

Minireview

***In ovo omnia*: diversification by duplication in fish and other vertebrates**

Ingo Braasch* and Walter Salzburger†

Addresses: *University of Würzburg, Physiological Chemistry I, Biocenter, Am Hubland, 97074 Würzburg, Germany. †Zoological Institute, University of Basel, Vesalgasse 1, 4055 Basel, Switzerland.

Correspondence: Walter Salzburger. Email: walter.salzburger@unibas.ch

Published: 5 March 2009

Journal of Biology 2009, **8**:25 (doi:10.1186/jbiol121)

The electronic version of this article is the complete one and can be found online at <http://jbiol.com/content/8/3/25>

© 2009 BioMed Central Ltd

Abstract

Gene and genome duplications are considered to be the main evolutionary mechanisms contributing to the unrivalled biodiversity of bony fish. New studies of vitellogenin yolk proteins, including a report in *BMC Evolutionary Biology*, reveal that the genes underlying key evolutionary innovations and adaptations have undergone complex patterns of duplication and functional evolution.

Since the publication of Charles Darwin's *The Origin of Species* a century and a half ago, evolutionary biologists have been concerned with the identification of the processes that govern the emergence of new species and, thus, of organismal diversity. Because of variation in the rate of speciation and extinction, evolution inevitably leads to an unequal distribution of morphological diversity and species-richness across taxonomic lineages. Some lineages have remained morphologically uniform and are species-poor, whereas others have diversified rapidly. It is these more 'successful' and species-rich lineages in particular that enable insights into the process of diversification.

In vertebrates, the most species-rich group is that of the fishes: at least one in two vertebrate species is a fish, or - more precisely - a teleost fish. There are at least 26,000 living teleost species [1], which show a remarkable variety of ecological, morphological and behavioral adaptations. Among the characteristics that distinguish the teleost cohort from the only 50 or so species of basal ray-finned fishes and the rest of the vertebrates are genomic features such as gene

and genome duplications and higher rates of chromosomal rearrangements and molecular evolution [2].

Are gene and genome duplications the fuel that drives biodiversity in fish?

A fish-specific genome duplication (also known as the 3R duplication) occurred in an ancestor of the teleost lineage around 300-350 million years ago [3]. This event, which endowed teleosts with additional new genes, has been hypothesized to be at least partly responsible for their biodiversity and species richness [2,4,5]. Not all genes that emerged from the duplication are still present, however. In fact, the majority of duplicated genes (about 70-90%) have since been degraded and/or lost (a process termed nonfunctionalization). But because this massive post-duplication gene loss followed different routes, different teleost lineages now have different complements of paralogous genes derived from the original genome duplication. This process is called divergent resolution [4,5]. Empirical support for divergent resolution between teleost lineages that diverged

very early comes from a recent comparative genome-wide analysis of paralog loss in zebrafish and the green spotted pufferfish [6].

In many cases where both copies have been maintained in a genome, the functions of the ancestral gene are now distributed among the duplicates - a process called subfunctionalization. Given that retention of duplication-derived gene copies also followed different routes and that subfunctionalization can be neutral and stochastic, the partitioning of gene functions can also occur lineage-specifically. Finally, it is possible that one of the duplicates continues to fulfill the ancestral functions while the other acquires a completely new function (neofunctionalization). Differential functional evolution between teleost lineages has so far been shown for zebrafish, stickleback and medaka [4].

Together, the fish-specific genome duplication and the divergent resolution, subfunctionalization and neofunctionalization that followed it created a large evolutionary playground within teleost genomes. The duplication-diversification hypothesis predicts that gene and genome duplication and subsequent reciprocal gene loss and/or differential paralog evolution in divergent populations leads to genomic incompatibilities between isolated populations and, consequently, to postzygotic isolation and speciation. That is how the fish-specific genome duplication might have facilitated the radiation of teleosts [4,5].

