Whole-Genome Duplication in Teleost Fishes and Its Evolutionary Consequences

Stella K. Glasauer[1] and Stephan C.F. Neuhauss[1]*

[1]University of Zurich, Institute of Molecular Life Sciences, Neuroscience Center Zurich and Molecular Life Science Program Zürich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

*Corresponding Author: email: stephan.neuhauss@imls.uzh.ch

Phone: +41 44 635 60 40

Fax: +41 44 635 68 97

Keywords:

gene pairs, paralogue, subfunctionalization, neofunctionalization evolution, radiation

Abstract

Whole genome duplication (WGD) events have shaped the history of many evolutionary lineages. One such duplication has been implicated in the evolution of teleost fishes, the by far most species-rich vertebrate clade. After initial controversy, evidence is now solid that such an event took place in the common ancestor of all extant teleosts, the so-called teleost-specific (TS) WGD. After WGD, duplicate genes have different fates. The most likely outcome is non-functionalization of one duplicate gene due to the lack of selective constraint on preserving both copies of the gene. Mechanisms that act on preservation of duplicates are subfunctionalization (partitioning of ancestral gene functions on the duplicates), neofunctionalization (assigning a novel function to one of the duplicates) and dosage-selection (preserving genes to maintain dosage-balance between interconnected components). Since the frequency of these mechanisms is influenced by the gene's properties, there are over-retained classes of genes, such as highly expressed ones and genes involved in neural function. The consequences of the TS-WGD, especially its impact on the massive radiation of teleosts, have been matter of controversial debate. It is evident that gene duplications are crucial for generating complexity and WGDs provide large amounts of raw material for evolutionary adaptation and innovation. However, it is less clear whether the TS-WGD is directly linked to their evolutionary success and radiation. Recent studies let us conclude that TS-WGD has been important in generating teleost complexity, but that more recent ecological adaptation only marginally related to TS-WGD might have played an even bigger role.

1. Introduction

During evolution, genes are often subject to duplication events. Duplications can affect single genes, a stretch of several genes, whole chromosomes or even whole genomes. Doubling of whole genomes initially leads to polyploidization (doubling of the whole chromosomal set) and can principally be achieved by nonreduction in meiosis or somatic doubling in mitosis, either in the parental germline or in the early embryo. However, on an evolutionary timescale polyploidy does not persist. Duplicated chromosomes accumulate changes until they become too different to pair as quadrivalents during meiosis. Eventually, when disomic inheritance of all chromosomes is restored, a fully diploid organism emerges. This process is called re-diploidization. While re-diploidized organisms are no longer polyploid, they still carry signs of the ancestral polyploidization event, such as genes that have been retained as duplicates.

Duplication of a gene results in two daughter genes, termed paralogues (resulting from a duplication event within the genome regardless of the mechanism they arose by). Immediately after duplication, paralogues are identical and functionally redundant. It was realized early on by Susumu Ohno that such redundant genes are attractive candidates to provide the genetic raw material for evolutionary innovation (Ohno, 1970a). By releasing genes from selective constraint in this way, one of the duplicates can be assigned a novel function, a process called neofunctionalization.

It has been suggested that whole genome duplications are especially important in generating novel genes, since during whole-genome duplication (WGD) the entire genetic repertoire of an organism is doubled.

Although the importance of polyploidization events has initially been realized in plants, it is now clear that also many animals experienced WGDs (Mable, 2004). Even within mammals, which are generally thought not to tolerate polyploidization, a duplicated genome was

identified in a rodent (Gallardo et al., 1999; Gallardo et al., 2004). There is now clear evidence that the radiation of vertebrates was preceded by two rounds of WGD, and it has been suggested that these events have contributed to diversification and evolutionary innovations within vertebrates (Canestro et al., 2013).

The subject of this review is to focus on a third round of WGD within vertebrates that occurred at the base of the teleost fish lineage, the so-called teleost-specific (TS) WGD. Teleosts comprise most extant bony fishes and are the by far most diverse vertebrate group.

First, we will summarize evidence for this WGD event. We will then discuss the fate that duplicate genes can undergo, especially in the context of genes that originated in the TS-WGD. Finally, we are asking which evolutionary consequences the TS-WGD had, focusing on its potential contribution to the massive radiation of teleost fishes.

2. WGDs have shaped teleost evolution

A WGD took place in the common ancestor of all teleosts

First evidences for TS-WGD emerged from the realization that many tetrapod genes have two orthologues in teleosts (Wittbrodt et al., 1998; Taylor et al., 2001). However, it was unclear whether these co-orthologues originated in a single WGD event or in a sequence of smaller duplications at the level of whole chromosomes or chromosomal pieces.

Several lines of evidence confirm that a WGD indeed took place at the root of the teleost lineage:

Early on, the developmentally important and well-conserved Hox gene clusters sparked interest. After identifying many supernumerous Hox genes to the already known ones from tetrapods (Njolstad et al., 1988; Misof and Wagner, 1996; Aparicio et al., 1997), the

systematic evaluations of Hox genes revealed seven Hox clusters in zebrafish (Amores et al., 1998; Prince et al., 1998) as opposed to the four found in tetrapods. Although duplicated Hox genes have also been identified in other teleosts, it was initially not clear whether the increase in the number of Hox clusters is universal for teleosts (Prohaska and Stadler, 2004). Recently, duplicated Hox gene clusters were found in the two most basal extant groups of teleost fishes, the Elopomorpha (including eels and tarpons) (Guo et al., 2009; Henkel et al., 2012). and Osteoglossomorpha (including bony tongues and elephantfish) (Chambers et al., 2009). Notably, the eels (European and Japanese eel) are currently the only fishes in which the complete set of the original eight Hox clusters has been observed (Guo et al., 2009; Henkel et al., 2012). Since Elopomorpha is the most basal teleost group (Arratia, 1997; Near et al., 2012) this strongly suggests that the ancestor of all living teleosts also possessed eight Hox clusters, consistent with a WGD at the base of teleost evolution (Figure 1).

Since the conservation of long stretches of gene order in the entire teleost lineage is an expected outcome of a WGD, the detection of conserved synteny (gene order on chromosomes) of Hox clusters and other genes in a number of teleost fishes was taken as strong evidence (Amores et al., 1998; Gates et al., 1999; Barbazuk et al., 2000; Taylor, 2003; Hoegg and Meyer, 2007; Sato et al., 2009). Additional support for TS-WGD has been gained by molecular clock analyses. A WGD is expected to result in the divergence of all the resulting paralogues at the same time. Indeed, in two molecular-clock analyses using the *Fugu* genome, clear peaks in age distribution of paralogous blocks of duplicated genes were observed (Christoffels, 2004; Vandepoele et al., 2004). Although the age estimates of the fish-specific genome duplication are slightly diverging (350 million years ago (mya) (Christoffels, 2004) versus 320 mya (Vandepoele et al., 2004)), both studies place it before the teleost radiation, which is consistent with a genome duplication at the base of ray-finned fishes.

Finally, whole-genome sequencing of a number of fish genomes provided conclusive evidence for at least one WGD in the whole teleost lineage (Aparicio et al., 2002; Jaillon et al., 2004; Kasahara et al., 2007; Howe et al., 2013; Schartl et al., 2013).

In summary, owing to convincing evidence on many levels, it is now near universally accepted that the TS-WGD took place.

Additional lineage-specific WGDs occurred in salmonids and some cyprinids

Additionally to the WGD at the base of teleost evolution, more recent genome duplications have shaped fish evolution. WGD events are well established for both salmonids and cyprinids.

The ancestor of all extant salmonids underwent a tetraploidization event (Johnson et al., 1987), according to most recent estimations probably between 80 and 50 mya (Alexandrou et al., 2013). From the 1970ies on (Ohno, 1970b), it has been suspected that the salmonid specific genome duplication event, which preceded the origin of migratory behavior, provided the genetic basis for this evolutionary innovation (Alexandrou et al., 2013).

Within Cyprinidae, the common carp and the goldfish have been suggested to be tetraploid (Ohno et al., 1967).

Analysis of microsatellite loci (David, 2003) and comparing the linkage map of the common carp to the zebrafish genome (Zhang et al., 2013) provided strong evidence for the duplication event in the common carp. Goldfish and the common carp are closely related and likely share the same tetraploid ancestor which underwent a genome duplication an estimated 5.6 – 11.3 mya (Wang et al., 2012).

Additional polyploidization events within the Cyprinidae have been described in some loaches (Cobitidae) (Ferris and Whitt, 1977a) and in suckers (Catostomidae) (Uyeno and

Smith, 1972). Once more fish species become subject to genomic analysis, we will likely see many more additional examples of WGDs in teleost sublineages.

The evolution of chromosome numbers after teleost WGDs

WGD initially leads to doubling of the chromosomal set. However, it is well known that chromosomes behave dynamically during evolution and undergo rearrangements, such as centric fusions by Robertsonian Translocation. This mechanism leads to two chromosomes being fused at their centromeres, resulting in a reduction of chromosome number. Approaches to infer the ancestral teleost prior to TS-WGD have consistently predicted a haploid chromosome number of 12 to 13 (Postlethwait et al., 2000; Jaillon et al., 2004; Kohn et al., 2006; Kasahara et al., 2007). Accordingly, TS-WGD resulted in a post-duplication ancestor with 24 or 26 chromosomes. More than 50% of all extant teleosts with data in the genome size database (http://wwwgenomesize.com) have indeed 24 or 25 chromosomes (Naruse et al., 2004), presumably representing the ancestral condition. Thus, the number of chromosomes remained nearly unchanged during evolution of most extant species. However, whereas the number of chromosomes remained fairly constant, the comparison of different teleost genomes to that of humans revealed a higher rate of chromosomal rearrangements other than fusions (Kasahara et al., 2007).

After the recent WGD in the ancestor of common carp and goldfish, chromosome numbers have also not been reduced. Both species have 50 chromosomes, twice as many as other Cyprinidae (Ohno et al., 1967).

Conserving chromosome numbers after WGD is not essential, as chromosome numbers of salmonids illustrate. Because the stem salmonid underwent WGD, unchanged chromosome numbers would result in extant salmonids with around 50 chromosomes, a number twice that

of their closest relatives. In contrast, although cells of salmonid fishes consistently have double the DNA content (Gregory et al., 2007) and chromosome arms (Phillips et al., 2009) as compared to their closest relatives, their chromosome numbers vary extensively between 26 and 51. Most of the species have a lower chromosome number than the original number after duplication (Supp. Fig.1). Therefore, chromosome fusions must have played a major role in shaping salmonid karyotypes, as proposed by Hartley (Hartley, 1986). Different modes of chromosome evolution appear to have acted in the evolution of different salmonid sublineages, leading to the diverse chromosome numbers observed (reviewed by Phillips and Rab, 2001).

3. The fates of duplicated genes after WGD

After WGD, all duplicated genes should be relieved from selective pressure and therefore would be expected to vanish over time. However, the fate of duplicated genes is more complicated and much more interesting (Figure 2).

