

COMMENTARY

Open Access

How the vertebrates were made: selective pruning of a double-duplicated genome

Gerard Manning* and Eric Scheeff

See research article: <http://www.biomedcentral.com/1741-7007/8/146/abstract>

Abstract

Vertebrates are the result of an ancient double duplication of the genome. A new study published in *BMC Biology* explores the selective retention of genes after this event, finding an extensive enrichment of signaling proteins and transcription factors. Analysis of their expression patterns, interactions and subsequent history reflect the forces that drove their evolution, and with it the evolution of vertebrate complexity.

A doubling of the genome, or whole genome duplication (WGD), is usually a cataclysmic event for an organism. Yet this polyploidy has been an important, if rare, event in the evolution of many plant groups, and has also occurred in yeasts, ciliates, fish and frogs [1]. It is now generally accepted that we and all other jawed vertebrates are the product of a remarkable two rounds of WGD, known as 2R [2], which duplicated every gene up to four-fold (fish and frog genomes have undergone a third duplication more recently). This opened the door to a tremendous expansion in functionality, and while most WGD duplicates, or ohnologs, were rapidly lost, this phenomenon was the genesis of almost one-third of all human genes. Establishing why these duplicates were retained and how they have evolved since then is an important way to advance the understanding of their current functions.

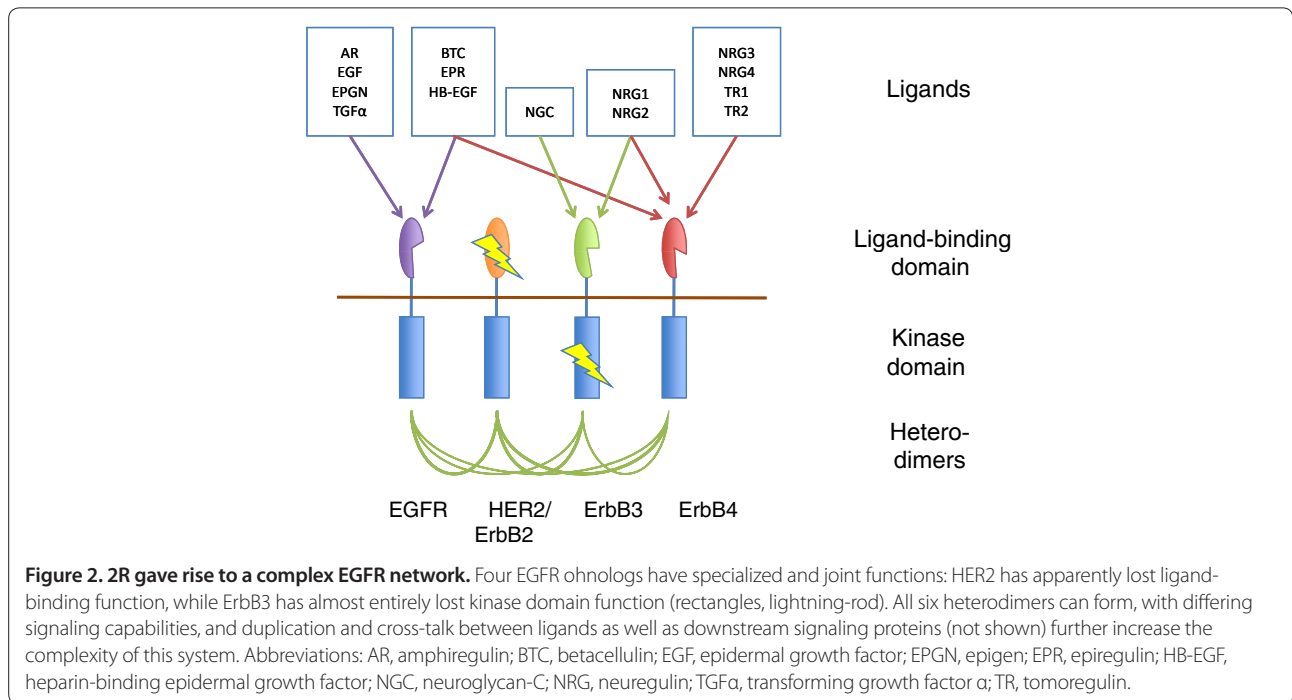
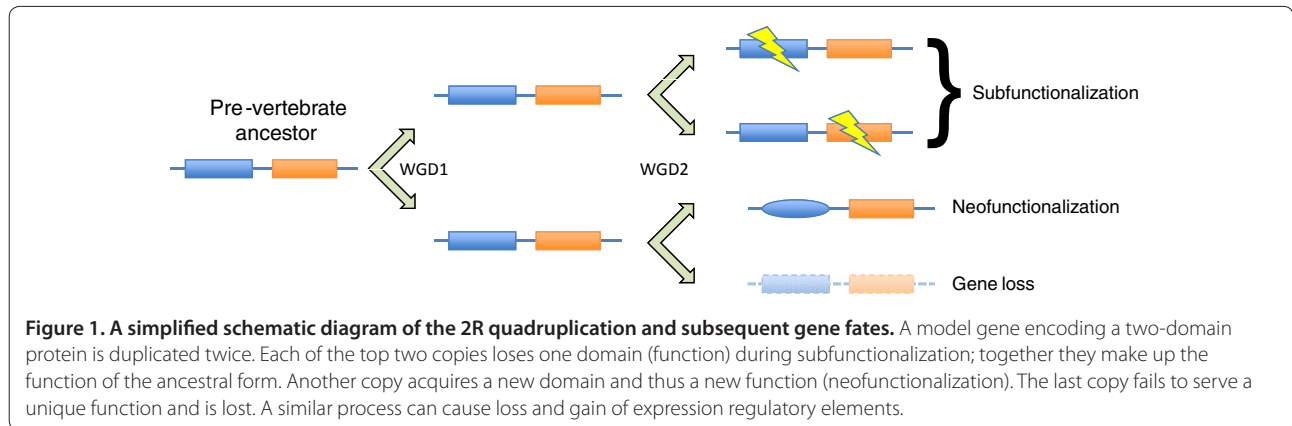
A study by Huminiecki and Heldin in *BMC Biology* [3] seeks to answer these questions through a global analysis of genes that survived the massive pruning that followed 2R. They identified 2R-derived gene pairs using a combination of sequence similarity (by comparing gene trees with the underlying species trees to identify duplications [4]) and chromosomal location, using syntenic

