



Zebrafish: unraveling genetic complexity through duplicated genes

Maliha Tasnim¹ · Preston Wahlquist¹ · Jonathon T. Hill¹

Received: 21 February 2024 / Accepted: 19 July 2024 / Published online: 30 July 2024
© The Author(s) 2024

Abstract

The zebrafish is an invaluable model organism for genetic, developmental, and disease research. Although its high conservation with humans is often cited as justification for its use, the zebrafish harbors oft-ignored genetic characteristics that may provide unique insights into gene structure and function. Zebrafish, along with other teleost fish, underwent an additional round of whole genome duplication after their split from tetrapods—resulting in an abundance of duplicated genes when compared to other vertebrates. These duplicated genes have evolved in distinct ways over the ensuing 350 million years. Thus, each gene within a duplicated gene pair has nuanced differences that create a unique identity. By investigating both members of the gene pair together, we can elucidate the mechanisms that underly protein structure and function and drive the complex interplay within biological systems, such as signal transduction cascades, genetic regulatory networks, and evolution of tissue and organ function. It is crucial to leverage such studies to explore these molecular dynamics, which could have far-reaching implications for both basic science and therapeutic development. Here, we will review the role of gene duplications and the existing models for gene divergence and retention following these events. We will also highlight examples within each of these models where studies comparing duplicated genes in the zebrafish have yielded key insights into protein structure, function, and regulation.

Keywords Gene duplication · Neofunctionalization · Subfunctionalization · Non-functionalization · Zebrafish · Genetic complexity · Protein-encoding genes

Introduction

Zebrafish are a valuable model for human disease studies due to the high degree of genetic conservation between the two species—approximately 70% of human protein-coding genes, including the majority of genes implicated in disease, have highly conserved orthologs in zebrafish (Adhish & Manjubala 2023; Barut & Zon 2000; Goldsmith & Jobin 2012; Howe et al. 2013; Santoriello & Zon 2012; Shehwana & Konu 2019). Genetic manipulation methods in the zebrafish, such as CRISPR mutagenesis (Hwang et al. 2013) and transgenic line generation (Streisinger et al. 1981), are also well-established and increasingly complemented by targeted gene insertion methods (Ata et al. 2018; Auer et al. 2014; Ranawakage et al.

2021). As a result, the zebrafish is one of the most commonly used model organisms and has been employed in studies on tissue and organ development, a wide range of human diseases, and drug discovery efforts (Chia et al. 2022; Choi et al. 2021; Patton et al. 2021; Teame et al. 2019). These studies have profoundly enhanced our knowledge of genetics, development, and disease, making zebrafish indispensable in biomedical research now and for the foreseeable future.

Despite the benefits of zebrafish's genetic parallels with humans, the species' greatest value as a model organism may come from a distinctive aspect of its genome—the remarkable abundance of duplicate genes within its genome. Most vertebrates share two rounds of genome duplication that occurred shortly after their split from invertebrates. These events, often referred to as the first (1R) and second (2R) rounds of genome duplication, are thought to have taken place approximately 500 to 600 million years ago, before the radiation of the extant vertebrate classes (Ohno 1970). In addition to the two vertebrate-wide genome duplication events, the teleost clade (which includes zebrafish) has undergone a third round of genome duplication known as the teleost-specific whole

Communicated by Kristen Panfilio.

✉ Jonathon T. Hill
jhill@byu.edu

¹ Department of Cell Biology and Physiology, Brigham Young University, 701 E. University Pkwy, Provo, UT 84602, USA

genome duplication (TS-3R WGD). This event is thought to have contributed to the considerable expansion and diversification of their genetic repertoire, resulting in an explosive evolutionary radiation that rapidly created the largest and most diverse vertebrate clade (Inoue et al. 2015; Postlethwait et al. 2000). This event is supported by multiple lines of evidence, including a detailed study of genomic architecture that showed that 3440 gene pairs—26% of the genes analyzed—exist within double-conserved synteny (DCS) blocks (Howe et al. 2013). In zebrafish, these duplications are supplemented by an apparent propensity for gene duplication. One study determined that zebrafish had the highest rate of tandem (duplicates located within 10 kb of each other) and intrachromosomal (copies on the same chromosome but more than 10 kb apart from one another) duplicates of the four teleost species studied (Lu et al. 2012). Together, whole genome and local duplication events have resulted in duplicates for approximately 5300 of the 26,206 protein-coding genes identified in the zebrafish (Howe et al. 2013).