WGD derived duplicate gene pairs can undergo different fates: One of the duplicates may be lost (non-functionalization), both duplicates may be retained basically unchanged, both duplicates may acquire changes so that the function of the ancestral gene is divided among the duplicates (a process called subfunctionalization), and finally one of the duplicate genes may acquire a new function (neofunctionalization). For clarity, we will describe the different scenarios as individual processes. But we like to stress that the categories are simplified, and that multiple scenarios may affect the evolution of individual genes. Different mechanisms can act successively to shape different phases of gene evolution. Furthermore, two or more mechanisms may act on the same duplicate gene pair simultaneously.

Non-functionalization

Immediately after WGD, the daughter genes of each ancestral gene are identical, and their functions are redundant. This suggests that selective constraint of maintaining both of them is low and that one of them is therefore free to disappear due to genetic drift. A classical model, first formulated by Ohno (Ohno, 1970a) predicts that loss of one paralogue is the most common outcome of duplicate gene evolution. This assumption is based on the simple fact that deleterious mutations are much more likely to occur than beneficial ones. Thus, one of the duplicates is expected to accumulate deleterious mutations, eventually leading to its silencing. Indeed, experimental studies have confirmed that non-functionalization is the most common scenario of duplicate gene evolution (Jaillon et al., 2004; Woods et al., 2005; Brunet et al., 2006).

Estimates have suggested that as few as 1-5% of duplicate genes have been retained in pufferfish (Aparicio et al., 2002; Jaillon et al., 2004). This is roughly in agreement with the first genome-wide, comparative analysis of five different fish species by Kassahn *et al*. (Kassahn et al., 2009) These authors found that in all five species examined, 3-4% of the genes show strong evidence for having originated in the ancient TS-WGD. Due to the design of the study, this number is probably underestimating the real abundance of gene retention after TS-WGD, and can be considered a minimum estimate, as pointed out by the authors themselves. The study also found that there is no difference in the percentage of duplicate gene retention between the five species analyzed. By looking at gene families, other studies have come to the conclusion that up to 20% of TS-WGD-duplicate genes may have been retained in zebrafish (Postlethwait et al., 2000; Postlethwait et al., 2004; Woods et al., 2005), which can be considered the maximum estimate of gene retention rate after the TS-WGD. Gene retention rates after the more recent WGD events in Salmonids and the lineage leading

to common carp are much higher: Probably more than 50% of all genes are still present in duplicates in the common carp and salmonids (Ferris and Whitt, 1977b; Allendorf, 1978).

Although non-functionalization is very frequent, Force and coauthors (Force et al., 1999) noted that the fraction of genes preserved after genome duplication events is higher than predicted by Ohno's classical model (Ohno, 1970a). In other words, the probability of a mutation being beneficial versus it being deleterious is too low to explain the number of duplicates observed after WGD events. Force *et al*. (Force et al., 1999) therefore proposed a seminal model of subfunctionalization with their Duplication-Degeneration-Complementatiion (DCC) model.

Subfunctionalization by Duplication-Degeneration-Complementation (DDC)

Genes usually have more than only one function. These functions can be represented by expression in different cell types or developmental stages. Different expression domains are regulated by transcription factors binding to distinct elements in regulatory regions of a gene.

The Duplication-Degeneration Complementation model (Force et al., 1999) proposes that duplicates can be conserved by complementary degenerative mutations in such regulatory regions. The degenerative mutations are neutral, because one gene still performs the ancestral function that was lost in the other one. By this mechanism, functions of a gene can be subdivided between the daughter genes, which – together – continue to perform the functions of their ancestral pre-duplication gene. After complementary loss of subfunctions, both genes will be fixed in the genome, because loss of either of them will disrupt the essential ancestral gene function. It is important to point out that this process can take place in the absence of selection. Genetic drift leading to complementary loss of subfunctions is sufficient to explain the DDC mechanism. A mathematical model provided by Force et al. demonstrates that the

higher the number of subfunctions of a gene pair, the higher the probability of fixation by the DDC model, and the lower the probability of non-functionalization (Force et al., 1999). Furthermore, the model predicts that the fate decision between non- and subfunctionalization is determined quickly on an evolutionary time scale (within a few million years). Taken together, the model suggests that the DDC mechanism has had a significant contribution to the proportion of retained duplicates after WGD that we can observe today. Indeed, numerous cases of TS-WGD duplicate gene pairs evolving by DDC have been reported (for example, see McClintock et al., 2001; Jovelin et al., 2007; Kassahn et al., 2009; Renninger et al., 2011; von Niederhäusern et al., 2013).

One demonstrative example of subfunctionalization is shown by the study of cellular retinaldehyde-binding proteins (CRALBP) in zebrafish (Fleisch et al., 2008). During light perception in photoreceptors, visual pigment absorbs a photon, leading to isomerization of its visual chromophore (11-cis-retinal). The visual pigment is replenished in two separate visual cycles, the canonical cycle located in the retinal pigment epithelium (RPE) and the non-canonical cycle in Müller glia cells (Fleisch and Neuhauss, 2010). The visual chromphore needs to be chaperoned by CRALBP, hence tetrapods express this protein in both RPE and Müller glia cells. In zebrafish there are two CRALBP paralogues, one expressed in Müller glia cells and one in the RPE (Collery et al., 2008; Fleisch et al., 2008). Functional analyses showed that they serve different functions in vision. Hence the ancestral function and expression domain is split up between two paralogues, who together make up the function of the presumptive ancestral gene duplicated at WGD.

Clear examples of subfunctionalization by changes on the amino acid level in animals are rare, but one exemplary case of Proopiomeloncortins in pufferfish duplicated by TS-WGD and subfunctionalized on the amino acid level has been documented (de Souza et al., 2005).

Some examples of subfunctionalization on the level of coding sequence after WGD have also been found in plant genomes (Cusack and Wolfe, 2007).

Subfunctionalization by escape from adaptive conflict (EAC)

Besides subfunctionalization by the DDC mechanism, an additional mode of subfunctionalization was first proposed by Hughes (Hughes, 1994) and has been later termed escape from adaptive conflict (EAC) by Des Marais and Rausher (DesMarais and Rausher, 2008). In this model, two duplicate genes evolve not solely by genetic drift, but when adaptive evolution is driving changes in both paralogues, leading to their divergence (DesMarais and Rausher, 2008). In this way, better adaptation of a different subfunction in each gene can be achieved. This scenario is expected to occur in cases when two subfunctions of a gene cannot be improved simultaneously, because optimization of one subfunction would negatively interfere with the other one. Duplication solves this conflict and one paralogue is free to acquire adaptive mutations to optimize one subfunction without compromising the performance of the other one, whereas the second paralogue can optimize another subfunction. Whether this model is able to explain a substantial fraction of retained paralogues depends on two factors: The abundance of multifunctional genes and the abundance of situation where optimization of one function impairs another one (Innan and Kondrashov, 2010). Physical modeling of amino acid chains suggests that EAC preferentially takes place under moderate selective pressure and therefore likely in genes that are not essential for survival, but that can substantially improve fitness if optimized (Sikosek et al., 2012). So far, duplicate gene evolution by EAC has been mainly a theoretical model and only few cases indicating EAC have been documented (DesMarais and Rausher, 2008; Deng et al., 2010; Huang et al., 2012). One reason for the scarcity of examples is related to the difficulty

to ascertain whether the criteria for EAC are fulfilled, in particular whether functions were improved compared to the ancestral gene and whether this improvement was really constrained before duplication (Barkman and Zhang, 2009).

Neofunctionalization

Besides non- and subfunctionalization, duplicate genes can also acquire novel functions. This is the classical model of neofunctionalization of one paralogue, again first formulated by Ohno (Ohno, 1970a). It is also referred to as “mutation during non-functionality” (MDN) model (Hughes, 1994; Conant and Wolfe, 2008). Due to the lack of selective constraint on maintaining both duplicates, one of them is free to acquire mutations conferring a new function. As discussed above, beneficial mutations occur only at a low rate. Therefore, this scenario is expected to be encountered less frequently than non- or subfunctionalization. Indeed, fewer instances of neofunctionalization have been confirmed. The technical difficulty to identify cases of neofunctionalization likely greatly contributes to the apparent scarcity of bona fide examples (Conant and Wolfe, 2008). Duplicate genes with divergent functions, one of which is new, might be gene pairs that underwent neofunctionalization. Instances of neofunctionalization that have been reported are frequently gain of novel expression domains, and therefore probably neofunctionalization by alterations in their regulatory regions (Kassahn et al., 2009), while changes in the coding sequence of a genes giving rise to a new function are rarer (Braasch et al., 2006; Douard et al., 2008).

One illustrative example of neofunctionalization in both regulatory and protein coding sequences of a TS-WGD duplicate gene pair is the co-option of a voltage-gated sodium channel to contribute to the origin and function of electric organs (Zakon et al., 2006; Arnegard et al., 2010). Both African mormyroid and South American gymnotiform fishes

possess organs to electrically generate communication signals. Interestingly, although the electric organs of those two groups are very similar (for example, they are both derived from skeletal muscle), they evolved independently (Alves-Gomes, 1999). The electric organ was therefore invented twice by convergent evolution. The function of electric organs highly depends on modified voltage-gated sodium channels that are needed to discharge electrocytes, the cells of the electric organs suitable to produce the communication signal. Teleost fishes possess *scn4aa* and *scn4ab*, two products of TS-WGD coding for alpha-subunits of voltage-gated sodium channels. In non-electrogenic fishes, both paralogues are expressed in skeletal muscle. But in all members of both groups of electrogenic fishes, expression of *scn4aa* was found to be lost from muscle and gained in the electric organ. The fact that this switch in expression is found even in the most basal electrogenic fishes in both groups suggests that *scn4aa* not only acts in, but also supported the formation of the electric organs (Arnegard et al., 2010). Furthermore, the authors found that selective pressure acted specifically on the protein-coding sequence of the *scn4aa* paralogue during phases when the electric organ was evolving, whereas selective pressure on *scn4ab* remained constantly low. Strikingly, selective forces showed to be particularly high in functionally important regions of the proteins, for example in extracellular loops that are thought to have an impact on the duration of electric organ discharge and therefore on the properties of the communication signal. These specific changes occurred in parallel in types of gymnotiform and mormyroid electric fishes that generate pulsed electric signals as opposed to the more uniform signals of their relatives. In summary, the innovation of electric organs in two distant related groups of weak electric fishes highlights two interesting aspects: Firstly, it demonstrates that genes which arose in genome duplications can acquire new functions, leading to the acquisition of evolution of new evolutionary traits, even complex ones such as new organs. Secondly, it illustrates an example of pre-adaptation and co-option: The duplicated sodium channel

existed in the genome of fishes and probably acted in muscle activity for around 100 million years (Arnegard et al., 2010) until, within an evolutionary short period of time, it was co-opted twice to function in the electric organ.

