# Patterns of Gene Deletion following Genome Duplication in Yeast

Jake K. Byrnes and Wen-Hsiung Li
*Department of Ecology and Evolution, University of Chicago*
*byrnes@uchicago.edu and wli@uchicago.edu*

## Abstract

*Whole genome duplication (WGD) is followed by massive duplicate deletion that reorganizes gene adjacencies. We compare the deletion patterns and adjacency reorganization following WGD in yeast with simulations. We find that deletion events alternate between paralogous chromosomes more often than expected under a random duplicate deletion model.*

## 1. Introduction

Whole genome duplication creates significant amounts of redundant genetic material, much of which is subsequently removed by deletion [3]. Deletion of gene duplicates from opposing duplicate chromosomes may separate gene pairs that were formerly adjacent and create new adjacencies. Adjacent gene pairs fall into four possible classes based on transcriptional orientation; divergent (<-- -->), convergent (--> <--), Watson tandem (--> -->) and Crick tandem (<-- <--). Previous studies have demonstrated that adjacent genes have regulatory interactions that are affected by orientation [2]. From the genomic sequences of six yeast species we quantify the percentage of adjacent gene pairs that are divergent or convergent (%DC) and the observed distribution of deleted block lengths. We compare these data to simulated results under two deletion models in which the duplicate copy of each gene that is deleted is randomly selected. First we estimate that prior to the duplication event, %DC was ~56%. Following duplication and deletion, this percentage dropped to 52%. In addition, the distribution of deleted block lengths is heavily biased toward small deletions. Our simulations reveal that neither model tested accounts for both the reduction in %DC and the deleted block length distribution. In particular, deletion events alternate duplicate chromosomes more than expected under either model. This suggests that duplicate deletions may not occur at random with respect to paralogous chromosome.

## 2. Materials and Methods

Six yeast species were analyzed for %DC. Three of these species diverged prior to the yeast genome duplication (*Kluveromyces waltii*, *Kluveromyces lactis*, and *Ashbya gossypii*) and three are on the lineage that experienced the duplication event (*Saccharomyces cerevisiae*, *Saccharomyces bayanus*, and *Candida glabrata*). An orthology map between *S. cerevisiae* and *A. gossypii* (Dietrich *et al*. 2004) was used to determine the pattern of duplicate deletion in *S. cerevisiae*.

We used a PERL simulation to examine the effects of two deletion models on %DC and the deletion pattern. First, we created a duplicated version of the *A. gossypii* genome, the most well-annotated genome not undergoing duplication. Next, a deletion model was applied to the genome. Each deletion event involves a series of steps. First, a prospective deletion block length, location, and paralogous chromosome copy (one or two) are chosen. Next, the deletion is tested to see if it removes the second copy of a duplicate pair. If so, the deletion is assumed deleterious and is not accepted, otherwise it is accepted. Deletion events continue until 90% of the gene pairs have returned to single copy (as seen in the real data). Finally we record the resulting %DC and the final observable pattern of deletion for the simulated genome. This completes one of ten thousand iterations used for each set of parameters examined.

## 3. Results

Analysis revealed that the %DC prior to WGD was ~56% while the %DC in the genomes that experienced WGD and deletion was reduced to ~52%.

We examined two different deletion models with different parameters settings. The first model involved drawing prospective deletions from a uniform distribution with a maximum block size of one, two, or three genes. The second model involved drawing prospective deletions from a Poisson distribution with a mean block length of one, two, or three genes. It is clear that deletion models composed of smaller deletions reduce the %DC more, but none of the tested

