Published online: 11 June 2018
https://doi.org/10.1038/s41588-018-0144-6

References

1. Yu, J. et al. *Nat. Genet.* **38**, 203–208 (2006).
2. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. *Nat. Genet.* **46**, 100–106 (2014).
3. Loh, P.-R. et al. *Nat. Genet.* **47**, 284–290 (2015).
4. Sudlow, C. et al. *PLoS Med.* **12**, 1–10 (2015).
5. Bycroft, C. et al. Preprint at *bioRxiv* https://www.biorxiv.org/content/early/2017/07/20/166298 (2017).
6. Listgarten, J. et al. *Nat. Methods* **9**, 525–526 (2012).
7. Bulik-Sullivan, B. K. et al. *Nat. Genet.* **47**, 291–295 (2015).
8. Gazal, S. et al. *Nat. Genet.* **49**, 1421–1427 (2017).
9. Canela-Xandri, O., Rawlik, K. & Tenesa, A. Preprint at *bioRxiv* https://www.biorxiv.org/content/early/2017/08/16/176834 (2017).
10. Zhou, W. et al. Preprint at *bioRxiv* https://www.biorxiv.org/content/early/2017/11/15/212357 (2017).

Acknowledgements

We are grateful to H. Finucane and Y. Reshef for helpful discussions. This research was conducted using the UK Biobank Resource under application 10438 and was supported by US National Institutes of Health grants R01 HG006399, R01 GM105857 and R01 MH107649 (A.L.P.), a Burroughs Wellcome Fund Career Award at the Scientific Interfaces and the Next Generation Fund at the Broad Institute of MIT and Harvard (P.-R.L.), and a Boehringer Ingelheim Fonds fellowship (A.P.S.). Computational analyses were performed on the Orchestra High-Performance Compute Cluster at Harvard Medical School, which is partially supported by grant NCRR 1S10RR028832-01.

Author contributions

P.-R.L. and A.L.P. designed the study. P.-R.L., G.K., S.G. and A.P.S. performed analyses. All authors wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0144-6.

Subfunctionalization versus neofunctionalization after whole-genome duplication

To the Editor: The question of what the predominant evolutionary fate is of genes after duplication events has been intensely debated for decades[1,2]. Two articles in *Nature* (Lien et al.[3]) and *Nature Genetics* (Braasch et al.[4]) investigated the regulatory fate of gene duplicates after the salmonid-specific (Ss4R) and teleost-specific (Ts3R) whole-genome duplication (WGD) events, respectively.