Gene dosage effects

A final mechanism of duplicate gene retention worth discussing here is retention due to dosage effects. Directly after WGD, all chromosomes and genes are doubled in every cell, and therefore it can be assumed that the duplicate gene pairs are all expressed at a higher level than the corresponding ancestral gene. Since this is true for every gene, relative gene dosage is not disrupted by WGD. Maintaining gene dosage balance seems to be crucial for some genes, and loss of so-called dosage-sensitive genes after WGD can be detrimental. Degenerative mutations of such genes disrupt balanced expression of genes interconnected in networks. Because relative gene dosages of such genes are important, reducing gene dosage by deleterious mutations in one paralogue can lead to negative developmental or physiological consequences. Genes where dosage is believed to be especially important are ribosomal genes, genes coding for proteins with a high number of interactions, and genes encoding proteins functioning in signaling pathways and networks. Requirement for gene dosage maintenance can lead to scenarios in which all genes of a network or pathway remain duplicated, and it has been suggested that retention of duplicate members in whole networks, can have broad evolutionary implications (Conant and Wolfe, 2007).

Support for this Gene Balance Hypothesis comes from comparing trends in gene retention between single-gene duplications and WGDs. Relative gene dosages are after a WGD initially not changed, while single-gene duplications instantaneously disrupt gene dosage balance and should therefore be selected against in dosage-sensitive systems. Indeed, it has

been shown in vertebrates and plants that highly interconnected genes, such as genes involved in transcription and signaling cascades, and genes coding for proteins with more than average protein-protein interactions are over-retained after WGDs, but not after small-scale duplications (Blomme et al., 2006; Freeling, 2008; Hufton et al., 2009).

Hufton et al. even propose that the gene balance hypothesis better explains gene duplication retention in vertebrate genomes than the DDC model (Hufton et al., 2009). In their study of phylogenetically conserved noncoding sequences, they showed that genes retained after WGD are rather marked by many protein interaction sites than by many conserved noncoding elements, as the DDC model would predict (Hufton et al., 2009). Additionally, it is plausible that an increased dose of some genes is beneficial even if they are not highly interactive. In such a case, both duplicates will be preferentially retained in the genome as well. Examples for genes that are required in high doses and that are therefore prone to be maintained as duplicates by positive selection are histones and ribosomal proteins (Sugino and Innan, 2006). However, both histones and ribosomal proteins are also dependent on the abundance of their interaction partners (other histones, other ribosomal proteins and ribosomal RNA), and might therefore be retained by both the benefits of an increased dose and the need to keep dosage balance.

As noted at the beginning of the section, the different mechanisms leading to different fates cannot be regarded as isolated processes since they can act together, resulting in complex evolutionary dynamics of duplicate genes. On top of that, outcomes of duplicate gene evolution are also affected by other interesting evolutionary mechanisms such as gene conversion.

Concerted evolution by gene conversion

Gene conversion has been described as nonreciprocal exchange of DNA fragments between homologous sequences within a genome. Gene conversion can also be regarded as a copy-and-paste event by which a gene fragment is replaced by a homologous sequence. When gene conversion is active between genes at a sufficiently high rate, those genes do not evolve independently anymore, but in a fashion called concerted evolution. The principal effect of concerted evolution is that affected genes remain more similar to each other than would be expected considering only divergent evolution without gene conversion.

Gene conversion is dependent on homology and sufficient sequence similarity (Ahn et al., 1988; Elliott et al., 1998), and can therefore also be expected to be active between paralogues arising through WGDs.

Several effects of gene conversion on duplicated genes have been suggested (reviewed by Innan, 2009): First, gene conversion is expected to make non-functionalization less likely, since deleterious mutations in one duplicate can be removed by “pasting” the corresponding sequence of its intact paralogue. Second, gene conversion can contribute to neofunctionalization, since beneficial mutations can be shared and also, novel combinations of allelic sequences can be created. In a theoretical model, Teshina and Innan have explored another interesting effect of gene conversion by which it counteracts neofunctionalization (Teshima and Innan, 2008). A DNA sequence conferring a novel function can be converted back to the ancestral sequence by gene conversion. Therefore, in genes undergoing conversion, neofunctionalization can only occur under strong selection. A second consequence suggested by this study is that gene conversion prevents complete fixation of a novel gene. As a result, fixation of neofunctionalization can only take place subsequently to mechanisms that terminate gene conversion, i.e. progressive sequence divergence or events of immediate large impact such as transposon insertions.

Clear examples of gene conversion are documented in plants, fungi and animals (for example, see Semple and Wolfe, 1999; Drouin, 2002; Rozen et al., 2003; Mondragon-Palomino and Gaut, 2005). It has also been shown that gene conversion can principally be active after WGD by studies in yeast (Wolfe and Shields, 1997; Kellis et al., 2004). However, the extent to which gene conversion contributes to duplicate gene conversion, in particular after WGDs, is still unclear due to difficulties in detecting gene conversion with current methods (Mansai and Innan, 2010).

Almost all gene conversion events discovered in teleosts have affected paralogues that do not stem from TS-WGD but rather from more recent duplications (mostly tandem duplications) in teleost sublineages (Bargelloni et al., 1999; McGuigan et al., 2004; Noonan et al., 2004; Gerrard and Meyer, 2007; Yu et al., 2007; Windsor and Owens, 2009; Weadick and Chang, 2012). Only one instance of gene conversion between paralogues generated by TS-WGD, namely rainbow trout sox9α2 and sox9 (Alfaqih et al., 2009) has so far been documented. However, the scarcity of examples does not necessarily imply that gene conversion was infrequent or unimportant after TS-WGD. Since gene conversion depends on sufficient sequence similarity between paralogues, it is expected to be most common directly following WGD events and becoming less frequent over the course of time. Therefore, many gene conversion events acting on paralogues duplicated by TS-WGD can be expected to be ancient and quite hard to detect.

Coding and noncoding regions in duplicate gene evolution

Most of the mechanisms we have discussed can either be achieved by mutations in coding- or noncoding regions. In the case of subfunctionalization, gene expression of duplicate genes can be divided between tissues by reciprocal loss of cis-regulatory elements, as Force *et al*.

initially postulated (Force et al., 1999). However, distribution of an ancestral function onto two paralogous genes can in principal also be achieved by reciprocal inactivation of functional domains. Similarly, a new function can be gained through the addition of another regulatory element resulting in a new expression domain, or through imposing a new function via changes within the coding sequence.

It is still lively debated whether changes in coding- or non-coding sequences are more relevant to the evolution of genes and new traits (Carroll, 2000; Hoekstra and Coyne, 2007; Wray, 2007; Lynch and Wagner, 2008; Wittkopp and Kalay, 2012). One of the difficulties is that it is straightforward to detect changes in coding regions, while cis-regulatory elements are small, interspersed with non-relevant sequences, often far away from the regulated gene, and less strictly conserved in sequence. Furthermore, their position can be changed or they can be inverted without functional consequences. Therefore meaningful changes in non-coding sequences are much harder to identify. Hence it is no surprise that studies having directly identified cis-regulatory sequence changes as the source of divergent duplicate gene expression are rare. However, the consequences of cis-regulatory mutations can easily be identified by expression analyses, such as (quantitative) reverse PCR and RNA *in-situ* hybridization. In fact, changes in expression patterns have often been interpreted by authors as alterations in cis-regulatory regions. This may often be the appropriate conclusion, however the methods used might not always detect alternatively spliced variants of a paralogue which might be prevalent in a certain tissue, and mRNA abundances can also be altered through changes in mRNA stability (Hoekstra and Coyne, 2007).

Comparative genome wide analyses of retained duplicates after the TS- WGD from five species (Kassahn et al., 2009) showed that expression patterns often diverge between paralogues: 87% of duplicate gene pairs showed distinct expression patterns (indicative of neo- and subfunctionalization events) in at least one developmental stage examined. This

number represents only a rough estimate since it might underestimate the actual number of divergent genes. Some paralogues might be differentially expressed in developmental stages not examined. This is especially relevant since adult stages are often not included in expression analyses, and the same survey showed that expression patterns of duplicate gene pairs get more distinct during development.

This study suggests that neofunctionalization through changes in regulatory regions might be more abundant than predicted by the classical model of neofunctionalization by Ohno (Ohno, 1970a). This discrepancy can be alleviated if the "Duplication Degeneration Innovation" (DDI) model proposed by Jiménez-Delgado and coauthors is taken into account (Jimenez-Delgado et al., 2009). In the DDI model, sub- and neofunctionalization act together on regulatory elements to achieve evolutionary innovation. After duplication and during degeneration, conserved non-coding elements (CNEs) become non-functional, but retain their structural enhancer properties. Therefore, expression in a new spatial and/or temporal manner can be achieved even by only subtle mutations in those degenerate CNEs, making neofunctionalization more likely to occur.

When contemplating evolutionary divergence of coding regions, the first events that come to mind are amino acid substitutions caused by non-synonymous point mutations. While this mechanism has received considerable attention for decades, there is accumulating evidence that other mechanisms, which have been started to be investigated more recently, are also significantly contributing to structural and functional divergence of proteins after genome duplications.

Divergence of coding regions can be achieved by a number of mechanisms other than point mutations. These include insertions/deletions (indels), exon gain/loss, exonization/pseudoexonization, exon suffling and divergence of alternative splicing.

Divergence in splicing and indels have been suggested to contribute substantially to evolution after WGD.

Divergence of splicing events can principally play a role after polyploidization as shown in a study in plants (Brassicaceae) (Zhou et al., 2011). In this study natural and resynthesized tetraploid species were compared to a closely related diploid species in terms of splicing patterns of paralogous genes. A substantial number (>20%) of paralogous gene pairs were found to have diverged in splicing events. The resynthesized tetraploids showed that those changes occur fast: Already after 5 generations, more than 20% of duplicate gene pairs showed divergent splicing patterns. The study also showed that the most common change in splicing is loss of one parental splicing event in a duplicate gene. Whether such a mechanism is equally prevalent in teleost evolution is not known, but provides an attractive alternative mechanism to explain duplicate gene retention due to partition of alternative splicing between paralogues.

Indels have also been shown to very frequently contribute to divergent gene evolution after the TS-WGD. Both members of a paralogous gene pair have experienced significantly more insertion and deletion events than genes not retained as duplicates. These indels mostly occurred shortly after the duplication event and are predicted to affect protein structure more than amino acid substitutions (Guo et al., 2012).

This and other reports (Brunet et al., 2006; Jiang and Blouin, 2007; Tian et al., 2008; Chen et al., 2009) have led to the idea that Indels have at least as much impact on duplicate gene evolution as nucleotide substitutions do.

