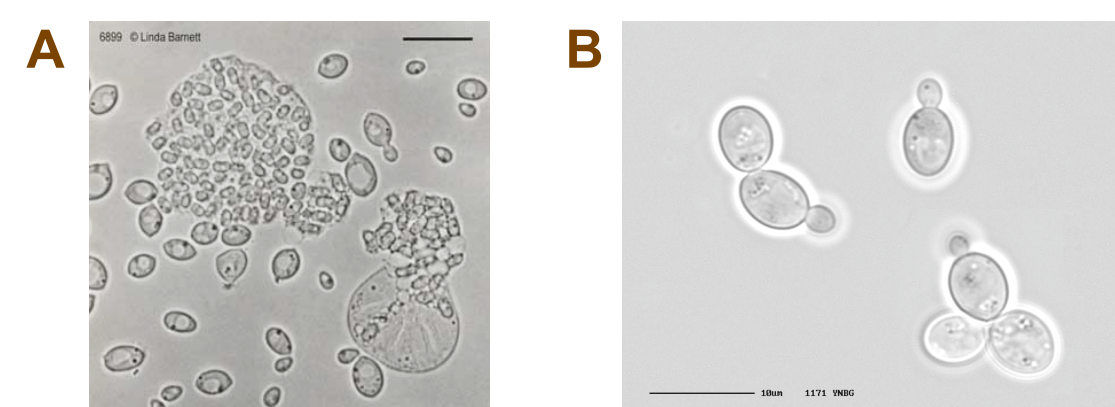
3. K. polysporus Draft Genome Sequence

Kluyveromyces polysporus (Figure 2) is a soil yeast, which is primarily haploid and undergoes extra post-meiotic mitoses to produce 70-100 spores per ascus.

The genome was sequenced to 7.8x coverage (by GATC Biotech) and assembled using the Irish Centre for High-End Computing (www.ichec.ie) facility.

- 290 contigs totaling 14.7Mb
- 40 scaffolds (N50 = 422kb)
- 5652 protein coding genes
- 251 tRNA genes
- > 39 LTR retrotransposons

Figure 2. Photos of yeast cells:
(A) *K. polysporus*
(B) *S. cerevisiae*



4. Massive Reciprocal Gene Loss

We have shown that Dobzhansky-Muller incompatibility by Reciprocal Gene Loss (RGL; filled circles in Figure 4) is associated with post-polyploidy speciations in yeasts [2]. Using our syntenic scoring method as described [2,3] we find massive RGL between *S. cerevisiae* and *K. polysporus* (Table 1 & Figure 3), more than sufficient for reproductive isolation and highlighting an almost entirely independent sorting-out of gene loss in the two genomes.

Table 1

	3252 ancestral loci	S. cer	
		2 copy	1 copy
<i>K. pol</i>	2 copy	212 (6%)	238 (7%)
	1 copy	221 (7%)	2581 (80%)

