Increasing Rates of Gene Loss and Biased Sub-Genome Evolution Post Whole Genome Duplication

Benjamin L. S. Furman, Utkarsh Dang, Ben J. Evans and G. Brian Golding

Department of Biology, McMaster University

Objectives

After whole genome duplication:

1. assess selective pressures shaping genes in different sub-genomes
2. calculate the rate of duplicate gene loss in different time periods

Introduction

Whole genome duplication (WGD) leads to complete genetic redundancy. As such, gene copies experience reduced selective pressure to be maintained, and mutations will accumulate. Most often gene copies are lost (pseudogenized) presumably soon after whole genome duplication, as gene functions are entirely redundant (*1*). However, recent analyses suggest that gene loss rates may change over time, and may not be highest early on (*2*).

Xenopus

African clawed frogs of the *xenopus* sub-genus are minimally tetraploid, formed by the union of two extinct diploid ancestors. Thus, these tetraploids contain two sub-genomes, referred to as the L and S. A recent genome sequence of *X. laevis* determined that the S and L are evolving differently (e.g. more gene loss in S) (*3*), but it is unclear when these differences accumulated, or if differences are a common feature across species.

Time Periods

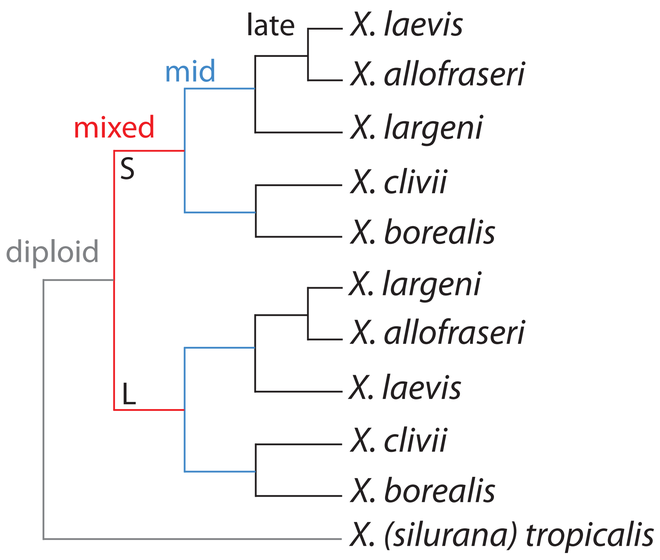

Figure 1: We constructed a starting topology using the 1417 genes (concatenated) (topology is similar to previous estimates; (*4*)). For the following analyses, the tree was divided into time periods pre and post duplication. The mixed lineages had periods of both diploidy and tetraploidy, as the moment of WGD is unable to be estimated.

Relaxation of Selection

The rate at which nonsynonymous (i.e. amino acid changing, called dN) and synonymous (i.e. non-amino acid changing, called dS) mutations are accumulating reflects the overall strength of purifying selection. Typically this rate is near zero, as nonsynonymous changes are often deleterious and removed. But, relaxation of selective pressures (since gene copies are redundant) allows nonsynonymous mutations to accumulate, bringing the ratio closer to one. We estimated this ratio for each sub-genome for the different time periods outlined in Fig. 1.

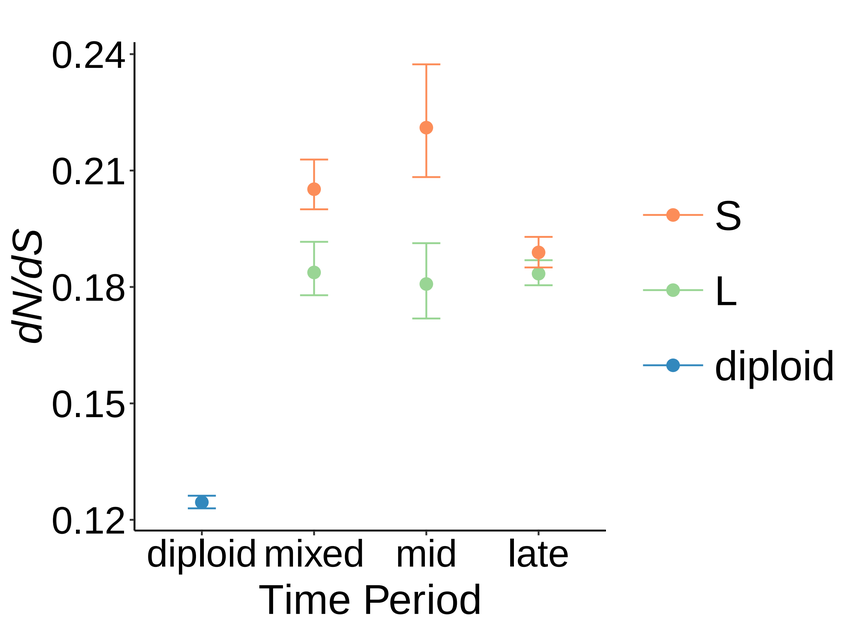

Figure 2: dN/dS estimates based on a concatenated alignment of 1417 genes with assigned sub-genome of origin, sampled in five species (14–64% missing data for each species). S and L reflect the two sub-genomes of tetraploid *Xenopus*. Error bars reflect 95% confidence intervals derived from bootstrap replicates of the sequence alignment.