A number of studies were conducted aiming to address evolution in both coding and noncoding regions of TS-WGD duplicates. In some cases, both noncoding and coding

sequences of certain paralogues were found to undergo divergent evolution. In particular, such scenarios were shown for Proopiomelanocortins, prohormones mostly expressed in the pituitary gland (de Souza et al., 2005), and Follistatins, TGF-β binding proteins involved in muscle development (Macqueen and Johnston, 2008). Other studies exclusively identified divergent evolution in noncoding regions, while the coding sequence or gene function remained highly conserved. This was true for a duplicated gene pair of the argonaute (*Ago*) family, encoding AGO proteins important for small RNA mediated gene silencing (McFarlane et al., 2011), and for IGFBP-2 genes, binding and regulating actions of Insulin-like growth factor (Zhou et al., 2011). None of these studies reported identical expression patterns of TS-WGD duplicates, emphasizing that changes in regulatory elements are very common after WGD.

Over-retained duplicates after WGDs

Having discussed the mechanisms for duplicate gene retention, an obvious follow-up question is if the retained genes are distributed equally among gene categories. Strong evidence, mainly obtained in plants, yeast and unicellular eukaryotes has been collected that duplicate genes with certain properties are over-retained after WGD events (for example, see Seoighe and Wolfe, 1999; Papp et al., 2003; Maere et al., 2005; Aury et al., 2006).

In agreement with the dosage-balance hypothesis, over-retained duplicate genes often encode proteins with more than average protein-protein interactions and proteins that function in complexes (Hakes et al., 2007). Correspondingly functional categories that have been over-retained include ribosomal proteins, protein kinases and transcription factors. Two recent studies have shown that gene expression is a factor highly correlated with duplicate gene retention. Gout *et al*. investigated the relation between gene expression and duplicate gene

retention on a genome-wide basis in the unicellular *Paramecium tetraurelia* (Gout et al., 2010). The evolutionary lineage leading to *P. tetraurelia* is marked by three rounds of WGDs. There was a strong positive relationship between expression level and duplicate gene retention. Similarly, Chain *et al*. found that expression level is the factor correlating most strongly with duplicate gene retention in tetraploid *Xenopus laevis*, again suggesting dosage sensitivity of retained duplicates (Chain et al., 2011). Evenness of expression was the second strongest factor positively associated with duplicate gene retention. “Evenness” of expression means activation in many tissues and therefore might be linked to pleiotropy (multifunctionality) and complexity of regulatory sequences, features that increase the chance of sub- and neo functionalization.

Another observation made in both *P. tetraurela* and *X. laevis* was that genes which have been evolving slowly before a genome duplication are more likely to be retained (Gout et al., 2010; Chain, Frederic J. J. et al., 2011). Consistent with this finding, Semon and Wolfe have previously argued that slowly evolving genes may tend to persist because they give sub- or neofunctionalization more time to take place before deleterious mutations occur (Semon and Wolfe, 2008). An observation that points to the same direction is that in pufferfish, well conserved genes with close homologues already present in invertebrates are overrepresented among duplicates (Kassahn et al., 2009).

By assigning functions to duplicate and non-duplicate genes of five teleost species, it was shown that a number of functional categories is strongly enriched among paralogues derived from TS-WGD (Kassahn et al., 2009). The categories most enriched are related to ion channel and transporter activity. Ion transport needs to be tightly regulated in any cell. However, neurons are the cells that most strongly rely on a repertoire of diverse ion channels and transporters. Consistent with this genome-wide analysis, also studies of protein families in zebrafish have shown that genes involved in neuronal function have often retained both

paralogues (Gesemann et al., 2010; Di Donato et al., 2013; Haug et al., 2013; Kastenhuber et al., 2013).

4. Consequences of TS-WGD for fish evolution

Teleost fishes represent the by far most diverse vertebrate clade, constituting more than 32,000 species of an estimated total number of 64,000 vertebrate species (Froese and Pauly, 2013). Teleost fishes populate a wide range of oceanic and freshwater habitats all over our planet, ranging from arctic to tropic regions. Without doubt, teleosts are an evolutionary highly successful group.

WGDs are found at the base of some other diverse taxa. The vertebrate stem is marked by two rounds of WGD (Dehal and Boore, 2005; Putnam et al., 2008; Kuraku et al., 2009). Similarly, ancient WGDs have been documented in flowering plants (Jaillon et al., 2007; Tang et al., 2008). Additionally in flowering plants, more recent WGDs at the base of diverse subgroups have been reported (reviewed by Soltis et al., 2009). These and similar observations made in Fungi and unicellular eukaryotes (Aury et al., 2006; Scannell et al., 2007) have led to the idea that there is a causal correlation between WGD, evolutionary success and radiation.

Here, we are going to discuss the impact of TS-WGD on teleost evolution, focusing on its role in their radiation. First, we will briefly summarize the mechanisms by which WGDs are thought to impose selective advantage and facilitate speciation. For an extensive review on this matter, the reader is referred to Van de Peer *et al*. (Van de Peer et al., 2009). Then, we will discuss the current and controversial state of evidence on teleost radiation driven by TS-WGD.

Mechanisms by which WGDs can contribute to evolutionary success and radiation

Although polyploidization most often leads to an evolutionary dead end, it seems that polyploid organisms sometimes have advantages over their diploid relatives. In particular, some polyploids have been suggested to be more robust to changing environments, therefore having reduced risk of extinction (Fawcett et al., 2009). Rapid genomic and epigenetic changes taking place after WGD (Osborn et al., 2003) probably enable polyploids to adapt faster than diploids. Furthermore, polyploids have been suggested to have increased mutational robustness, meaning that redundant genes copies can transiently mask the effect of deleterious mutations in their paralogue (Otto and Whitton, 2000).

WGDs have also been suggested to directly facilitate speciation by reciprocal gene loss, where different paralogues are lost in different populations, ultimately leading to genetic isolation and speciation of these populations (Scannell et al., 2006). There is indeed evidence for reciprocal gene loss in teleost lineages. It was found that 8% of gene loci of *Tetraodon nigroviridis* (green spotted puffer) and zebrafish underwent reciprocal gene loss (Sémon and Wolfe, 2007). Similar to reciprocal gene loss, also subfunctionalization has the potential to lead to genetic isolation of populations (Lynch and Force, 2000; Postlethwait et al., 2004; Volff, 2005).

Also evolutionary innovations made possible by WGD provide a path to evolutionary success. Gene duplication of any kind is a crucial generator of raw material for evolutionary innovation. As discussed before, however, WGDs uniquely enable duplication of dosage sensitive genes. Such genes include regulatory genes, thought to eminently contribute to the emergence of evolutionary innovations. Regulatory gene have also been over-retained in fishes (Blomme et al., 2006; Brunet et al., 2006). Additionally, dosage sensitivity is expected

to result in the retention of whole transcriptional networks that can as a whole get assigned a novel function (Freeling and Thomas, 2006; Freeling, 2009).

Finally, WGD leads to rapid expansion of whole gene families that can be retained to evolve and generate many genes with similar but not identical function. Thus, it is reasonable to assume that WGD particularly enables fine tuning and optimization of already existing functions.

State of evidence for TS-WGD causing evolutionary success and radiation

At first, we want to stress that WGDs are neither necessary for nor predictive of diversification and radiation. Although WGDs are found at the base of radiating clades such as vertebrates and flowering plants, there are also many species-rich lineages that show no signs of WGD, for instance in Coleoptera (beetles), the most diverse group of insects, consisting of over 360,000 species. Conversely lineages that are not unusually species-rich have undergone WGD, such as the Salmoniformes. Although this teleost lineage underwent an additional round of WGD, they are with 222 species comparatively species poor (Froese and Pauly, 2013). A conclusive assessment of the correlation between occurrence of WGDs and evolutionary rate or radiation requires a more complete picture on WGD across the tree of life. The current data may very well be biased in favour of a causative role of WGD, simply due to the unequal number of species in different lineages. We expect that many more WGDs across the whole tree of life will be revealed by future genome analyses.

Due to the lack of clear correlation between WGD and diversification, the sole fact that TS-WGD took place cannot be taken as evidence for it generating teleost diversity.

A way to address the question whether TS-WGD enabled teleosts to radiate is to closely look at the timing of TS-WGD and the rate of teleost diversification. A recent study estimating the

timing of diversification rates in teleosts indeed revealed a prominent diversification event at the base of teleost evolution (Santini et al., 2009). This result supports a role of WGD in the diversification of teleosts. However, two additional and more recent diversification events were detected, preceding the radiation of Percomorpha and Ostariophysi, two particularly species-rich teleost clades. The delay between TS-WGD and more recent occurrence of diversification puts a causal link between TS-WGD and diversification into question. Since around 88% of species richness stems from the two more recent diversification events (Santini et al., 2009), it is suggested that it was not the primary factor generating teleost diversity. Also recent time-calibrated phylogeny of ray-finned fishes showed that the major teleost lineages originated late after TS-WGD, in the late Mesozoic and early Cenozoic (Near et al., 2012).

Interestingly, similar patterns of WGD followed by delayed radiation were found in several large groups of flowering plants that independently underwent WGD. Such lineages show greater biodiversity than lineages of flowering plants without WGD (Soltis et al., 2009). However, the major shifts in diversification did not immediately follow WGD, but took place in later emerging subclades (Smith et al., 2011). At least six species-rich families of flowering plants show a pattern of WGD followed by the emergence of both radiating and species-poor subclades (Schranz et al., 2012). Also the phylogenetic tree of teleosts shows such a pattern with the species-poor Elopomorpha and Osteoglossomorpha at the base and later evolving highly diverse clades such as Cyprinidae and Percomorpha. These similarities led Schranz *et al.* to propose that a temporal delay between WGD and radiation is a pattern generally observed and called it “time-lag model” (Schranz et al., 2012). Also in salmonids, a gap of 40-50 million years between WGD (which likely took place 88-103 mya) and diversification (although low in comparison to some other teleost groups) was reported (Macqueen and Johnston, 2014).

Since such patterns of delayed diversification seem to be common, they are likely not coincidental. It is possible that WGDs enable radiation long after the duplication event occurred. Indeed reciprocal gene loss and subfunction partitioning have been shown to take place long after WGD in many instances (Scannell et al., 2006; Sémon and Wolfe, 2007).

If TS-WGD did not directly drive teleost diversification, which factors did and how are they related to TS-WGD? There is not enough data available to answer these questions conclusively, but a picture is emerging.

The most prominent phase marked by radiation came with the appearance of the Acanthomorpha, the most diverse group of teleosts, including the species-rich Percomorpha. The appearance of Acanthomorpha 100-150 mya (Near et al., 2012) preceded the so-called teleost explosion, the phase when a dramatic number of new fish species emerged.

Acanthomorpha radiated in the oceans, but descended from freshwater ancestors. This transition from fresh- to saltwater is a major adaptive step due to different osmoregulatory demands of the marine environment. Once this adaptive hurdle is taken, oceans provide a rich biotope to diverge into. Therefore adaptation of Acanthomorpha to the high salinity is a likely cause of their massive oceanic radiation. Eggs principally have the same osmolarity as the maternal body fluids and are therefore hypoosmotic to sea water. If such an egg is spawned in the hyperosmotic ocean, it will suffer from osmotic water efflux. It has been shown that marine fishes increase the osmolarity of their eggs by cleaving yolk proteins, which in turn are derived from Vitellogenin (Vtg). This cleavage is especially prominent in pelagic eggs, resulting in a large amount of free amino acids driving their hydration and thus making the eggs even float (Amores et al., 1998; Finn et al., 2002).