Vitellogenin gene duplications and marine teleost radiations

Besides the overall impact of gene and genome duplication on reproductive isolation and thus on speciation, neofunctionalization of a duplicated gene copy can lead to the origination of a key evolutionary innovation that enables a group to radiate, for example in a new environment. In two new articles, one in *BMC Evolutionary Biology* [7] and the other in *Molecular Biology and Evolution* [8], Finn and colleagues examine an example of a cluster of genes that emerged by duplication and that apparently has enabled a whole group of fishes to diversify.

Finn and Kristoffersen had already in earlier studies [1] reconstructed the evolution of the vitellogenin (*vtg*) gene family in teleost fishes. Vitellogenins are yolk proteins synthesized in the liver and deposited in the maturing oocyte. Finn and Kristoffersen [1] suggested that neofunctionalization of the *vtgAa* gene in acanthomorphs, the most species-rich group of teleosts (comprising about 16,000 species, 78% of which are marine), was an important step towards adapting to a new spawning strategy in the marine realm. Proteolysis of the VtgAa yolk protein leads to

an increase in the levels of free amino acids in the maturing oocyte and causes water influx. In this way, the hydrated eggs are protected against leakage of water into the hyperosmolar marine environment, so that the eggs float on the water surface. This is an important adaptation that makes pelagic ('floating') spawning strategies possible.

The initial phylogenetic analysis of teleost vitellogenins [1] suggested that the three *vtg* genes in acanthomorphs, *vtgAa*, *vtgAb* and *vtgC*, evolved through a progressive series of gene duplications and subsequent gene losses, involving the fish-specific genome duplication and the two earlier rounds of whole genome duplication in vertebrates (called 1R and 2R), and also an acanthomorph-specific duplication of the *vtgA* gene that generated the *vtgAa* and *vtgAb* duplicates. According to this scenario, lineage-specific neofunctionalization of the newly arising *vtgAa* paralog in acanthomorphs facilitated their conquest of the marine ecosystem from their original habitats in freshwaters.

New data presented by the same group in *BMC Evolutionary Biology* [7], as well as an earlier article by Babin [9], take the location of vitellogenin genes in vertebrate genomes into account and turn the duplication history of teleost *vtg* genes upside down. In acanthomorphs, *vtgAa*, *vtgAb*, and *vtgC* are located close to each other on the same chromosome. This is consistent with the arrangement of vitellogenin genes in other teleosts and in more distantly related vertebrate lineages, such as frog and chicken [7,9]. The most parsimonious explanation for this arrangement is thus that a vitellogenin gene cluster consisting of three genes (*Vtg1*, called *vtgC* in fish, *Vtg2*, called *vtgAb* in fish, and *Vtg3*, called *vtgAa* in fish) was already present in the last common ancestor of fish and tetrapods about 450 million years ago (Figure 1). An ancestral vitellogenin gene (proto *Vtg*) was duplicated, giving rise to *Vtg1* and *Vtg2/3*. The latter gene was then duplicated in tandem, generating *Vtg2* and *Vtg3* (Figure 1a). In the fish lineage, two vitellogenin gene clusters were present after the fish-specific genome duplication, but one of them degenerated so that this round of genome duplication did not increase the number of functional *vtg* genes.

In theory, phylogenetic reconstruction of the vitellogenin gene or protein family should reveal these three ancestral gene duplications. However, published vitellogenin phylogenies [1,7,8,10] consistently suggest that the different vertebrate *Vtg2* and *Vtg3* genes have been generated in parallel but independently through lineage-specific tandem duplications (Figure 1b). One explanation for the failure of phylogenies to reconstruct the common duplication of the *Vtg2/3* precursor could be that gene conversion has occurred between *Vtg2* and *Vtg3*, keeping them alike. The new results