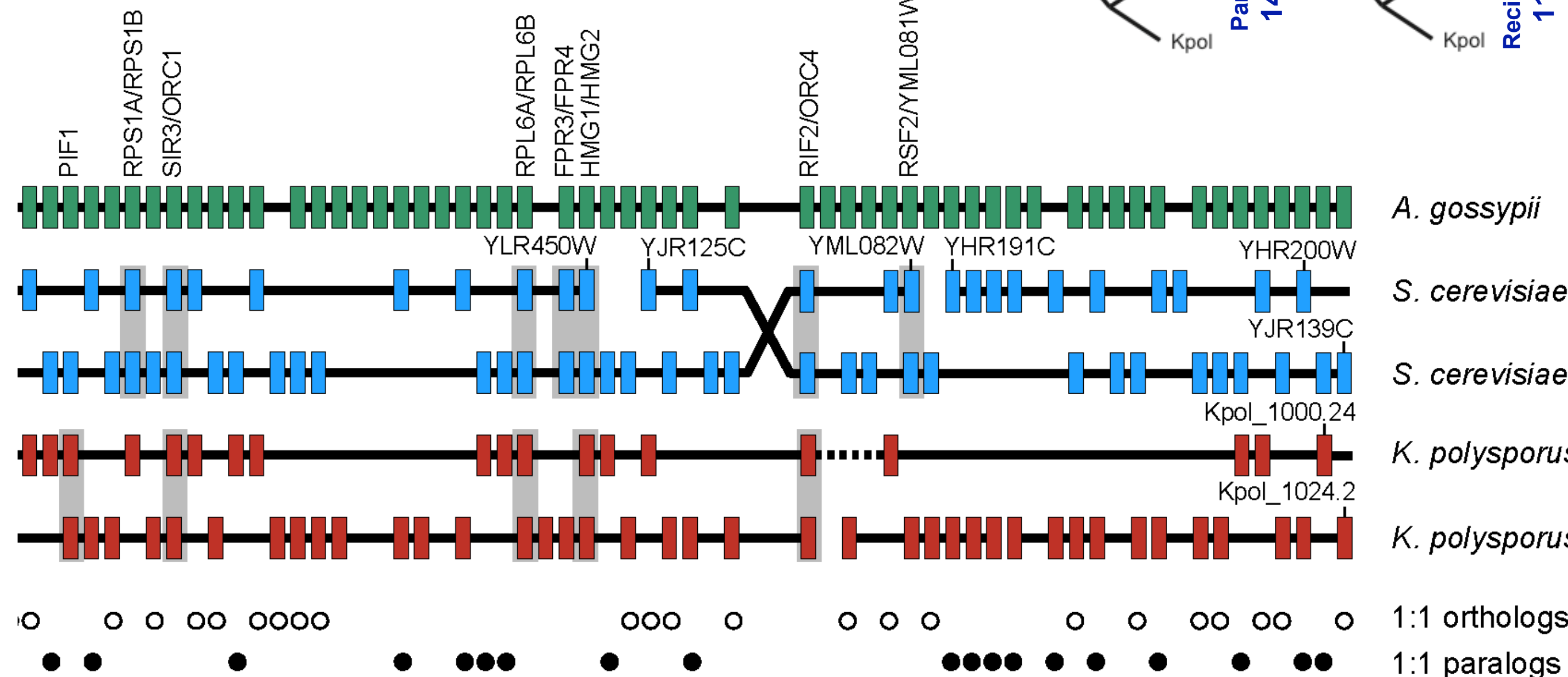
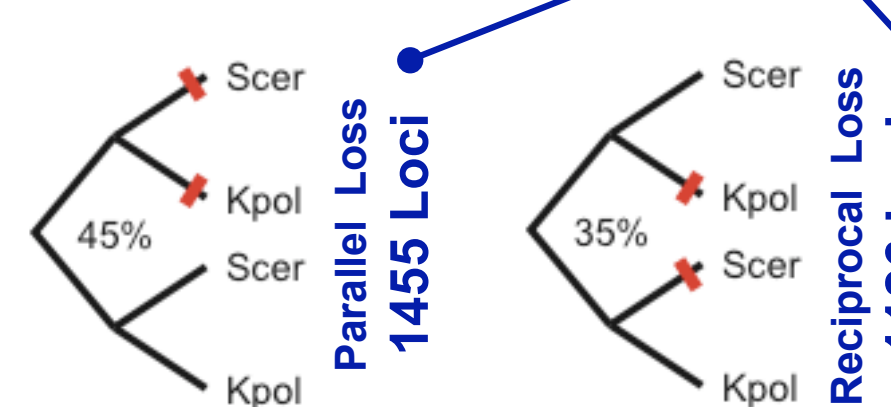