Phylogenetic analysis showed that Vtgs are evolutionary derived from a large-lipid transfer molecule that predated the origin of Bilateria (Finn and Kristoffersen, 2007). Vtg evolution in vertebrates is marked by numerous duplication events (both WGD events and local

duplications) as well as gene losses. Teleost Vtgs can be assigned to three different types, and Acanthopterygii show a lineage specific duplication of one of the Vtgs (VtgA), resulting in VtgAa and VtgAb (Finn and Kristoffersen, 2007). The free amino acid pool in pelagic eggs mainly stems from VtgAa paralogues (Matsubara et al., 1999), while VtgAb is essentially not degraded (La Fleur et al., 2005). VtgAb thus functionally represents the ancestral state. This ancestral state of VtgAb versus the derived state of VtgAa could also be confirmed by evolutionary rate analysis between the two paralogues (Finn and Kristoffersen, 2007). Finn and Kristoffersson conclude that the hydration of marine eggs and therefore the oceanic radiation of Acanthomorpha were made possible by a post-WGD event. TS-WGD only contributed indirectly by the expansion of the Vtg gene family that preceded the crucial duplication event in Acanthomorpha.

In salmonids, climatic changes have recently been suggested to have caused their diversification. Most salmonid lineages and species only formed in the last 10 myr, with two clades independently evolving anadromy (migratory behavior from freshwater to the oceans and back for reproduction). Macqueen and Johnston found that the shift in diversification correlates with climatic cooling, and argue that this climate change might have provided a selective advantage for anadromous behavior, since marine productivity exceeds that of freshwater in a temperate climate, providing more abundant food sources (Macqueen and Johnston, 2014). Migrating to the oceans also offered new freshwater habitats. Via estuaries, salmonids were now able to enter new river systems with new ecological demands, stimulating speciation. It has been speculated that anadromy was made possible by WGD at the base of salmonids, but clear evidence is still missing (Alexandrou et al., 2013).

In summary, there are good reasons to believe that TS-WGD has been important in generating teleost complexity. However, the time delay between TS-WGD and phases of

extensive speciation suggest that TS-WGD has not been the direct factor generating teleost diversity. Ecological changes followed by adaptations also had a large impact. TS-WGD probably provided teleost fishes with diversification potential that can be utilized when needed, even after tens of millions of years. In other words, TS-WGD may very well set the stage for important ecological adaptations.

Acknowledgements:

We acknowledge funding from the Swiss National Science Foundation (31003A_135598).

References

Ahn, B.Y., Dornfeld, K.J., Fagrelius, T.J., and Livingston, D.M. (1988). Effect of limited homology on gene conversion in a Saccharomyces cerevisiae plasmid recombination system. Mol Cell Biol *8*, 2442–2448.

Alexandrou, M.A., Swartz, B.A., Matzke, N.J., and Oakley, T.H. (2013). Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. Molecular Phylogenetics and Evolution *69*, 514–523.

Alfaqih, M.A., Steele, C.A., Morris, R.T., and Thorgaard, G.H. (2009). Comparative genome mapping reveals evidence of gene conversion between Sox9 paralogs of rainbow trout (Oncorhynchus mykiss). Comp Biochem Physiol Part D Genomics Proteomics *4*, 147–153.

Allendorf, F.W. (1978). Protein polymorphism and the rate of loss of duplicate gene expression. Nature *272*, 76–78.

Alves-Gomes (1999). Systematic biology of gymnotiform and mormyriform electric fishes: phylogenetic relationships, molecular clocks and rates of evolution in the mitochondrial rRNA genes. J Exp Biol *202*, 1167–1183.

Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., Westerfield, M., Ekker, M., and Postlethwait, J.H. (1998). Zebrafish hox clusters and vertebrate genome evolution. Science *282*, 1711–1714.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., Gelpke, M., Roach, J., Oh, T., Ho, I., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S., Clark, M., Edwards, Y., Doggett, N., Zharkikh, A., Tavtigian, S., Pruss, D., Barnstead, M., Evans, C., Baden, H., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, D., and Brenner, S. (2002). Whole-Genome Shotgun Assembly and Analysis of the Genome of Fugu rubripes. Science *297*, 1301–1310.

Aparicio, S., Hawker, K., Cottage, A., Mikawa, Y., Zuo, L., Venkatesh, B., Chen, E., Krumlauf, R., and Brenner, S. (1997). Organization of the Fugu rubripes Hox clusters: evidence for continuing evolution of vertebrate Hox complexes. Nat Genet *16*, 79–83.

Arnegard, M.E., Zwickl, D.J., Lu, Y., and Zakon, H.H. (2010). Old gene duplication facilitates origin and diversification of an innovative communication system--twice. Proc Natl Acad Sci U S A *107*, 22172–22177.

Arratia, G. (1997). Basal teleosts and teleostean phylogeny (Pfeil: München).

Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Camara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A.-M., Kissmehl, R., Klotz, C., Koll, F., Le Mouel, A., Lepere, G., Malinsky, S., Nowacki, M., Nowak, J.K., Plattner, H., Poulain,

J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Betermier, M., Weissenbach, J., Scarpelli, C., Schachter, V., Sperling, L., Meyer, E., Cohen, J., and Wincker, P. (2006). Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature *444*, 171–178.

Barbazuk, W.B., Korf, I., Kadavi, C., Heyen, J., Tate, S., Wun, E., Bedell, J.A., McPherson, J.D., and Johnson, S.L. (2000). The syntenic relationship of the zebrafish and human genomes. Genome Res *10*, 1351–1358.

Bargelloni, L., Scudiero, R., Parisi, E., Carginale, V., Capasso, C., and Patarnello, T. (1999). Metallothioneins in antarctic fish: evidence for independent duplication and gene conversion. Mol Biol Evol *16*, 885–897.

Barkman, T., and Zhang, J. (2009). Evidence for escape from adaptive conflict? Nature *462*, E1; discussion E2-3.

Blomme, T., Vandepoele, K., Bodt, S. de, Simillion, C., Maere, S., and Van de Peer, Yves (2006). The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol *7*, R43.

Braasch, I., Salzburger, W., and Meyer, A. (2006). Asymmetric evolution in two fish-specifically duplicated receptor tyrosine kinase paralogons involved in teleost coloration. Mol Biol Evol *23*, 1192–1202.

Brunet, F.G., Roest Crollius, H., Paris, M., Aury, J.-M., Gibert, P., Jaillon, O., Laudet, V., and Robinson-Rechavi, M. (2006). Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Mol Biol Evol *23*, 1808–1816.

Canestro, C., Albalat, R., Irimia, M., and Garcia-Fernandez, J. (2013). Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. Semin Cell Dev Biol *24*, 83–94.

Carroll, S.B. (2000). Endless forms: the evolution of gene regulation and morphological diversity. Cell *101*, 577–580.

Chain, Frederic J. J., Dushoff, J., and Evans, B.J. (2011). The odds of duplicate gene persistence after polyploidization. BMC Genomics *12*, 599.

Chambers, K.E., McDaniell, R., Raincrow, J.D., Deshmukh, M., Stadler, P.F., and Chiu, C.-h. (2009). Hox cluster duplication in the basal teleost Hiodon alosoides (Osteoglossomorpha). Theory Biosci. *128*, 109–120.

Chen, J.-Q., Wu, Y., Yang, H., Bergelson, J., Kreitman, M., and Tian, D. (2009). Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. Mol Biol Evol *26*, 1523–1531.

Christoffels, A. (2004). Fugu Genome Analysis Provides Evidence for a Whole-Genome Duplication Early During the Evolution of Ray-Finned Fishes. Molecular Biology and Evolution *21*, 1146–1151.

Collery, R., McLoughlin, S., Vendrell, V., Finnegan, J., Crabb, J.W., Saari, J.C., and Kennedy, B.N. (2008). Duplication and Divergence of Zebrafish CRALBP Genes Uncovers Novel Role for RPE-

and Muller-CRALBP in Cone Vision. Investigative Ophthalmology & Visual Science *49*, 3812–3820.

Conant, G.C., and Wolfe, K.H. (2007). Increased glycolytic flux as an outcome of whole-genome duplication in yeast. Mol Syst Biol *3*, 129.

Conant, G.C., and Wolfe, K.H. (2008). Turning a hobby into a job: how duplicated genes find new functions. Nat. Rev. Genet. *9*, 938–950.

de Souza, Flavio S. J., Bumaschny, V.F., Low, M.J., and Rubinstein, M. (2005). Subfunctionalization of expression and peptide domains following the ancient duplication of the proopiomelanocortin gene in teleost fishes. Mol Biol Evol *22*, 2417–2427.

Dehal, P., and Boore, J.L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol *3*, e314.

Deng, C., Cheng, C.-H.C., Ye, H., He, X., and Chen, L. (2010). Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. Proc Natl Acad Sci U S A *107*, 21593–21598.

DesMarais, D.L., and Rausher, M.D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature *454*, 762–765.

Di Donato, V., Auer, T.O., Duroure, K., and Del Bene, F. (2013). Characterization of the calcium binding protein family in zebrafish. PLoS ONE *8*, e53299.

Douard, V., Brunet, F., Boussau, B., Ahrens-Fath, I., Vlaeminck-Guillem, V., Haendler, B., Laudet, V., and Guiguen, Y. (2008). The fate of the duplicated androgen receptor in fishes: a late neofunctionalization event? BMC Evol Biol *8*, 336.

Drouin, G. (2002). Characterization of the gene conversions between the multigene family members of the yeast genome. J Mol Evol *55*, 14–23.

Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J.A., and Jasin, M. (1998). Gene conversion tracts from double-strand break repair in mammalian cells. Mol Cell Biol *18*, 93–101.

Fawcett, J.A., Maere, S., and Van de Peer, Yves (2009). Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc Natl Acad Sci U S A *106*, 5737–5742.

Ferris, S.D., and Whitt, G.S. (1977a). Duplicate gene expression in diploid and tetraploid loaches (Cypriniformes, Cobitidae). Biochem Genet *15*, 1097–1112.

Ferris, S.D., and Whitt, G.S. (1977b). The evolution of duplicate gene expression in the carp (Cyprinus carpio). Experientia *33*, 1299-1301.

Finn, R.N., and Kristoffersen, B.A. (2007). Vertebrate vitellogenin gene duplication in relation to the "3R hypothesis": correlation to the pelagic egg and the oceanic radiation of teleosts. PLoS ONE *2*, e169.

Finn, R.N., Ostby, G.C., Norberg, B., and Fyhn, H.J. (2002). In vivo oocyte hydration in Atlantic halibut (Hippoglossus hippoglossus); proteolytic liberation of free amino acids, and ion transport, are driving forces for osmotic water influx. J Exp Biol *205*, 211–224.