Result 1: Biased Sub-Genome Evolution

1. Purifying selection is weaker post-WGD. After ~35 my (*3*), the strength of purifying selection has not returned to diploid like levels.
2. Relaxation of selection was more pronounced soon after WGD
3. Surprisingly, the S-sub genome experience greater relaxation of selective pressures than the L, and this effect was limited to shortly after WGD.

Rates of Gene Loss

Model Methods

Using a modified method of (*5*), we used a maximum likelihood model to estimate the rate of gene loss (pseudogenization) for different time-points with presence/absence data of gene copies. Due to modeling constraints, we were only able to compare "mid" and "late" time-points. Missing gene copies for a species could be the result of pseudogenization or missed for technical reasons, and our model estimates both.

Model Results

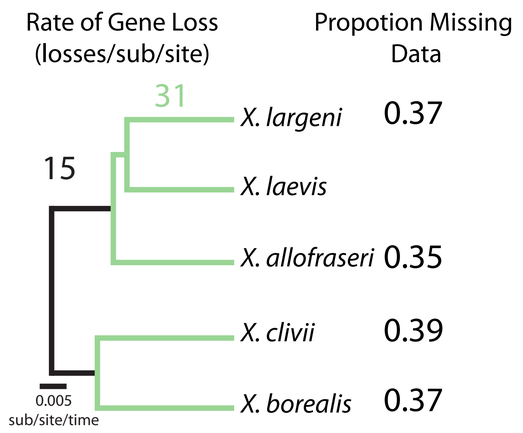

Figure 3: Values on branches reflect time specific estimates of the rate of gene loss. 95% confidence intervals do not overlap.

Result 2: Increasing Rate of Gene Loss

1. Despite maximal gene redundancy immediately after WGD, gene loss rates are not highest soon after WGD
2. Gene loss has significantly increased in recent lineages ("late") compared to "mid", likely related to the increased pressure of purifying selection (Fig. 2)

Other Results

1. Using divergence from 750 human orthologs, we estimate the divergence time of the sub-genomes to be 33 my, similar to other estimates using different data and calibrations (*3*).
2. Transcripts from the S sub-genome are significantly shorter than transcripts from the L. Possibly reflecting accumulation of premature stop codons.

Conclusion

Genome restructuring post-WGD is still ongoing in *Xenopus*. Selective pressures have not returned to pre-WGD levels, which will lead to continued gene loss and potentially divergence of gene duplicates. The biased sub-genome evolution is surprising and it is uncertain what is driving this difference (*3*), but our analyses reveal that this bias occurred soon after WGD and may be resolved in extant lineages.

The rate of gene loss had increased over time, contrary to most predictions. Recently, a new model of evolution post-WGD was proposed where at first the two gene copies are maintained for a constant total expression level (*2*). As mutations accumulate, due to relaxed purifying selection, the expression levels diverge (one going higher and the other lower), selection pressures change for each copy, and reach an extreme with silencing of one copy. As more mutations accumulate, more genes will be lost. Work in *Paramecium* with WGD support this model (*2*) and it may be the case for *Xenopus* as well, suggesting that increasing rates of gene loss may be a common outcome of WGD events.

References

1. M. Lynch, J. S. Conery, *Science* **290**, 1151–1155 (2000).
2. J. Gout, M. Lynch, *Molecular biology and evolution*, msv095 (2015).
3. A. M. Session, Y. Uno, T. Kwon, J. A. Chapman, A. Toyoda, S. Takahashi, A. Fukui, A. Hikosaka, A. Suzuki, M. Kondo, *et al.*, *Nature* **538**, 336–343 (2016).
4. B. L. S. Furman, B. J. Evans, *G3: Genes| Genomes| Genetics* **6**, 3625–3633 (2016).
5. U. J. Dang, A. M. Devault, T. D. Mortimer, C. S. Pepperell, H. N. Poinar, G. B. Golding, *Genetics* **204**, 513–529 (2016).

Acknowledgements

Thanks to NSERC for project and personal funding.