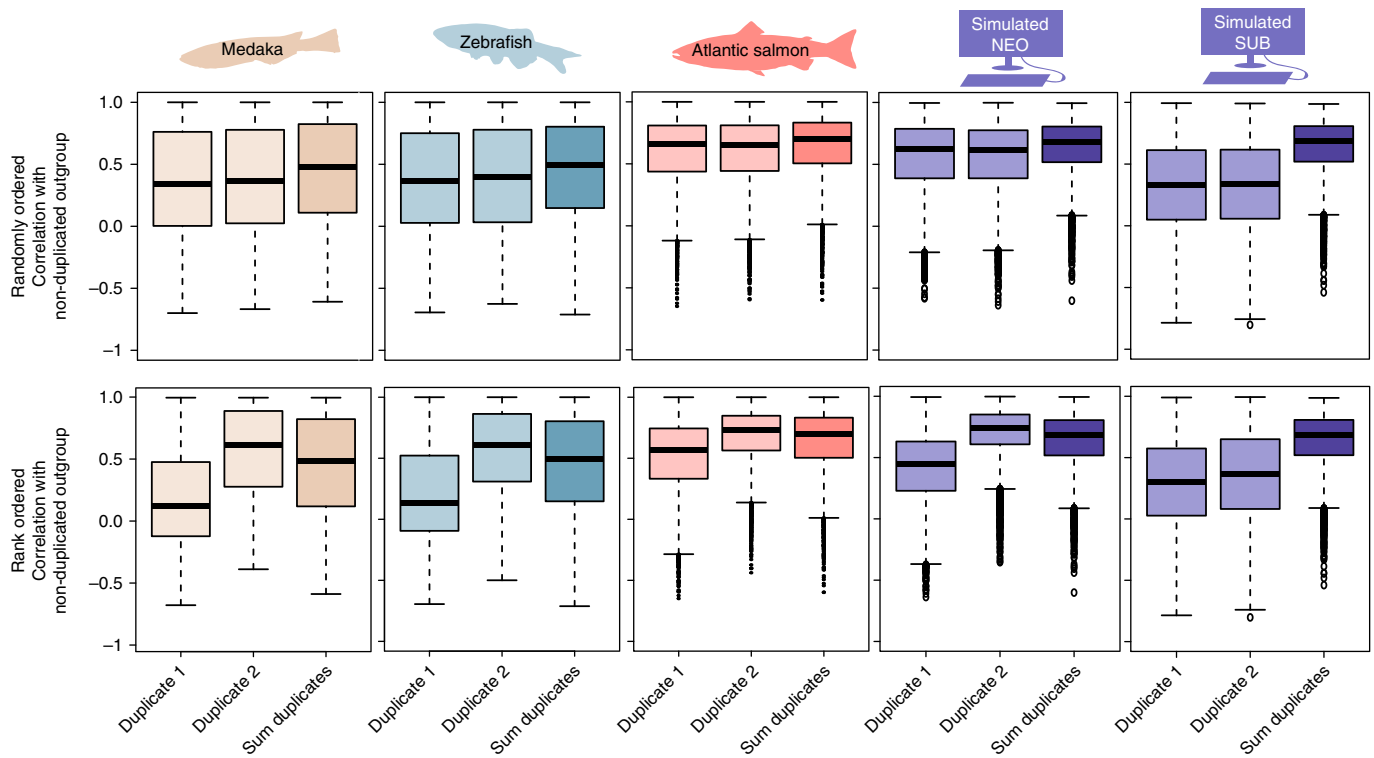

Fig. 1 | Tissue expression divergence in real and simulated data. Tissue expression correlation between duplicates in medaka or zebrafish and the corresponding orthologs in spotted gar (1,606 and 1,315 triplets, respectively) and between duplicates in Atlantic salmon and the orthologs in Northern pike (8,070 triplets). In the upper row, duplicated genes are assigned labels 'duplicate 1' and 'duplicate 2' randomly, while in the lower row the duplicates are ranked so that duplicate 1 has the lowest correlation with the ortholog and duplicate 2 has the highest. 'Sum duplicates' represents the correlation between the summed expression of the two duplicates and the ortholog in the unduplicated species. All correlations were computed using the Pearson correlation coefficient on the original expression data from the two publications in the first three columns and simulated data in the two last columns. All pairwise comparisons were statistically significant ($P < 5 \times 10^{-11}$, Wilcoxon signed-rank test, two-sided), with the exception of comparisons between duplicate 1 and duplicate 2 in the upper row (randomly ordered). Box plots were produced using the function 'boxplot' in R with default settings. The boxes indicate upper and lower quartiles with the horizontal lines marking the medians. The lines extending vertically from the boxes (whiskers) indicate the maximum and minimum values excluding outliers. Outliers are plotted as open circles. The expression data are available in the Supplementary Data.

Both studies relied on tissue expression atlases to estimate regulatory divergence and used closely related unduplicated sister taxa (Northern pike and spotted gar, respectively) as proxies for the ancestral expression state. Surprisingly, the two studies reach very different conclusions about the evolutionary mechanisms impacting gene expression after WGD. Braasch et al.[4] concluded that expression divergence was consistent with partitioning of tissue regulation between duplicates (subfunctionalization), whereas Lien et al.[3] concluded that most divergences in tissue regulation were consistent with one copy maintaining ancestral tissue regulation while the other diverged (in line with neofunctionalization). Here we show that this discrepancy in the conclusions of the two studies is a consequence of the analytical approach and is not related to underlying differences in the data.

To evaluate the underlying cause of the discrepancies between the two studies, we reanalyzed the data from Braasch et al.[4] using the approach of Lien et al.[3], and vice versa. Both studies computed expression correlation to the unduplicated sister taxon as a measure of divergence—for the two duplicates individually (duplicate 1 and duplicate 2) and for the summed expression of the duplicates (sum duplicates). The only aspect of the analyses that differed was how the individual genes within the duplicate pairs were ranked. In Braasch et al.[4], the duplicate pairs were ordered randomly (i.e., the two genes were assigned labels 'duplicate 1' and 'duplicate 2' at random; see Chapter 13.1 (p. 36) of the supplementary material of Braasch et al.[4]). In contrast, Lien et al.[3] ranked the two genes in each duplicate pair as 'most diverged' (i.e., the gene with the lowest expression correlation with the ortholog in the unduplicated sister taxon; duplicate 1) and 'most conserved' (i.e., the gene with the highest expression correlation with the ortholog; duplicate 2).