Fleisch, V.C., and Neuhauss, S.C. (2010). Parallel visual cycles in the zebrafish retina. Prog Retin Eye Res *29*, 476–486.

Fleisch, V.C., Schonthaler, H.B., Lintig, J. von, and Neuhauss, Stephan C F (2008). Subfunctionalization of a retinoid-binding protein provides evidence for two parallel visual cycles in the cone-dominant zebrafish retina. J. Neurosci. *28*, 8208–8216.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics *151*, 1531–1545.

Freeling, M. (2008). The evolutionary position of subfunctionalization, downgraded. Genome Dyn *4*, 25–40.

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol *60*, 433–453.

Freeling, M., and Thomas, B.C. (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res *16*, 805–814.

Froese, R., and Pauly, D. (2013). FishBase. www.fishbase.org, version (12/2013).

Gallardo, M.H., Bickham, J.W., Honeycutt, R.L., Ojeda, R.A., and Kohler, N. (1999). Discovery of tetraploidy in a mammal. Nature *401*, 341.

Gallardo, M.H., Klausel, G., Jiménez, Al., Bacquet, C., González, C., Figueroa, J., Köhler, N., and Ojeda, R. (2004). Whole-genome duplications in South American desert rodents (Octodontidae). Biol J Linn Soc *82*, 443–451.

Gates, M.A., Kim, L., Egan, E.S., Cardozo, T., Sirotkin, H.I., Dougan, S.T., Lashkari, D., Abagyan, R., Schier, A.F., and Talbot, W.S. (1999). A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. Genome Res *9*, 334–347.

Gerrard, D.T., and Meyer, A. (2007). Positive selection and gene conversion in SPP120, a fertilization-related gene, during the East African cichlid fish radiation. Mol Biol Evol *24*, 2286–2297.

Gesemann, M., Lesslauer, A., Maurer, C.M., Schönthaler, H.B., and Neuhauss, Stephan C F (2010). Phylogenetic analysis of the vertebrate excitatory/neutral amino acid transporter (SLC1/EAAT) family reveals lineage specific subfamilies. BMC Evol. Biol. *10*, 117.

Gout, J.-F., Kahn, D., and Duret, L. (2010). The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. PLoS Genet *6*, e1000944.

Gregory, T.R., Nicol, J.A., Tamm, H., Kullman, B., Kullman, K., Leitch, I.J., Murray, B.G., Kapraun, D.F., Greilhuber, J., and Bennett, M.D. (2007). Eukaryotic genome size databases. Nucleic Acids Res *35*, D332-8.

Guo, B., Gan, X., and He, S. (2009). Hox genes of the Japanese eel Anguilla japonica and Hox cluster evolution in teleosts. J. Exp. Zool. *9999B*, n/a.

Guo, B., Zou, M., and Wagner, A. (2012). Pervasive indels and their evolutionary dynamics after the fish-specific genome duplication. Mol Biol Evol *29*, 3005–3022.

Hakes, L., Pinney, J.W., Lovell, S.C., Oliver, S.G., and Robertson, D.L. (2007). All duplicates are not equal: the difference between small-scale and genome duplication. Genome Biol *8*, R209.

Hartley, S. (1986). The chromosomes of salmonid fishes. Biol. Rev. *62*, 197–214.

Haug, M.F., Gesemann, M., Mueller, T., and Neuhauss, Stephan C F (2013). Phylogeny and expression divergence of metabotropic glutamate receptor genes in the brain of zebrafish (Danio rerio). J Comp Neurol *521*, 1533–1560.

Henkel, C.V., Burgerhout, E., Wijze, D.L. de, Dirks, R.P., Minegishi, Y., Jansen, H.J., Spaink, H.P., Dufour, S., Weltzien, F.-A., Tsukamoto, K., van den Thillart,Guido E. E. J. M., and Schubert, M. (2012). Primitive Duplicate Hox Clusters in the European Eel's Genome. PLoS ONE *7*, e32231.

Hoegg, S., and Meyer, A. (2007). Phylogenomic analyses of KCNA gene clusters in vertebrates: why do gene clusters stay intact? BMC Evol Biol *7*, 139.

Hoekstra, H.E., and Coyne, J.A. (2007). The locus of evolution: evo devo and the genetics of adaptation. Evolution *61*, 995–1016.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J.C., Koch, R., Rauch, G.-J., White, S., Chow, W., Kilian, B., Quintais, L.T., Guerra-Assuncao, J.A., Zhou, Y., Gu, Y., Yen, J., Vogel, J.-H., Eyre, T., Redmond, S., Banerjee, R., Chi, J., Fu, B., Langley, E., Maguire, S.F., Laird, G.K., Lloyd, D., Kenyon, E., Donaldson, S., Sehra, H., Almeida-King, J., Loveland, J., Trevanion, S., Jones, M., Quail, M., Willey, D., Hunt, A., Burton, J., Sims, S., McLay, K., Plumb, B., Davis, J., Clee, C., Oliver, K., Clark, R., Riddle, C., Eliott, D., Threadgold, G., Harden, G., Ware, D., Mortimer, B., Kerry, G., Heath, P., Phillimore, B., Tracey, A., Corby, N., Dunn, M., Johnson, C., Wood, J., Clark, S., Pelan, S., Griffiths, G., Smith, M., Glithero, R., Howden, P., Barker, N., Stevens, C., Harley, J., Holt, K., Panagiotidis, G., Lovell, J., Beasley, H., Henderson, C., Gordon, D., Auger, K., Wright, D., Collins, J., Raisen, C., Dyer, L., Leung, K., Robertson, L., Ambridge, K., Leongamornlert, D., McGuire, S., Gilderthorp, R., Griffiths, C., Manthravadi, D., Nichol, S., Barker, G., Whitehead, S., Kay, M., Brown, J., Murnane, C., Gray, E., Humphries, M., Sycamore, N., Barker, D., Saunders, D., Wallis, J., Babbage, A., Hammond, S., Mashreghi-Mohammadi, M., Barr, L., Martin, S., Wray, P., Ellington, A., Matthews, N., Ellwood, M., Woodmansey, R., Clark, G., Cooper, J., Tromans, A., Grafham, D., Skuce, C., Pandian, R., Andrews, R., Harrison, E., Kimberley, A., Garnett, J., Fosker, N., Hall, R., Garner, P., Kelly, D., Bird, C., Palmer, S., Gehring, I., Berger, A., Dooley, C.M., Ersan-Urun, Z., Eser, C., Geiger, H., Geisler, M., Karotki, L., Kirn, A., Konantz, J., Konantz, M., Oberlander, M., Rudolph-Geiger, S., Teucke, M.,

Osoegawa, K., Zhu, B., Rapp, A., Widaa, S., Langford, C., Yang, F., Carter, N.P., Harrow, J., Ning, Z., Herrero, J., Searle, Steve M J, Enright, A., Geisler, R., Plasterk, Ronald H A, Lee, C., Westerfield, M., de Jong, Pieter J, Zon, L.I., Postlethwait, J.H., Nusslein-Volhard, C., Hubbard, Tim J P, Roest Crollius, H., Rogers, J., and Stemple, D.L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature *496*, 498–503.

Huang, R., Hippauf, F., Rohrbeck, D., Haustein, M., Wenke, K., Feike, J., Sorrelle, N., Piechulla, B., and Barkman, T.J. (2012). Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. Proc Natl Acad Sci U S A *109*, 2966–2971.

Hufton, A.L., Mathia, S., Braun, H., Georgi, U., Lehrach, H., Vingron, M., Poustka, A.J., and Panopoulou, G. (2009). Deeply conserved chordate noncoding sequences preserve genome synteny but do not drive gene duplicate retention. Genome Res *19*, 2036–2051.

Hughes, A.L. (1994). The evolution of functionally novel proteins after gene duplication. Proc Biol Sci *256*, 119–124.

Innan, H. (2009). Population genetic models of duplicated genes. Genetica *137*, 19–37.

Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet *11*, 4.

Jaillon, O., Aury, J., Brunet, F., Petit, J., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biémont, C., Skalli, Z., Cattolico, L., Poulain, J., Berardinis, V. de, Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K., McEwan, P., Bosak, S., Kellis, M., Volff, J., Guigó, R., Zody, M., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quétier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E., Weissenbach, J., and Roest Crollius, H. (2004). Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature *431*, 946–957.

Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.-F., Weissenbach, J., Quetier, F., and Wincker, P. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature *449*, 463–467.

Jiang, H., and Blouin, C. (2007). Insertions and the emergence of novel protein structure: a structure-based phylogenetic study of insertions. BMC Bioinformatics *8*, 444.

Jimenez-Delgado, S., Pascual-Anaya, J., and Garcia-Fernandez, J. (2009). Implications of duplicated cis-regulatory elements in the evolution of metazoans: the DDI model or how simplicity begets novelty. Brief Funct Genomic Proteomic *8*, 266–275.

Johnson, K.R., Wright, J. E. Jr., and May, B. (1987). Linkage Relationships Reflecting Ancestral Tetraploidy in Salmonid Fish. Genetics *116*, 579–591.

Jovelin, R., He, X., Amores, A., Yan, Y.-L., Shi, R., Qin, B., Roe, B., Cresko, W.A., and Postlethwait, J.H. (2007). Duplication and divergence of fgf8 functions in teleost development and evolution. J Exp Zool B Mol Dev Evol *308*, 730–743.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., Jindo, T., Kobayashi, D., Shimada, A., Toyoda, A., Kuroki, Y., Fujiyama, A., Sasaki, T., Shimizu, A., Asakawa, S., Shimizu, N., Hashimoto, S.-I., Yang, J., Lee, Y., Matsushima, K., Sugano, S., Sakaizumi, M., Narita, T., Ohishi, K., Haga, S., Ohta, F., Nomoto, H., Nogata, K., Morishita, T., Endo, T., Shin-I, T., Takeda, H., Morishita, S., and Kohara, Y. (2007). The medaka draft genome and insights into vertebrate genome evolution. Nature *447*, 714–719.

Kassahn, K.S., Dang, V.T., Wilkins, S.J., Perkins, A.C., and Ragan, M.A. (2009). Evolution of gene function and regulatory control after whole-genome duplication: Comparative analyses in vertebrates. Genome Res *19*, 1404–1418.

Kastenhuber, E., Gesemann, M., Mickoleit, M., and Neuhauss, Stephan C F (2013). Phylogenetic analysis and expression of zebrafish transient receptor potential melastatin family genes. Dev Dyn *242*, 1236–1249.

Kellis, M., Birren, B.W., and Lander, E.S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature *428*, 617–624.

Kohn, M., Hogel, J., Vogel, W., Minich, P., Kehrer-Sawatzki, H., Graves, Jennifer A M, and Hameister, H. (2006). Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. Trends Genet *22*, 203–210.

Kuraku, S., Meyer, A., and Kuratani, S. (2009). Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? Mol Biol Evol *26*, 47–59.