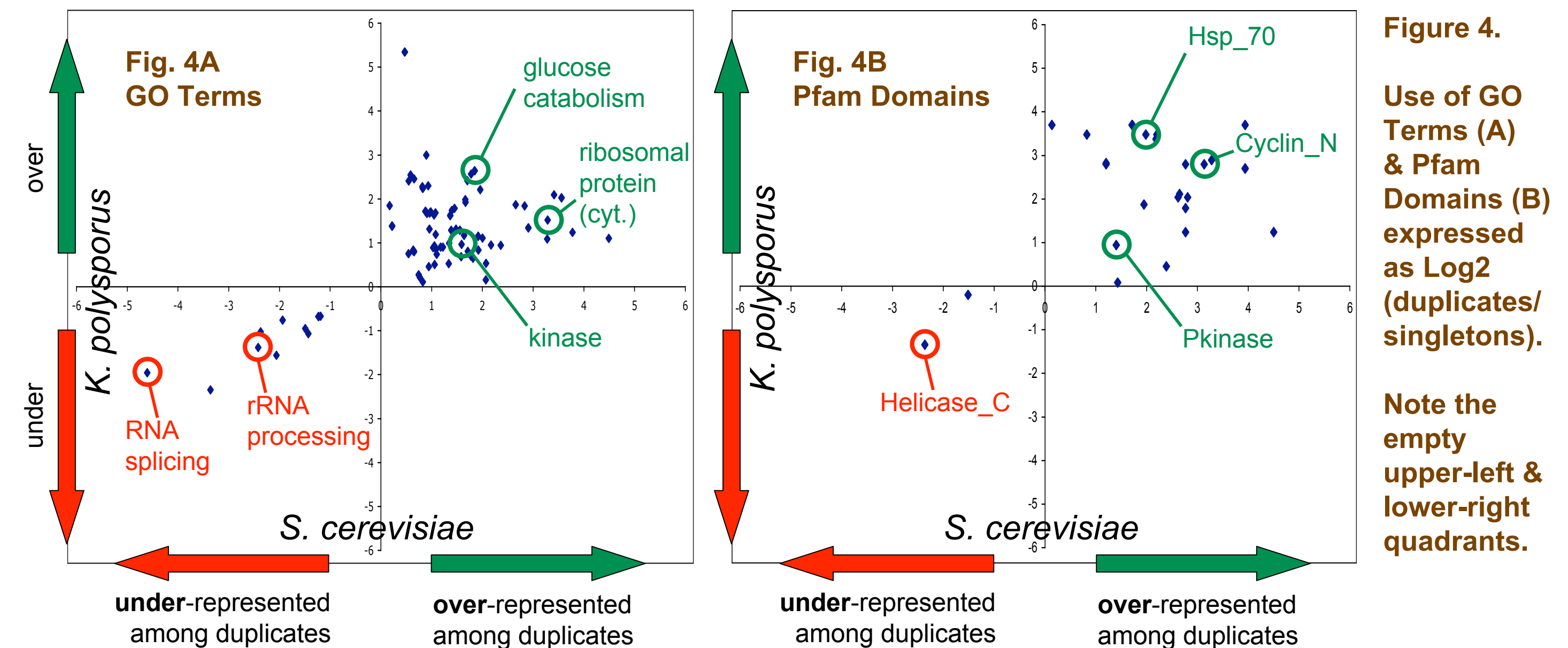