Although considered a nuisance by many (Doyle & Croll 2022; Otis et al. 2015; Vaz et al. 2019), this rich reservoir of duplicated genes offers an abundant collection of genetic data for researchers. By examining duplicated genes, it is possible to gain a greater understanding of the evolution of gene domains, the diversification of protein functions, and the complex mechanisms of gene regulation that have developed over the course of millennia. Unfortunately, researchers often focus solely on one copy of a duplicated gene, perhaps the one with the most homology to its human ortholog or the one that results in a stronger phenotype. For instance, in the case of the red-sensitive opsin genes, studies on the *lws-1* gene have been emphasized, while *lws-2* has received less attention (Xu et al. 2024). Similarly, certain transducin gene duplicates have distinct roles in vision and circadian rhythms, but those with more subtle functions linked to circadian regulation in the pineal complex are less studied compared to their counterparts involved in vision (Lagman et al. 2015). Finally, among the *cyp26* paralogs involved in retinoic acid metabolism, *cyp26a1* is more thoroughly investigated than *cyp26b1* and *cyp26c1* (Rodríguez-Marí et al. 2013). By focusing on a single paralog, we miss out on a 350-million-year natural experiment that would be difficult to reproduce in the lab.

Regardless of the disease being modeled, understanding the evolutionary history of genes duplicated in zebrafish and modeling their divergence over time is essential for creating meaningful and informative experiments. While this may require some effort, the information gleaned from these studies can inform the researcher on what insights into gene regulation and protein function can be obtained while also determining what experiments should be conducted and how their outcomes can be interpreted. Here, we will review the role of gene duplications in evolution and the existing

models for gene divergence and retention. We will also demonstrate how studies that incorporate careful analyses of the post-duplication divergence of zebrafish paralogs have yielded key insights into their protein structure, function, and regulation.

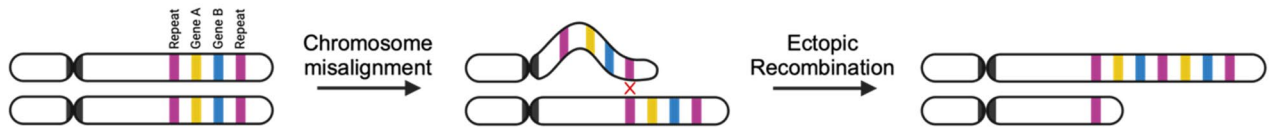
Gene duplication in evolutionary history

Gene duplication is one of the most important mechanisms driving adaptive radiation and evolutionary innovation (Arnegard et al. 2010; Crow & Wagner 2006; Kondrashov 2012; Roth et al. 2007; A. Wagner 2008). In 1970, Ohno first proposed that gene duplication is an essential mechanism driving the creation of novelty through evolution (Ohno 1970). In the context of a single-copy gene, evolutionary forces impose stringent selection pressures to maintain the integrity and functionality of the gene product. However, gene duplication results in a redundant copy, which allows for the evolution of novel functions, regulatory mechanisms, or adaptations to changing environments in the absence of stabilizing selective pressures (Force et al. 1999; Lynch & Force 2000; Magadum et al. 2013; Rochette et al. 2001). Over time, these mutations can lead to advantageous novel gene functions or increased gene specialization, leading to their long-term retention in the genome. Ohno suggested that these events occur frequently and randomly, providing a substrate for evolutionary innovation (Ohno 1970).