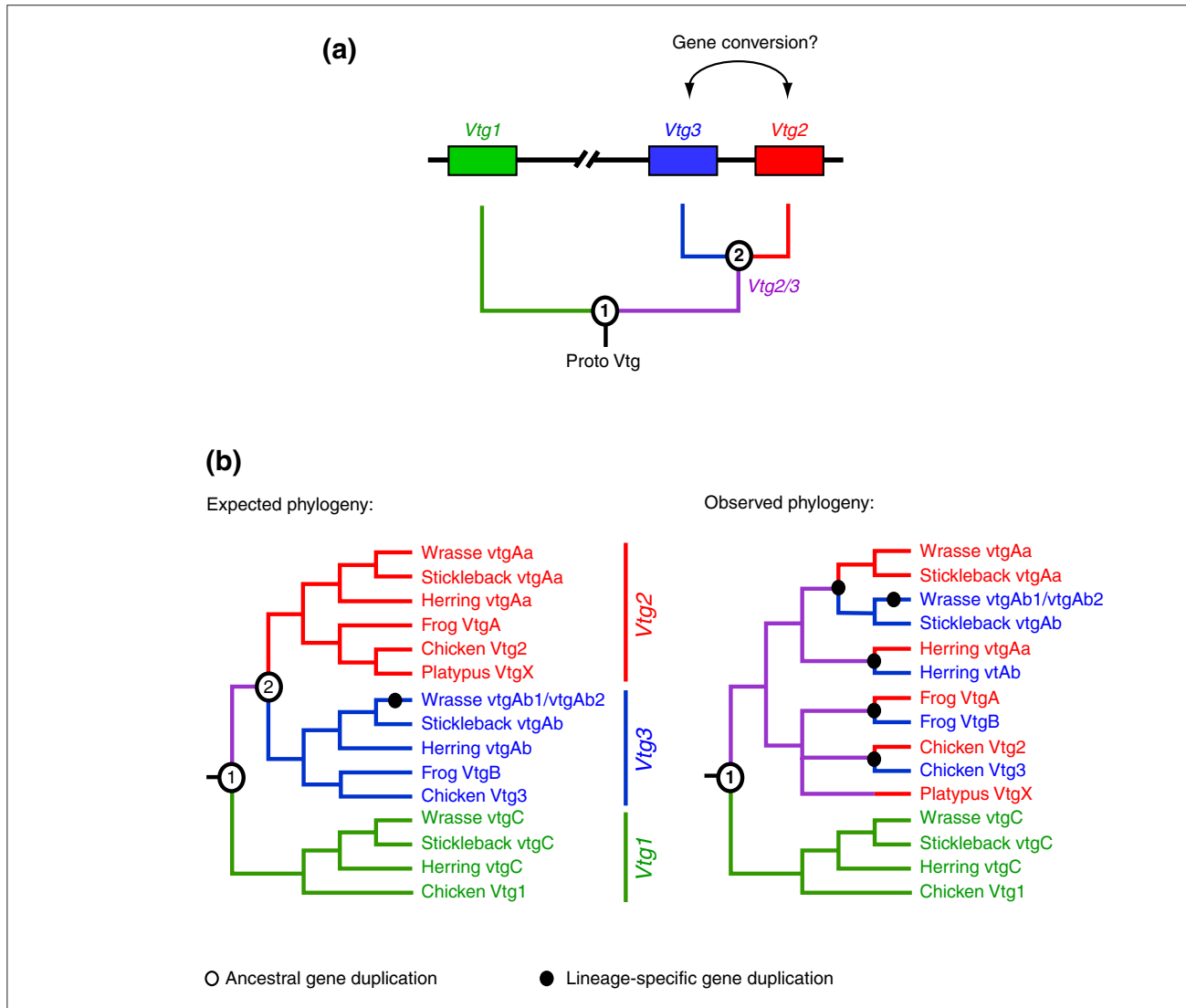


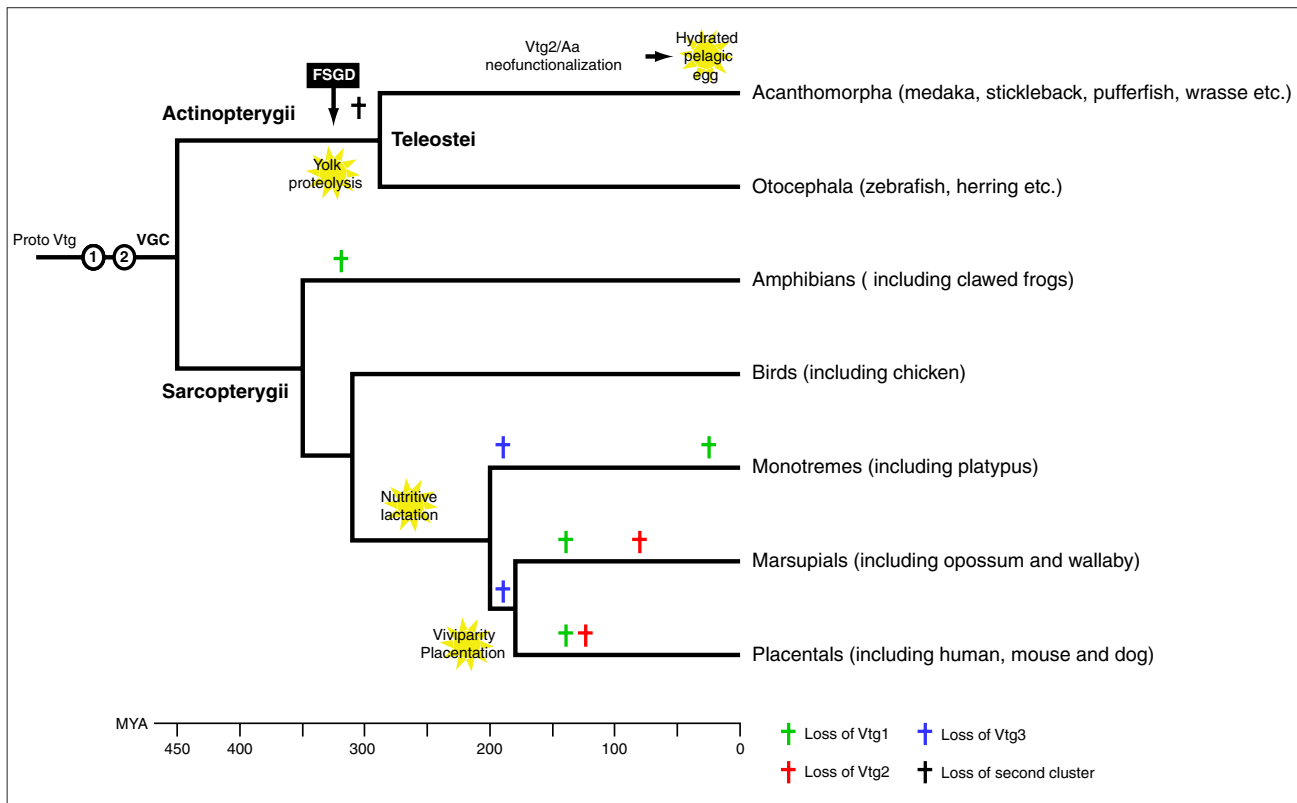
Figure 1 Evolution of the vertebrate vitellogenin cluster. **(a)** The vertebrate vitellogenin cluster was generated by two ancestral gene duplications (1 and 2). **(b)** The phylogeny of vertebrate Vtgs should reconstruct the ancestral gene duplications correctly (left), but observed phylogenies (right, merged and deduced from [1,7,8,10]) indicate multiple, independent duplications (black circles) of *Vtg2/3*. Gene names are as used in the literature. A unifying nomenclature is shown to the right of the expected phylogeny. The remaining functional platypus *VtgX* gene is most likely a *Vtg2* [9,10].

by Finn *et al.* [7] and Babin [9] therefore illustrate how important it is to include synteny data for the correct inference of gene family evolution.

Use it or lose it (or duplicate or delete)

The evolutionary significance of vitellogenins is further substantiated by the high frequency of true lineage-specific duplication events in teleost fishes. In acanthomorphs, *Vtg2/vtgAa* has been duplicated in medaka, whereas

Vtg3/vtgAb has multiple copies in marine labrids (wrasses). In the zebrafish, an ostariophysian, both *Vtg3/vtgAb* and *Vtg2/vtgAa* have been duplicated, the latter being present in as many as five copies [7-9]. Nevertheless, acanthomorphs are special in their processing of the *Vtg2/VtgAa* protein and the exceptionally high expression of *Vtg2/vtgAa* in marine, pelagically spawning species [7]. Although yolk proteolysis evolved before the divergence of Acanthomorpha and Otocephala (such as zebrafish and herring), it was not until the neofunctionalization of *Vtg2/vtgAa* in the acanthomorph