Figure 3. Schematic view of the genomic region near the duplicate pair SIR3/ORC1 based on YGOB output [3]. The green track represents the pre-WGD *A. gossypii* genome and the two blue and red tracks represent the two homologous sister regions in *S. cerevisiae* and *K. polysporus* respectively. Notice the massive amount of reciprocal gene loss - there are 23 single-copy orthologs (columns with open circles underneath) and 19 single-copy paralogs (filled circles). Grey bars highlight retained duplicates in a genome, black lines denote adjacency.

5. Very Similar in Coarse-Grained Analysis

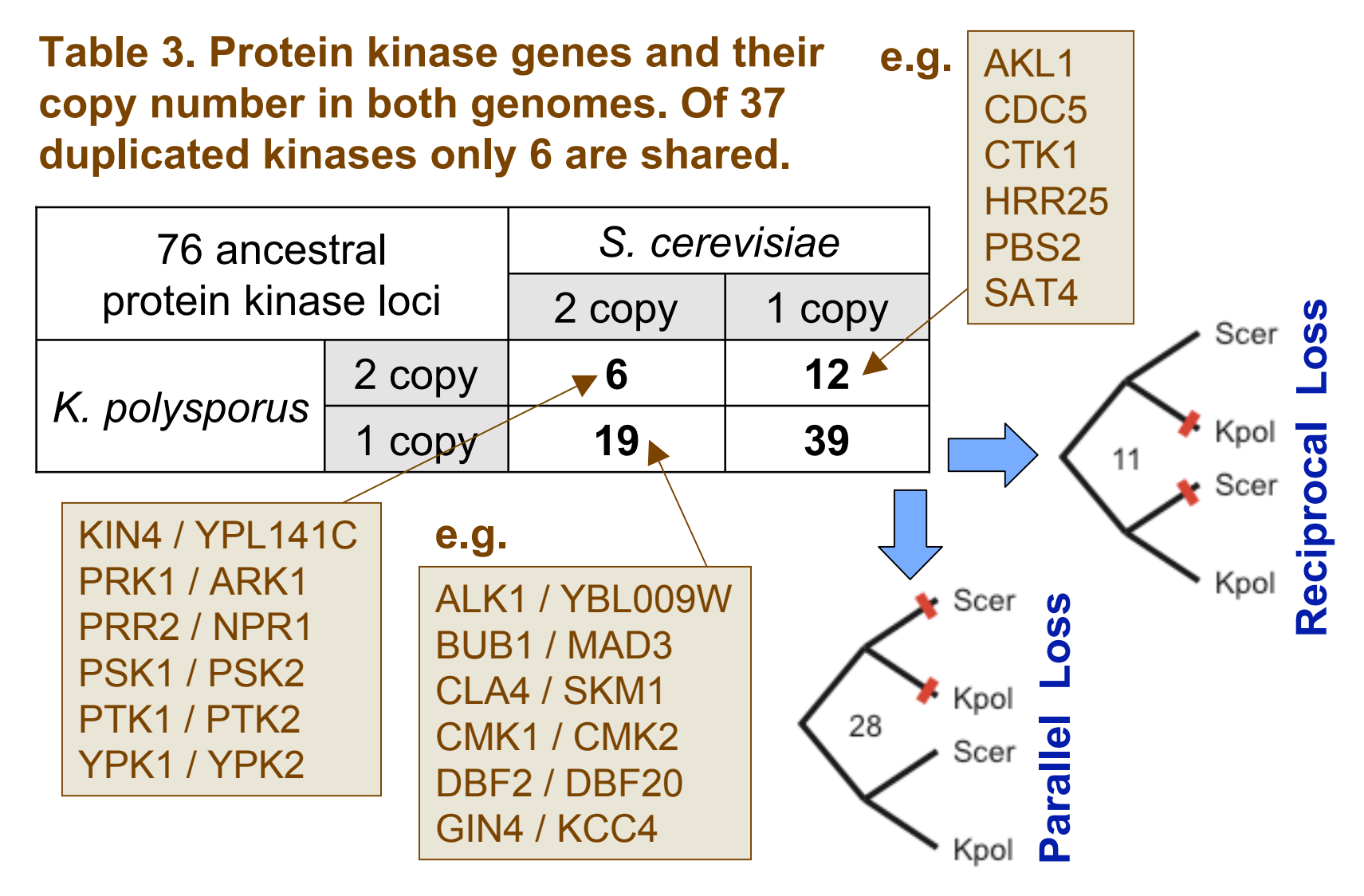
At a high-level we see similar types of retained duplicates in both genomes, with the same functional biases (Table 2 & Figure 4). The numbers of retained duplicates are similar too:

<i>K. polysporus</i>	450 pairs / 3252 loci	13.8%			
<i>S. cerevisiae</i>	433 pairs / 3252 loci	13.3%	Single Copy	<i>K. polysporus</i> Duplicates	<i>S. cerevisiae</i> Duplicates
Table 2. The three Gene Ontology (GO) terms most significantly over-represented among the retained duplicates in both genomes.			Ribosome (cytosolic)	4%	7%
			Carbohydrate metabolism	3%	7%
			Kinase activity	3%	6%
					7%



6. Very Different in Fine-Grained Analysis

Although from a distance the number and type of retained duplicates is very similar in both genomes, nevertheless the actual genes retained are very different. Overall 6% of the traceable loci are duplicated in both genomes, out of the 13% that are duplicated in either.