chromosomal regions, in which runs of related gene pairs occur in different loci. They then explore the history of most vertebrate genes through 2R and subsequent gains and losses. They find that retained ohnologs are highly biased towards signaling genes and transcription factors and argue that this large pool of new genes would have enabled the complex regulation required for the development and function of the vertebrate body plan. They integrate these results with expression and pathway data to show that retained ohnologs play important roles in functional categories, such as those required by the nervous system and for locomotion, that are crucial to complex vertebrates.

After WGD, each new ohnolog enters a race to develop an essential function before succumbing to deletion [1]. Some develop variant new functions (neofunctionalization), while other ohnolog pairs reciprocally lose some of their functions or expression pattern (subfunctionalization) (Figure 1). Others survive through gene dosage balance, in which the toxicity of having a double dose of one gene can be offset by the retention of a duplicate of an interacting gene [5]. The relative role of these and more esoteric mechanisms is debated. Multiple mechanisms may act on individual genes: for example, dosage balance may buy time for novel functions to evolve.

The importance of dosage balance is supported by two other findings from this paper. First, small scale duplications (SSDs) that have occurred after 2R show a very different functional bias from that of WGD duplicates: they contain far fewer signaling proteins and transcription factors, but are enriched in immune functions and chromatin modifiers. This suggests that individual duplication of signaling proteins may be toxic or non-functional, requiring the dosage balance of a WGD to survive. A similar bias is also seen in other studies of SSD following WGD, and ohnologs are also underrepresented in copy number variations in human populations, further reflecting their dosage sensitivity [5]. Second, they show that retained ohnologs are more

*Correspondence: manning@salk.edu
Razavi Newman Center for Bioinformatics, Salk Institute for Biological Studies,
10010 North Torrey Pines Road, La Jolla, CA 92037, USA



highly connected in pathway and protein interaction maps, further suggesting that they may be required for dosage balance.

The simplest gene dosage models are based on stoichiometric balance between subunits of a stable protein complex. The Humniecki and Heldin study highlights the limitations of the simple model, since signaling proteins and transcriptional regulators tend to make relatively transient interactions, consistent with their role in information transfer. This suggests that dynamic balancing of signal flux may be as important as structural balances in protein complexes. For instance, duplication of a phosphatase might balance the increased flux from duplication of a corresponding kinase; accordingly, retained ohnologs are specifically enriched for negative regulatory interactions [2]. Dosage balance may also

operate in a positive sense: rather than blocking toxicity, the co-duplication of many interacting genes may aid the development of novel pathways and functions.