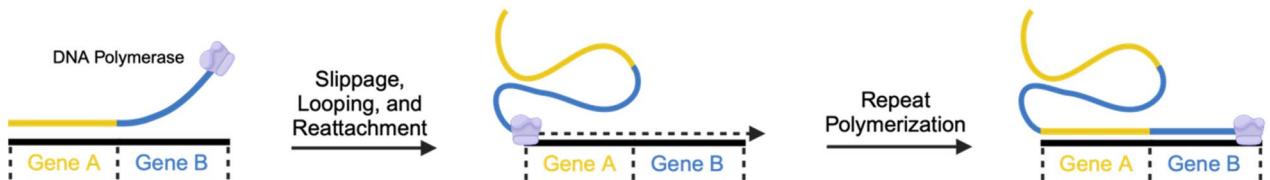
Studies across many taxa, including plants, animals, and bacteria, have corroborated Ohno's theory that speciation events often stem from gene duplications (Anatskaya & Vinogradov 2022; Crow & Wagner 2006; Roth et al. 2007; Singh & Krumlauf 2022). For example, gene duplication is common in plants, where multiple species have undergone genome duplication several times throughout history, and many species are very tolerant of polyploidization in agricultural breeding programs (del Pozo & Ramirez-Parra 2015; Panchy et al. 2016; Qiao et al. 2019). In animals, gene duplication events are less common and usually occur on a small scale, including ectopic recombination (Christiaens et al. 2012) (Fig. 1a), replication slippage (Viguera et al. 2001) (Fig. 1b), and retrotransposition (Huang et al. 2010) (Fig. 1c), but they can also involve larger genomic regions via aneuploidy (Koo et al. 2018) (Fig. 1d) or polyploidy (Blanc & Wolfe 2004) (Fig. 1e). The events with the most impact on evolution have been large-scale duplications, including whole genome duplication (WGD) events (Fig. 2), which have been associated with adaptive radiation in multiple clades (Meyer & Schartl 1999).

Vertebrates appear to share two rounds of WGD. The first (1R) may predate the Cambrian explosion (Meyer & Schartl 1999), and the second (2R) happened at the base of the vertebrates or after the agnathan/gnathostome split (Holland & Ocampo Daza 2018) (Fig. 2). The 1R

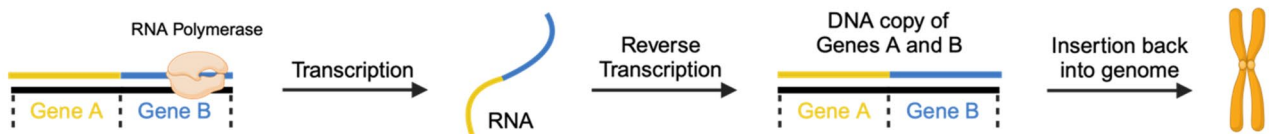
a) Ectopic Recombination



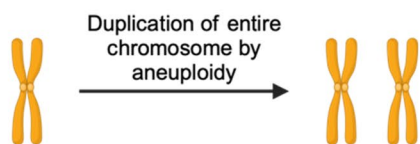
b) Replication Slippage



c) RNA Retrotransposition



d) Aneuploidy



e) Polyploidy

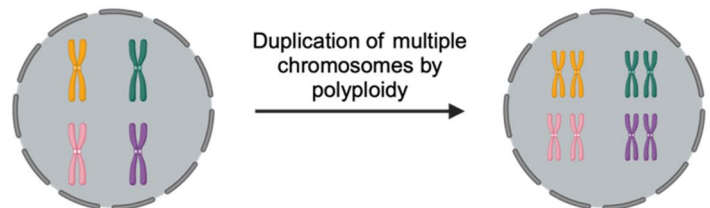


Fig. 1 The process of gene duplication can occur through various mechanisms: **a** ectopic recombination: repeat regions (purple segments) near genes (represented by yellow and blue segments) undergo recombination following misalignment of the repeats, leading to the exchange of genetic material and the creation of duplicated genes on one of the sister chromatids. **b** Replication slippage: a DNA polymerase temporarily detaches and then reattaches to the template DNA strand. The genes will become duplicated if the DNA polymer-