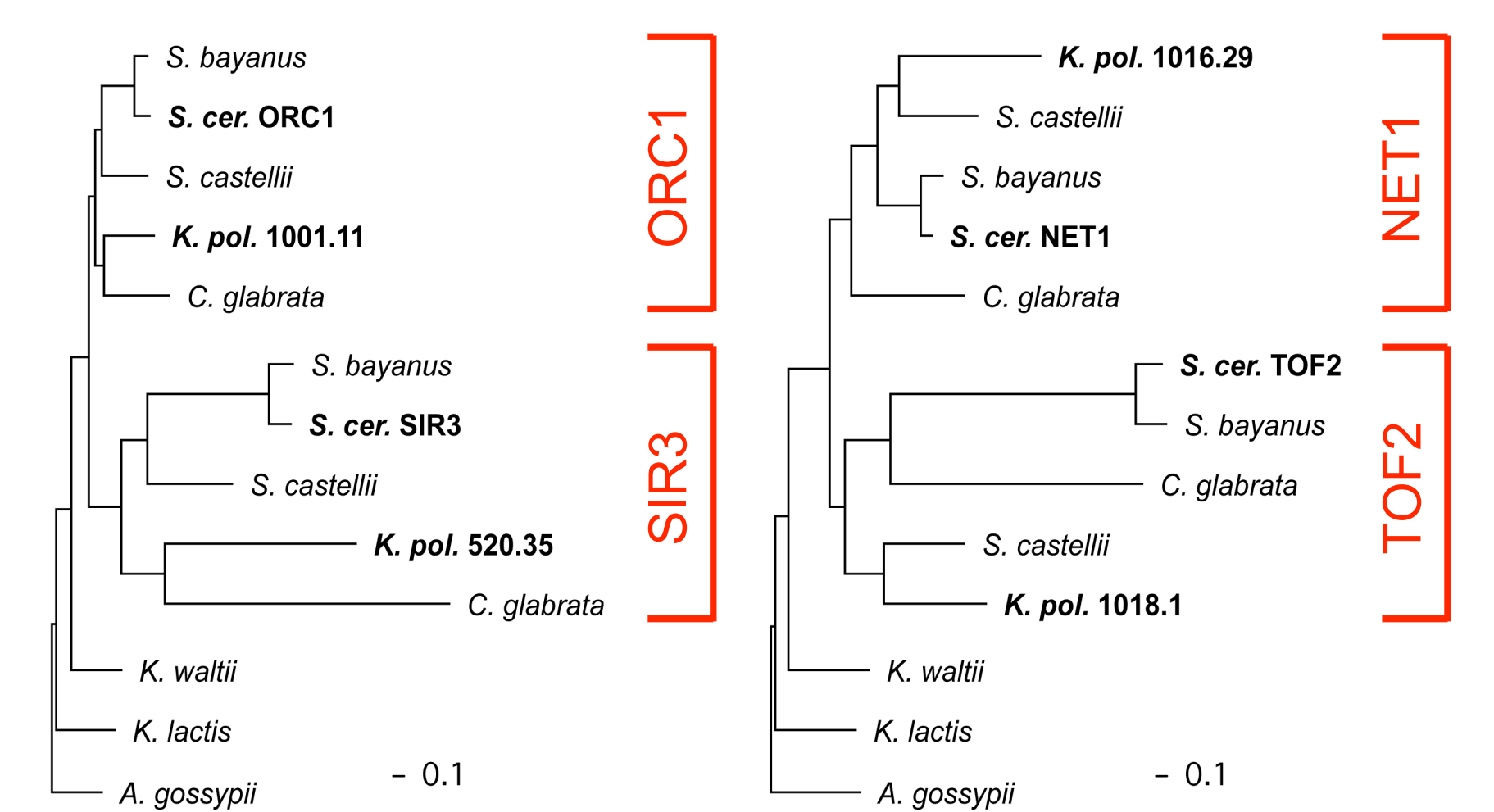
However the differences are most striking if we look at particular groups of genes, for example protein kinase genes, which are very significantly overrepresented among retained duplicates in both species.