Duplicates as an innovation factory

While dosage balance may explain the initial selective retention of WGD duplicates, development of new functions or expression patterns is the norm in most well studied human gene families. Humniecki and Heldin observe a divergence of mRNA and protein expression patterns between duplicate genes generated by either WGD or SSD, correlated with age of duplication. In signaling proteins, divergent functions are also common. For instance, all four ohnologs of the epidermal growth factor receptor (EGFR) tyrosine kinase were retained after 2R, giving rise to a complex array of homo- and

hetero-dimeric receptors [6] (Figure 2). Subfunctionalization is evident in the almost total loss of catalytic activity in ErbB3 and the apparent loss of ligand binding in ErbB2/HER2, while the concurrent duplications of ligands and downstream signaling genes has further expanded the complexity of this signaling system. EGFRs have proven refractory to SSD in metazoans, and indeed, amplification of the HER2 locus is a major driver for breast cancer, with other EGFR amplicons also reported to be associated with cancer, suggesting that negative selection may be operating on duplication of at least some family members.

Another receptor tyrosine kinase (RTK) family, the Ephs, has expanded by WGD and SSD from one gene in invertebrates to 14 in human, giving rise to a similar explosion in complexity through heterodimerization and ligand cross-talk. This richness is used extensively in developmental patterning, and demonstrates continued evolvability. For instance, in chicken, graded expression of EphA3 across the retina provides the basis for spatial mapping of retinal ganglion cells projecting to the tectum [7]. However, in mouse, EphA3 is not expressed in these cells, and instead EphA5 and EphA6 fulfill this role, suggesting that new and swapped functions can emerge from duplicates long after they have acquired essential roles, and that WGD can represent a quantum leap in the potential for new complexity and evolvability within the vertebrates. We estimate that, excluding the Ephs, 2R caused an expansion of RTKs from 20 to 46 but only two new human RTKs have emerged since then (ES and GM, unpublished): the two rounds of WGD thus seem to have been crucially important in shaping human RTK signaling.

One notable aspect of the patterns reported by Huminiecki and Heldin is how similar they are to those seen in other WGD events [8-10]. Enrichment in signaling proteins and transcription factors has also been seen in WGD from yeast, plants, and fish. Conversely, other genes (mostly those involved in basic cellular processes) preferentially return to singleton status, and similarities in these loss patterns can also be detected across kingdoms. While SSDs show more lineage-specific variability, there are also similarities, such as the increased SSD rate in plant secondary metabolic genes involved in pathogen defense [8] mimicking the increased vertebrate SSD in immune genes.

2R: the future

It is tempting to speculate from these observations that WGD produces a consistent drive towards higher complexity [11], and the two rounds of vertebrate WGD

doubly so. However, it is a vexed question exactly what is meant by complexity. It is not clear, for example, that fish and frogs, which have undergone an extra round of genome duplication, are more complex than humans, which have not.

The kind of molecular archaeology pursued by Huminiecki and Heldin is not just of academic interest: detailed comparison of ohnologs from many species can provide the unique sequence signatures underlying their specific functions, and patterns of gain or loss can help us to understand functional interactions between genes. As more vertebrate genomes become available, we will gain greater precision in determining orthology, synteny and post-2R changes. Knowing the trends in ohnolog retention and the history of human genes will help us to better understand their dosage sensitivity, and the shared and unique functions of all ohnologs.

Published: 13 December 2010

References

1. Sémon M, Wolfe KH: Consequences of genome duplication. *Curr Opin Genet Dev* 2007, **17**:505-512.
2. Kasahara M: The 2R hypothesis: an update. *Curr Opin Immunol* 2007, **19**:547-552.
3. Huminiecki L, Heldin C-H: 2R and modeling of vertebrate signal transduction engine. *BMC Biol* 2010, **8**:146
4. Li H, Coghlan A, Ruan J, Coin LJ, Hériché JK, Osmotherly L, Li R, Liu T, Zhang Z, Bolund L, Wong GK, Zheng W, Dehal P, Wang J, Durbin R: TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res* 2006, **34**(Database issue):D572-580.
5. Makino T, McLysaght A: Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Natl Acad Sci U S A* 2010, **107**:9270-9274.
6. Bublil EM, Yarden Y: The EGF receptor family: spearheading a merger of signaling and therapeutics. *Curr Opin Cell Biol* 2007, **19**:124-134.
7. Lemke G, Reber M: Retinotectal mapping: new insights from molecular genetics. *Annu Rev Cell Dev Biol* 2005, **21**:551-580.
8. Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y: The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol* 2006, **7**:R43.
9. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y: Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A* 2005, **102**:5454-5459.
10. Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC: Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. *Trends Genet* 2006, **22**:597-602.
11. Freeling M, Thomas BC: Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res* 2006, **16**:805-814.

doi:10.1186/1741-7007-8-144

Cite this article as: Manning G, Scheeff E: How the vertebrates were made: selective pruning of a double-duplicated genome. *BMC Biology* 2010, **8**:144.