**Figure 2**

Evolution of reproductive modes and vitellogenins in bony vertebrates. White circles indicate the ancestral gene duplications (1 and 2) that led to the establishment of the vitellogenin cluster (VGC). Yellow stars indicate innovations in the reproductive mode; crosses indicate *Vtg* gene losses. FSGD, fish-specific genome duplication; MYA, million years ago. The timing of establishment of the vitellogenin cluster in relation to the emergence of vertebrates and the occurrence of the 1R/2R genome duplications remain elusive and will require additional data from cartilaginous fishes, agnathans and non-vertebrate chordates. Adapted from [10] and revised and expanded using fish data from [7,8].

lineage that highly hydrated marine pelagic eggs were made possible, thereby triggering the teleost radiation in the oceans. This happened at least 400 million years after the evolution of the *Vtg2/vtgAa* gene itself [7,8].

In another part of the vertebrate phylogeny, some lineages evolved that do not seem to have any use for yolk proteins such as vitellogenins: mammals have evolved placentation and lactation to nourish their offspring [10]. It therefore does not come as a surprise that all three vitellogenin genes have been lost from the evolutionary lineage leading to the placental mammals and marsupials. Only the egg-laying monotremes have retained a single functional *Vtg* gene (Figure 2) [10]. The evolution of vitellogenins in vertebrates nicely demonstrates an association between gene duplication and functional need. It also shows that adaptively very important genes underlying key evolutionary innovations can lose their relevance once a new innovation arises, with the consequence that such

genes can vanish entirely from a genome. 'Use it or lose it' is the motto, or - in the context of genome evolution - duplicate it or delete it. An intriguing question remains: were there functional necessities of reproduction that were associated with the duplications of the vertebrate proto *Vtg* gene in the first place? The answer might, once more, be found in the oceans, where ancestral vertebrates used to spawn.

References

1. Finn RN, Kristoffersen BA: **Vertebrate vitellogenin gene duplication in relation to the "3R hypothesis": correlation to the pelagic egg and the oceanic radiation of teleosts.** *PLoS ONE* 2007, **2**:e169.
2. Ravi V, Venkatesh B: **Rapidly evolving fish genomes and teleost diversity.** *Curr Opin Genet Dev* 2008, **18**:544-550.
3. Meyer A, Van de Peer Y: **From 2R to 3R: evidence for a fish-specific genome duplication (FSGD).** *Bioessays* 2005, **27**:937-945.
4. Postlethwait J, Amores A, Cresko W, Singer A, Yan YL: **Subfunction partitioning, the teleost radiation and the annotation of the human genome.** *Trends Genet* 2004, **20**:481-490.

5. Volff JN: **Genome evolution and biodiversity in teleost fish.** *Heredity* 2005, **94**:280-294.
6. Semon M, Wolfe KH: **Reciprocal gene loss between *Tetraodon* and zebrafish after whole genome duplication in their ancestor.** *Trends Genet* 2007, **23**:108-112.
7. Finn RN, Kolarevic J, Kongshaug H, Nilsen F: **Evolution and differential expression of a vertebrate vitellogenin gene cluster.** *BMC Evol Biol* 2009, **9**:2.
8. Kristoffersen BA, Nerland A, Nilsen F, Kolarevic J, Finn RN: **Genomic and proteomic analyses reveal non-neofunctionalized vitellogenins in a basal clupeocephalan, the Atlantic herring, and point to the origin of maturational yolk proteolysis in marine teleosts.** *Mol Biol Evol* 2009, doi:10.1093/molbev/msp014.
9. Babin PJ: **Conservation of a vitellogenin gene cluster in oviparous vertebrates and identification of its traces in the platypus genome.** *Gene* 2008, **413**:76-82.
10. Brawand D, Wahli W, Kaessmann H: **Loss of egg yolk genes in mammals and the origin of lactation and placentation.** *PLoS Biol* 2008, **6**:e63.