On the basis of the three distributions of expression correlations between the ortholog in the unduplicated sister taxon (the assumed ancestral state) and duplicate 1, duplicate 2 and the sum of the duplicates in the duplicated species, conclusions can be drawn about broad transcriptome-wide evolutionary trends. If subfunctionalization is the dominant driver of expression divergence, we expect the sum of the duplicates to correlate more strongly with the ancestral state than the expression of the individual gene duplicates. Conversely, if neofunctionalization is the major evolutionary mechanism in play, we expect the duplicate with the most conserved expression patterns (duplicate 2 in the method ranking duplicates) to exhibit the highest expression correlation to the unduplicated ortholog.

Our reanalyses of the two datasets show that the conclusions from the two papers would be identical if the same data analysis approach were used (Fig. 1). So, do the results of Braasch et al.[4] and Lien et al.[3] support sub- or neofunctionalization as the dominant mechanism?

To answer this question, we simulated neo- and subfunctionalized duplicates and applied both the randomly ordered and rank-ordered methods. For the simulations, we used the pike transcriptome as an unduplicated reference. For the duplicates, we simulated neofunctionalization by adding a small error term ~N(0,1) for the conserved duplicate and a larger error term ~N(0,5) for the diverged duplicate. Subfunctionalization was simulated by randomly partitioning the preduplication tissue expression across the two duplicates and then adding a small error term, ~N(0,1). While the ranked approach correctly identified patterns consistent with sub- and neofunctionalization in the simulated data (Fig. 1, bottom row), random ordering of the duplicates obscured the signal of neofunctionalization, resulting in both simulations exhibiting patterns consistent with subfunctionalization (Fig. 1, top row).

From this analysis, we conclude that the prevailing fate for gene duplicates from the Ts3R and Ss4R WGD events is that one duplicate is under stronger purifying selection pressure to maintain ancestral regulation than the other duplicate. This is more consistent with regulatory neofunctionalization than with subfunctionalization. Nevertheless, the observed asymmetric gene expression divergence among duplicates could be a result of relaxed purifying selection (neutral evolution) rather than adaptive selection for new regulation (neofunctionalization). The crux of this can be addressed in future studies by comparing the likelihood of observed duplicate expression divergence data across multiple species under a model of regulatory neofunctionalization and under a model of neutral evolution[5]. Such a test would improve power when including more informative species and would provide stronger evidence for neofunctionalization if duplicate divergence is conserved across several species with the same WGD. Finally, by only evaluating global patterns of expression evolution, we neglect the fact that selection acts differently on different individual genes. It is therefore of the utmost importance to adapt existing phylogenetic methods[5–8] and to develop new techniques to evaluate whether the data support sub- or neofunctionalization on a gene-by-gene basis. Now that genome sequences and associated functional data are accumulating, this will open up new and exciting avenues for answering long-standing questions regarding genome evolution following WGD.

Data availability

All data analyzed during this study are included in the Supplementary Data. ❐

Simen R. Sandve[1]*, Rori V. Rohlfs[2] and Torgeir R. Hvidsten[3,4]*

[1]Center for Integrative Genetics (CIGENE), Faculty of Biosciences, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, Norway. [2]Department of Biology, San Francisco State University, San Francisco, CA, USA. [3]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway. [4]Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Umeå, Sweden.

**e-mail: simen.sandve@nmbu.no; torgeir.r.hvidsten@nmbu.no*

Published online: 28 June 2018
https://doi.org/10.1038/s41588-018-0162-4

References

1. Lynch, M. & Conery, J. S. *Science* **290**, 1151–1155 (2000).
2. Ohno, S. *Evolution by Gene Duplication* (Springer, Berlin, 1970).
3. Lien, S. et al. *Nature* **533**, 200–205 (2016).
4. Braasch, I. et al. *Nat. Genet.* **48**, 427–437 (2016).
5. Rohlfs, R. V. & Nielsen, R. *Syst. Biol.* **64**, 695–708 (2015).
6. Felsenstein, J. *Am. Nat.* **125**, 1–15 (1985).
7. Martins, E. P. & Hansen, T. F. *Am. Nat.* **149**, 646–667 (1997).
8. Revell, L. J. *Methods Ecol. Evol.* **3**, 217–223 (2012).

Acknowledgements

S.R.S. was supported by Norwegian Research Council projects 221734 and 244164.

Author contributions

S.R.S. and T.R.H. conceived the study. T.R.H. analyzed the expression data, and R.V.R. performed the simulations. All authors interpreted the results and wrote and approved the manuscript for publication.

Competing interests

The authors declare no competing interests.
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0162-4.