ase realigns to an area before the yellow and blue genes are located. **c** RNA retrotransposition: genes may be transcribed and then reverse-transcribed into double-stranded DNA and inserted back into the genome. **d** Aneuploidy: entire duplication of a particular chromosome due to non-disjunction errors. **e** Polyploidy: non-disjunction during meiosis results in a failure to create haploid cells, resulting in multiple copies of the genome in the offspring. Created with [BioRender.com](#)

duplication likely provided the raw genetic material that led to the development of several fundamental vertebrate features, including the complex organ systems and sophisticated neural circuitry characteristic of the phylum (Dehal & Boore 2005). The subsequent 2R duplication is believed to have resulted in further diversification of gene functions, giving rise to the vast array of vertebrate species seen today by enabling the specialization of physiological processes and morphological structures (Panopoulou & Poustka 2005). Additional WGD events have occurred more recently in vertebrate evolution, including TS-3R WGD approximately 350 million years ago, which resulted

in rapid and extensive radiation of the ray-finned fishes—one of the largest vertebrate clades today.

This event is responsible for most of the genes we consider to be duplicated in the zebrafish genome. Several pieces of evidence support the TS-3R WGD (Glasauer & Neuhaus 2014). For example, many Teleostei, such as medaka, puffer fish, and zebrafish, have seven Hox clusters (Crow et al. 2006). In contrast, the Sarcopterygii, or lobed-finned fishes, and most other vertebrates have only 4 clusters of Hox genes (Ozernyuk & Schepetov 2022). There is also considerable genomic structure and comparative analysis data supporting a TS-3R WGD, including conserved synteny, zebrafish

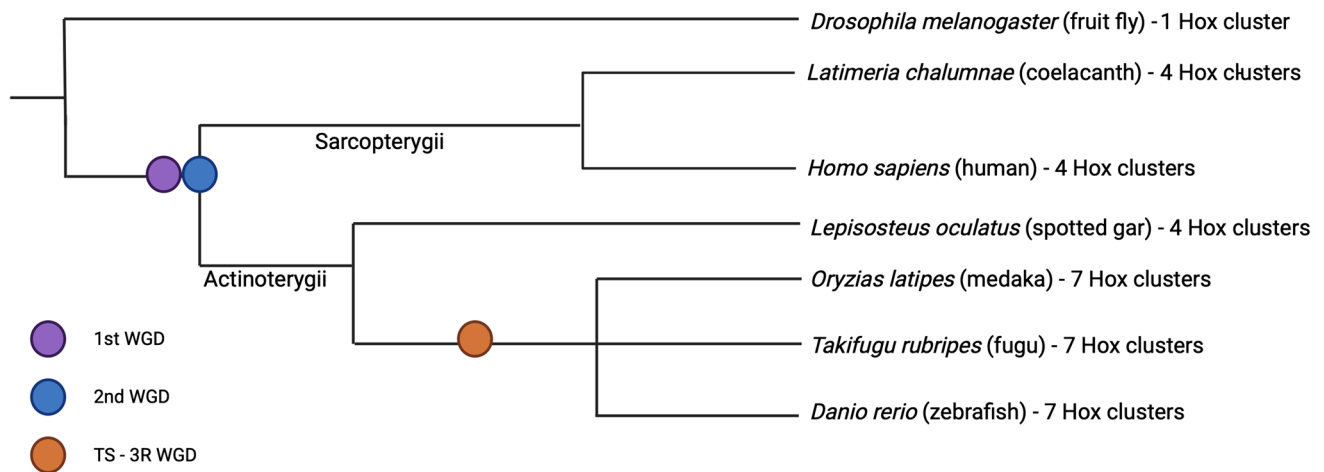


Fig. 2 Key WGD events in vertebrate evolution. There have been at least three whole genome duplication (WGD) events in the teleost clade. The first WGD event occurred around 500 million years ago (mya) and possibly preceded the Cambrian explosion. The second WGD occurred at the base of the vertebrates or after the agnathan/gnathostome split (humans and coelacanths are shown as examples).

The teleost-specific whole-genome duplication (TS-3R WGD) event occurred approximately 350 mya and led to the enormous diversity and adaptability of the teleost fish group (medaka, puffer, and zebrafish are only a few examples of this class). Created with [BioRender.com](#)

gene-mapping studies, and phylogenetic analyses of teleost genes, among others (Glasauer & Neuhauss 2014; Molinari et al. 2008; Taylor et al. 2001; Vandepoele et al. 2004).