La Fleur, G. J. Jr., Raldua, D., Fabra, M., Carnevali, O., Denslow, N., Wallace, R.A., and Cerda, J. (2005). Derivation of major yolk proteins from parental vitellogenins and alternative processing during oocyte maturation in Fundulus heteroclitus. Biol Reprod *73*, 815–824.

Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. Genetics *154*, 459–473.

Lynch, V.J., and Wagner, G.P. (2008). Resurrecting the role of transcription factor change in developmental evolution. Evolution *62*, 2131–2154.

Mable, B. (2004). 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. Biol J Linn Soc *82*, 453–466.

Macqueen, D.J., and Johnston, I.A. (2008). Evolution of follistatin in teleosts revealed through phylogenetic, genomic and expression analyses. Dev Genes Evol *218*, 1–14.

Macqueen, D.J., and Johnston, I.A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc Biol Sci *281*, 20132881.

Maere, S., Bodt, S. de, Raes, J., Casneuf, T., van Montagu, M., Kuiper, M., and Van de Peer, Yves (2005). Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci U S A *102*, 5454–5459.

Mansai, S.P., and Innan, H. (2010). The power of the methods for detecting interlocus gene conversion. Genetics *184*, 517–527.

Matsubara, T., Ohkubo, N., Andoh, T., Sullivan, C.V., and Hara, A. (1999). Two forms of vitellogenin, yielding two distinct lipovitellins, play different roles during oocyte maturation and early development of barfin flounder, Verasper moseri, a marine teleost that spawns pelagic eggs. Dev Biol *213*, 18–32.

McClintock, J.M., Carlson, R., Mann, D.M., and Prince, V.E. (2001). Consequences of Hox gene duplication in the vertebrates: an investigation of the zebrafish Hox paralogue group 1 genes. Development *128*, 2471–2484.

McFarlane, L., Svingen, T., Braasch, I., Koopman, P., Schartl, M., and Wilhelm, D. (2011). Expansion of the *Ago* gene family in the teleost clade. Dev Genes Evol *221*, 95-104

McGuigan, K., Phillips, P.C., and Postlethwait, J.H. (2004). Evolution of sarcomeric myosin heavy chain genes: evidence from fish. Mol Biol Evol *21*, 1042–1056.

Misof, B.Y., and Wagner, G.P. (1996). Evidence for four Hox clusters in the killifish Fundulus heteroclitus (teleostei). Mol Phylogenet Evol *5*, 309–322.

Mondragon-Palomino, M., and Gaut, B.S. (2005). Gene conversion and the evolution of three leucine-rich repeat gene families in Arabidopsis thaliana. Mol Biol Evol *22*, 2444–2456.

Naruse, K., Tanaka, M., Mita, K., Shima, A., Postlethwait, J., and Mitani, H. (2004). A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. Genome Res *14*, 820–828.

Near, T.J., Eytan, R.I., Dornburg, A., Kuhn, K.L., Moore, J.A., Davis, M.P., Wainwright, P.C., Friedman, M., and Smith, W.L. (2012). Resolution of ray-finned fish phylogeny and timing of diversification. Proc Natl Acad Sci U S A *109*, 13698–13703.

Niederhäusern, V. von, Kastenhuber, E., Stäuble, A., Gesemann, M., and Neuhauss, Stephan C F (2013). Phylogeny and expression of canonical transient receptor potential (TRPC) genes in developing zebrafish. Dev Dyn *242*, 1427–1441.

Njolstad, P.R., Molven, A., Hordvik, I., Apold, J., and Fjose, A. (1988). Primary structure, developmentally regulated expression and potential duplication of the zebrafish homeobox gene ZF-21. Nucleic Acids Res *16*, 9097–9111.

Noonan, J.P., Grimwood, J., Schmutz, J., Dickson, M., and Myers, R.M. (2004). Gene conversion and the evolution of protocadherin gene cluster diversity. Genome Res *14*, 354–366.

Ohno, S. (1970a). Evolution by gene duplication (Springer, New York).

Ohno, S. (1970b). The Enormous Diversity in Genome Sizes of Fish as a Reflection of Nature's Extensive Experiments with Gene Duplication. T Am Fish Soc *99*, 120–130.

Ohno, S., Muramoto, J., Lawrence, C., and Atkin, N.B. (1967). Diploid-tetraploid relationship among old-world members of the fish family Cyprinidae. Chromosoma *23*, 1–9.

Osborn, T.C., Pires, J.C., Birchler, J.A., Auger, D.L., Chen, Z.J., Lee, H.-S., Comai, L., Madlung, A., Doerge, R.W., Colot, V., and Martienssen, R.A. (2003). Understanding mechanisms of novel gene expression in polyploids. Trends Genet *19*, 141–147.

Otto, S.P., and Whitton, J. (2000). Polyploid incidence and evolution. Annu. Rev. Genet. *34*, 401–437.

Papp, B., Pal, C., and Hurst, L.D. (2003). Dosage sensitivity and the evolution of gene families in yeast. Nature *424*, 194–197.

Phillips, R., and Rab, P. (2001). Chromosome evolution in the Salmonidae (Pisces): an update. Biol Rev Camb Philos Soc *76*, 1–25.

Phillips, R.B., Keatley, K.A., Morasch, M.R., Ventura, A.B., Lubieniecki, K.P., Koop, B.F., Danzmann, R.G., and Davidson, W.S. (2009). Assignment of Atlantic salmon (Salmo salar) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (Oncorhynchus mykiss). BMC Genet *10*, 46.

Postlethwait, J., Amores, A., Cresko, W., Singer, A., and Yan, Y.-L. (2004). Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet *20*, 481–490.

Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y.L., Kelly, P.D., Chu, F., Huang, H., Hill-Force, A., and Talbot, W.S. (2000). Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res *10*, 1890–1902.

Prince, Y., Joly, L., Ekker, M., and Ho, R. (1998). Zebrafish hox genes: genomic organization and modified colinear expression patterns in the trunk. Development *125*, 407–420.

Prohaska, S.J., and Stadler, P.F. (2004). The Duplication of the Hox Gene Clusters in Teleost Fishes. Theory Biosci. *23*, 89–110.

Putnam, N.H., Butts, T., Ferrier, David E K, Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., Benito-Gutierrez, E.L., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, Pieter J, Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-

Fraser, M., Fujiyama, A., Holland, L.Z., Holland, Peter W H, Satoh, N., and Rokhsar, D.S. (2008). The amphioxus genome and the evolution of the chordate karyotype. Nature *453*, 1064–1071.

Renninger, S.L., Gesemann, M., and Neuhauss, Stephan C F (2011). Cone arrestin confers cone vision of high temporal resolution in zebrafish larvae. Eur J Neurosci *33*, 658–667.

Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., and Page, D.C. (2003). Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature *423*, 873–876.

Santini, F., Harmon, L.J., Carnevale, G., and Alfaro, M.E. (2009). Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol Biol *9*, 194.

Sato, Y., Hashiguchi, Y., and Nishida, M. (2009). Temporal pattern of loss/persistence of duplicate genes involved in signal transduction and metabolic pathways after teleost-specific genome duplication. BMC Evol Biol *9*, 127.

Scannell, D.R., Butler, G., and Wolfe, K.H. (2007). Yeast genome evolution--the origin of the species. Yeast *24*, 929–942.

Scannell, D.R., Byrne, K.P., Gordon, J.L., Wong, S., and Wolfe, K.H. (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature *440*, 341–345.

Schartl, M., Walter, R.B., Shen, Y., Garcia, T., Catchen, J., Amores, A., Braasch, I., Chalopin, D., Volff, J.-N., Lesch, K.-P., Bisazza, A., Minx, P., Hillier, L., Wilson, R.K., Fuerstenberg, S., Boore, J., Searle, S., Postlethwait, J.H., and Warren, W.C. (2013). The genome of the platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and several complex traits. Nat Genet *45*, 567–572.

Schranz, M.E., Mohammadin, S., and Edger, P.P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Curr Opin Plant Biol *15*, 147–153.

Semon, M., and Wolfe, K.H. (2008). Preferential subfunctionalization of slow-evolving genes after allopolyploidization in Xenopus laevis. Proc Natl Acad Sci U S A *105*, 8333–8338.

Sémon, M., and Wolfe, K.H. (2007). Rearrangement rate following the whole-genome duplication in teleosts. Mol. Biol. Evol. *24*, 860–867.

Semple, C., and Wolfe, K.H. (1999). Gene duplication and gene conversion in the Caenorhabditis elegans genome. J Mol Evol *48*, 555–564.

Seoighe, C., and Wolfe, K.H. (1999). Yeast genome evolution in the post-genome era. Curr Opin Microbiol *2*, 548–554.

Sikosek, T., Chan, H.S., and Bornberg-Bauer, E. (2012). Escape from Adaptive Conflict follows from weak functional trade-offs and mutational robustness. Proc Natl Acad Sci U S A *109*, 14888–14893.

Smith, S.A., Beaulieu, J.M., Stamatakis, A., and Donoghue, M.J. (2011). Understanding angiosperm diversification using small and large phylogenetic trees. Am J Bot *98*, 404–414.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Depamphilis, C.W., Wall, P.K., and Soltis, P.S. (2009). Polyploidy and angiosperm diversification. Am J Bot *96*, 336–348.

Sugino, R.P., and Innan, H. (2006). Selection for more of the same product as a force to enhance concerted evolution of duplicated genes. Trends Genet *22*, 642–644.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., and Paterson, A.H. (2008). Synteny and collinearity in plant genomes. Science *320*, 486–488.

Taylor, J.S. (2003). Genome Duplication, a Trait Shared by 22,000 Species of Ray-Finned Fish. Genome Res *13*, 382–390.

Taylor, J.S., Van de Peer, Y, Braasch, I., and Meyer, A. (2001). Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond, B., Biol Sci *356*, 1661–1679.

Teshima, K.M., and Innan, H. (2008). Neofunctionalization of duplicated genes under the pressure of gene conversion. Genetics *178*, 1385–1398.

Tian, D., Wang, Q., Zhang, P., Araki, H., Yang, S., Kreitman, M., Nagylaki, T., Hudson, R., Bergelson, J., and Chen, J.-Q. (2008). Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. Nature *455*, 105–108.

Uyeno, T., and Smith, G.R. (1972). Tetraploid Origin of the Karyotype of Catostomid Fishes. Science *175*, 644–646.

Van de Peer, Y, Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. Nat Rev Genet *10*, 725–732.

Vandepoele, K., Vos, W. de, Taylor, J.S., Meyer, A., and Van de Peer, Yves (2004). Major events in the genome evolution of vertebrates: Paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci U S A *101*, 1638–1643.

Volff, J.-N. (2005). Genome evolution and biodiversity in teleost fish. Heredity (Edinb) *94*, 280–294.

Wang, J.-T., Li, J.-T., Zhang, X.-F., and Sun, X.-W. (2012). Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (Cyprinus carpio). BMC Genomics *13*, 96.