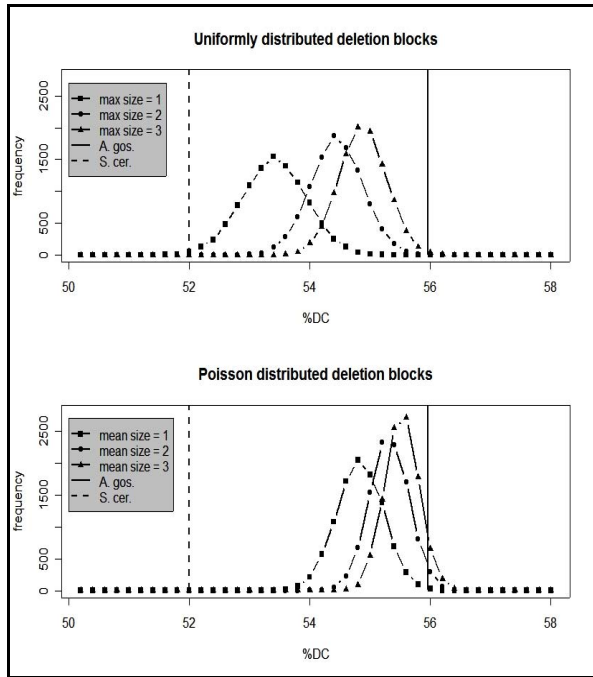models could reduce the %DC from 56% to 52% as seen in the real data (Figure 1).



**Fig. 1. %DC hists under the simulation.** Neither uniform block deletion nor Poisson block deletion could reduce the %DC as seen in the real data.

Our simulations also demonstrated that the observable deletion distribution is significantly different from the distribution of the underlying mutational process. Thus, we compared the resulting observable deletion distribution under our two models with the pattern from *S. cerevisiae* (Figure 2). The deletion distributions were significantly different for both models than that for *S. cerevisiae*. In particular, the simulated genomes contained a paucity of small deletions, and an abundance of large deletions.

## 4. Conclusions

Duplicate deletion alternates between paralogous chromosomes far more often than expected by any model based on a random choice of which duplicate gene copy to delete. This is apparent in both the change in %DC and the distribution of deleted blocks. Neither deletion model based on random duplicate deletion reduces the %DC by the amount observed in the real data, nor produces the deletion pattern of the real data. This suggests that the deletion of duplicates is not occurring randomly. We propose that the alternation of deletions may be selectively

advantageous in creating greater amounts of intergenic space and decoupling the transcription of adjacent genes. Future work will address this possibility.
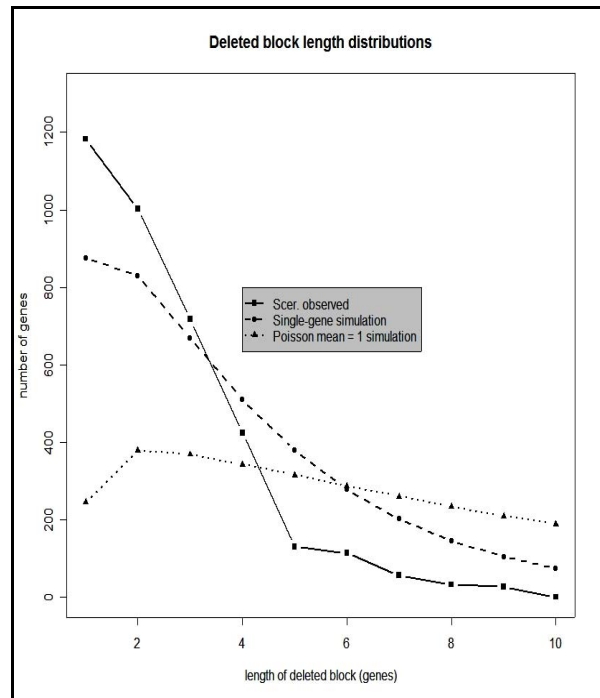


**Fig. 2. Deletion patterns for the single gene model, the Poisson mean = 1 model and *S. cerevisiae*.** Simulations contain too many large deletions and too few small deletions when compared to the data for *S. cerevisiae*

## 10. References

[1] F.S. Dietrich, *et* al., "The *Ashbya gossypii* Genome as a Tool for Mapping the Ancient *Saccharomyces cerevisiae* Genome." *Science*, 2004, 304:304-307.

[2] S. Kruglyak and H. Tang, "Regulation of adjacent yeast genes."*Trends in Genetics*, 2000, 16:109-111.

[3] K. H. Wolfe, "Yesterday's Polyploids and the Mystery of Diploidization." *Nature Rev.*, 2001 2:333-34.