Gene duplication, regardless of its scale or mechanism, results in one of two initial outcomes: the loss of one copy or the persistence of both. Loss of the gene, termed non-functionalization, is the most common result of gene duplication. This occurs when random mutations in key cis-regulatory or coding regions cause the gene to degenerate into a pseudogene (Evans-Yamamoto et al. 2023; Force et al. 1999; Moriyama & Koshiba-Takeuchi 2018; Rastogi & Liberles 2005). When duplicates persist, they typically exhibit one or more types of functional divergence. Subfunctionalization can split the original functions between the copies, or neofunctionalization can give rise to new functions (Birchler & Yang 2022; Force et al. 1999; Lynch & Force 2000; Qiu et al. 2020). Additionally, certain genes, like those coding for ribosomal RNA, may not undergo functional changes but are retained simply because having more copies is beneficial for the organism (Copley 2020; Hakes et al. 2007; Hallin & Landry 2019; Kuzmin et al. 2022; Xu et al. 2023). Gene families with broad expression profiles, especially those involved in developmental pathways, can also evolve after duplication to have specialized domains (Marlétaz et al. 2018). This specialization leads to more complex regulatory landscapes, particularly in neural tissues. This increase in regulatory complexity supports the idea that the retention of certain genes is advantageous because having more copies allows for greater specialization, providing significant developmental and evolutionary benefits.

Determining the evolutionary fate of gene duplicates is a dynamic interplay between hypothesis generation and experimentation. However, first determining whether a gene pair exhibits shared, split, or entirely new functions will guide what experiments to run. For instance, if both genes remain active, it might be useful to explore how their functions differ. Or, if one gene has a new role, one can look for unique traits or behaviors it supports. The experimental findings can then confirm or refute the initial hypotheses about the duplicated genes.

The study of the *elastin b* (*elnb*) gene in zebrafish provides a clear illustration of this principle (Moriyama et al. 2016). Researchers discovered that *elnb*, which was created during the TS-3R WGD, neofunctionalized and became essential for the proper development of the bulbus arteriosus (BA) by regulating the differentiation of cardiac precursor cells into smooth muscle. Rescue experiments with *elnb* mRNA mitigated the knockdown phenotype. However, when attempting to rescue the *elnb* morphant phenotype with the ancestral *elastin* gene from *Polypterus senegalus*, a basal actinopterygian and non-teleost fish considered to possess an ancestral-like gene due to its evolutionary position, the experiments did not yield the same restorative effects, implying a novel function for the duplicated *elnb* gene in zebrafish BA development. This process elucidates the creation of an experimental feedback loop where initial experimental data shapes subsequent hypotheses and directs future research, deepening our comprehension of how genetic evolution influences an organism's adaptability and development. Below, we will examine examples of studies conducted on genes exhibiting each of the major post-duplication

models and the insights gained from each (Table 1). These selected studies highlight key aspects of gene duplication research, demonstrating how such studies contribute to our understanding of genetic diversity and functionality across species.