7. Two Separate WGDs or One Shared WGD?

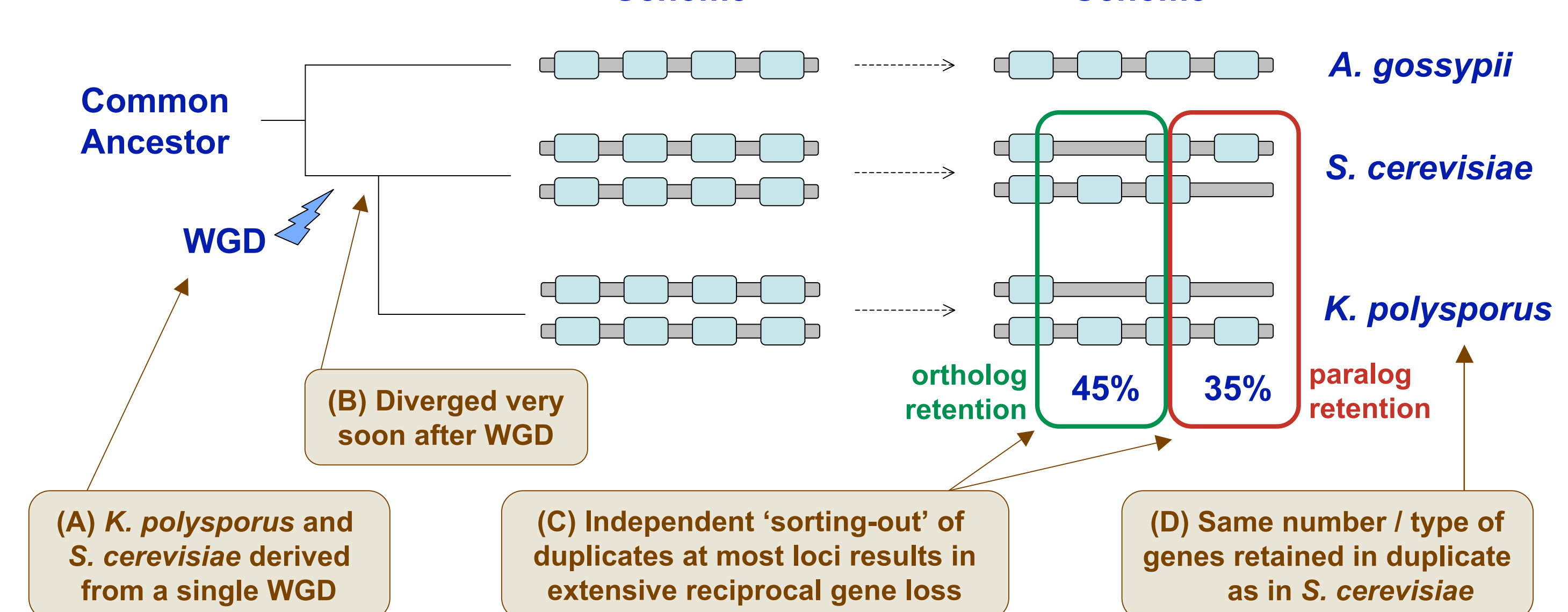
The very different patterns of gene loss in *K. polysporus* as compared to *S. cerevisiae* mean we have to ask whether they underwent two separate polyploidizations as opposed to one common polyploidization. A number of pieces of evidence support the latter hypothesis.

- Most convincingly, trees drawn for shared duplicates have a shared branch (Figure 5)
- Our probabilistic model of gene loss (not discussed here) has a significant branch between a common WGD and speciation
- Trivially, one WGD is the most parsimonious explanation

Figure 5. Examples of shared duplicates gene trees, supporting the hypothesis that there was one common WGD event.



8. Summary



References

1. Wolfe, K.H. and D.C. Shields. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387: 708-713.
2. Scannell D.R., K.P. Byrne*, J.L. Gordon, S. Wong and K.H. Wolfe. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440: 341-345.
3. Byrne, K.P. and K.H. Wolfe. 2005. The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* 15: 1456-1461.
4. Kurtzman, C.P. and C.J. Robnett. 2003. Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. *FEMS Yeast Res.* 3: 417-432.

Acknowledgements

Many thanks to B Cusack, J Gordon, N Khaldi, M Woolfit, M Webster & M Sémon for discussion. This study was supported by Science Foundation Ireland.