Weadick, C.J., and Chang, B.S. (2012). Complex patterns of divergence among green-sensitive (RH2a) African cichlid opsins revealed by Clade model analyses. BMC Evol Biol *12*, 206.

Windsor, D.J., and Owens, G.L. (2009). The opsin repertoire of Jenynsia onca: a new perspective on gene duplication and divergence in livebearers. BMC Res Notes *2*, 159.

Wittbrodt, J., Meyer, A., and Schartl, M. (1998). More genes in fish? Bioessays *20*, 511–515.

Wittkopp, P.J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet *13*, 59–69.

Wolfe, K.H., and Shields, D.C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. Nature *387*, 708–713.

Woods, I.G., Wilson, C., Friedlander, B., Chang, P., Reyes, D.K., Nix, R., Kelly, P.D., Chu, F., Postlethwait, J.H., and Talbot, W.S. (2005). The zebrafish gene map defines ancestral vertebrate chromosomes. Genome Res *15*, 1307–1314.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat Rev Genet *8*, 206–216.

Yu, W.-P., Yew, K., Rajasegaran, V., and Venkatesh, B. (2007). Sequencing and comparative analysis of fugu protocadherin clusters reveal diversity of protocadherin genes among teleosts. BMC Evol Biol *7*, 49.

Zakon, H.H., Lu, Y., Zwickl, D.J., and Hillis, D.M. (2006). Sodium channel genes and the evolution of diversity in communication signals of electric fishes: convergent molecular evolution. Proc Natl Acad Sci U S A *103*, 3675–3680.

Zhang, X., Zhang, Y., Zheng, X., Kuang, Y., Zhao, Z., Zhao, L., Li, C., Jiang, L., Cao, D., Lu, C., Xu, P., and Sun, X. (2013). A Consensus Linkage Map Provides Insights on Genome Character and Evolution in Common Carp (Cyprinus carpio L.). Mar Biotechnol *15*, 275–312.

Zhou, R., Moshgabadi, N., and Adams, K.L. (2011). Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. Proc Natl Acad Sci U S A *108*, 16122–16127.

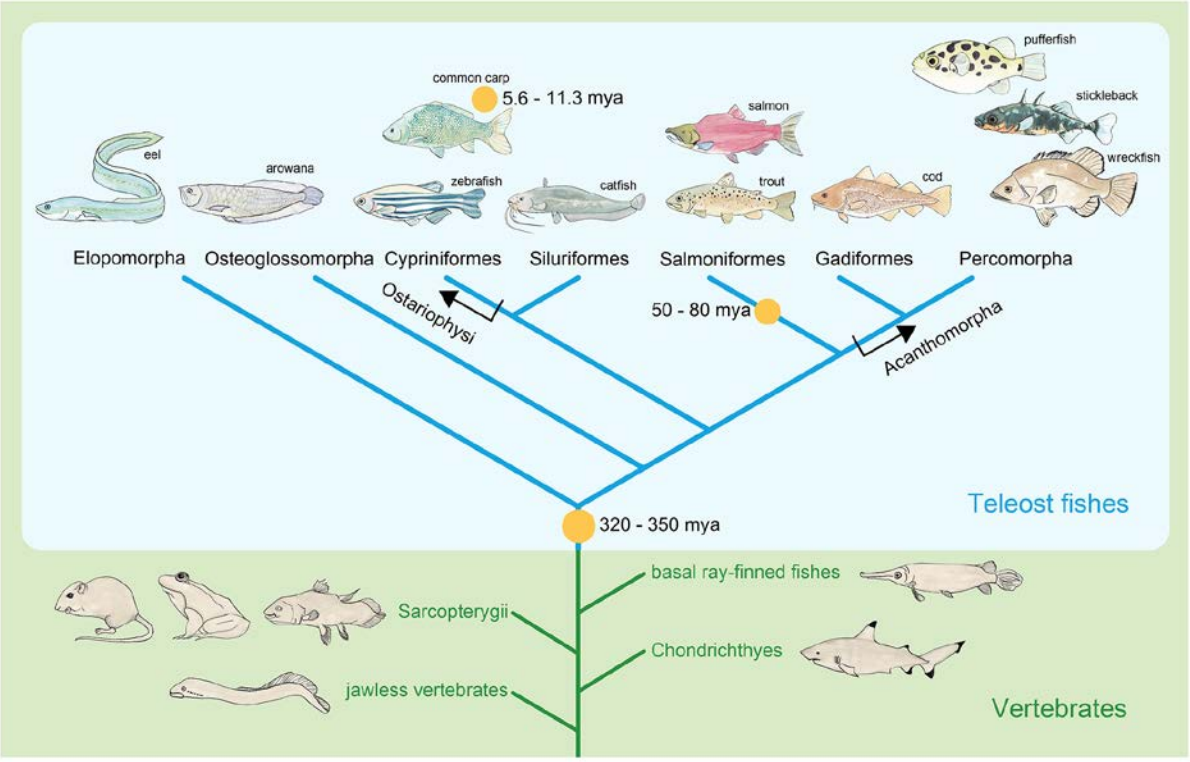
Fig.1 Simplified phylogeny of teleost fishes. The teleost lineage split from basal ray-finned fishes and started to diverge after a WGD event that took place 320-350 mya. Additional WGDs occured at the base of Salmoniformes 50-80 mya and in a closely related ancestor of the common carp less than 16 mya. For the sake of clarity, only a selection of teleost taxa is presented. Orange circles depict WGD events within teleost evolution. WGD events outside the teleosts are not shown. Mya, million years ago.

Fig.2 Fates of duplicated genes after WGD. A WGD event results in the formation of two identical duplicates of every gene. Duplicate genes can undergo different fates. Non-functionalization: Deleterious mutations occur in one of the duplicates, eventually leading to loss of expression (pseudogenization). Mutations continue to accumulate until the structural features of the gene have disappeared completely. Subfunctionalization: Complementary degenerative mutations in paralogous genes lead to preservation of both duplicates. Neofunctionalization: One of the genes acquires a novel function. Dosage selection: Dosage-sensitive genes remain basically unchanged after WGD. Although the initial mutations are depicted in regulatory regions, also changes in coding sequence can lead to the different scenarios. For sake of simplicity, introns were omitted, and regulatory regions are depicted only 5' of the transcription start site.

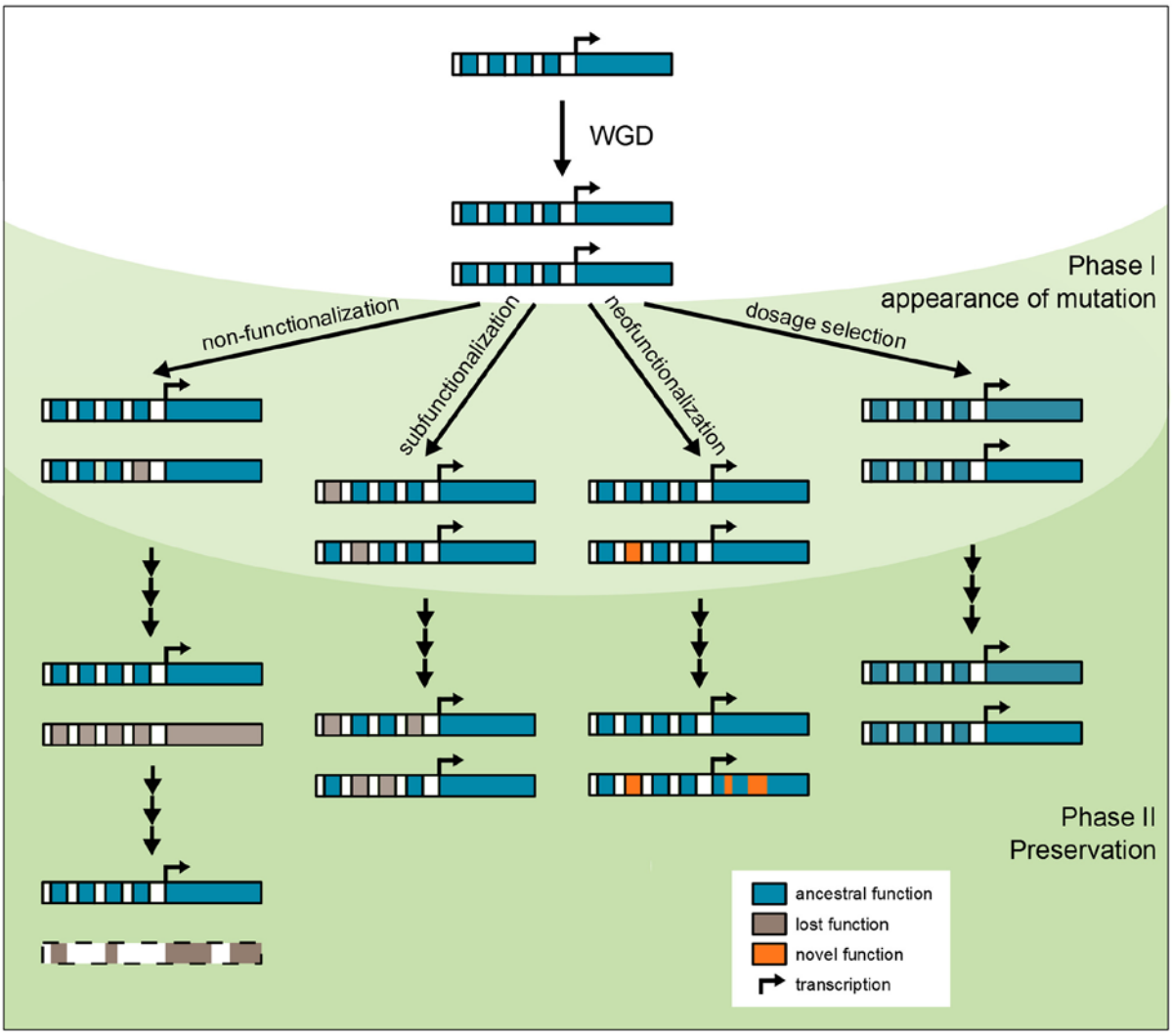
Supp. Fig 1 Chromosome number distribution of salmonids. Haploid chromosome numbers are shown. Data is obtained from the Animal genome size database (http://www.genomesize.com). All 27 salmonid species' chromosome numbers available in the database were analyzed.

Figure

1


pufferfish
stickleback
wreckfish
common carp
5.6 - 11.3 mya
salmon
eel
arowana
zebrafish
catfish
trout
cod
Elopomorpha
Osteoglossomorpha
Cypriniformes
Siluriformes
Salmoniformes
Gadiformes
Percomorpha
Ostariophysi
50 - 80 mya
Acanthomorpha
Teleost fishes
320 - 350 mya
basal ray-finned fishes
Sarcopterygii
Chondrichthyes
jawless vertebrates
Vertebrates

Figure


WGD
Phase I
appearance of mutation
non-functionalization
subfunctionalization
neofunctionalization
dosage selection
Phase II
Preservation
ancestral function
lost function
novel function
transcription

Supplementary Figure 1