Models for post-duplication evolution

Non-functionalization

The most common outcome following a gene duplication event is the formation of a pseudogene. While duplication relaxes selective pressure and allows for the rapid exploration of the evolutionary space, this often simply leads to the accumulation of deleterious mutations—leading to a loss of gene function (Fig. 3: non-functionalization). Pseudogenes are sequences of DNA that resemble functional genes but have lost their gene expression or protein-coding ability due to these mutations. In many cases, this results in a return to the ancestral state, but non-functionalization can also lead to novel traits. The teleost globin superfamily is a striking depiction of the evolutionary innovation from non-functionalization (Hoffmann et al. 2021). Multiple functional globin isoforms have accommodated unique environmental and developmental challenges in many niches among the teleost species (Storz et al. 2020). These specialized isoforms appear to have arisen from repeated rounds of tandem duplication, independent evolution of the two copies, and non-functionalization of the less fit duplicate (Opazo et al. 2013; Storz et al. 2013; Tiedke et al. 2011). For example, the η -globin gene is a pseudogene in all primates. This classification is based on sequence analyses that show accumulated mutations like frameshifts and premature stop codons, which disrupt the gene's ability to produce a functional protein. Similarly, the δ -globin gene is often found to be a pseudogene in many eutherian species due to gene conversion events where sequences from the β -globin gene are copied into the δ -globin gene locus (Hardison 2012). Although these conversions result in non-functional genes, the evolutionary interval where duplicated genes retain their original function before any significant divergence or loss occurs, referred to as lag, allows for the duplicated genes to be subjected to genetic drift and varying selective pressures (Lynch & Conery 2000). Thus, due to the lag between duplication and non-functionalization, traits can evolve under conditions with lower selective pressure and then be tested for fitness via the non-functionalization of one of the duplicates (Conant & Wolfe 2008; Innan & Kondrashov 2010; Zhang 2003).

Even when pseudogenization occurs, these former duplicated copies are worth exploring. Once thought to be molecular “fossils” with no function, pseudogenes play a

role in many cellular processes by expressing small interfering RNA, competitive endogenous RNA, or antisense transcripts (Groen et al. 2014). Thus, they are involved in complex genetic regulation both at the transcriptional and post-transcriptional levels. For example, they play a critical role in cancer progression, particularly in kidney cancer, and display tissue-specific expression patterns, indicating unique regulatory roles in different tissues (Nakamura-García & Espinal-Enríquez 2023; Qian et al. 2022; Welch et al. 2015). Disease-related changes in cellular processes can also lead to the reactivation or altered expression of pseudogenes (Pink et al. 2011), providing the potential for disease subtyping and personalized treatment decisions (Chen et al., 2020). Specific pseudogene expression changes are often highly specific to certain diseases or disease subtypes, contributing to high diagnostic accuracy and predictive power (Glenfield & McLysaght 2018; Y. Ma et al. 2021; Roberts & Morris 2013).

One key aspect to consider when studying pseudogenes is their cis-regulatory elements. Highly conserved cis-regulatory modules (CRMs) tend to be preserved alongside gene copies that retain their functional roles. In contrast, when a gene duplication results in a non-functional copy or a pseudogene, we often see a concomitant degradation of associated CRMs. Thus, the presence and conservation of CRMs in the vicinity of pseudogenes may indicate that these genetic elements are actively involved in regulating nearby functional genes within the same regulatory network (Kikuta et al. 2007). This suggests that the pseudogenes themselves may not be the direct targets but rather bystanders, with the CRMs playing a crucial role in the regulation of other genes. On the other hand, degradation of CRMs suggests a diminished role in the organism's current genetic regulation and evolutionary progression.

Subfunctionalization

Subfunctionalization, where the ancestral gene expression pattern or domain functions are partitioned between the duplicates, is the most common fate that results in the long-term preservation of functional paralogs in eukaryotic genomes (Lynch & Force 2000). Several cases in zebrafish have been reported where each paralog in a duplicate gene pair performs a subset of the functions performed by a single gene in tetrapods.

The most well-studied mechanism for this division of labor is expression pattern divergence (Fig. 3: expression level subfunctionalization). This commonly occurs by changing spatial expression domains, exemplified by *OXR1* (oxidation resistance gene 1) (*oxr1a* and *oxr1b* in zebrafish) and *PAX6* (Paired box protein Pax-6) (*pax6a* and *pax6b* in zebrafish). In mammals, *OXR1* is highly expressed in the brain, particularly in regions such as the cerebral cortex,

Table 1 Comparative analysis of duplicate genes. Similarity, expression patterns, and functional divergence

Duplicate genes	Human ortholog	Amino acid similarity	Overlapping expression	Differential tissue distribution	Origin of the paralogs	Functional divergence	Compensatory expression	Reference
<i>elna</i> and <i>elnb</i>	<i>ELN</i>	–	Y	Y	TS-3R WGD	Y	N	(Moriyama et al. 2016)
<i>Sox9a</i> and <i>sox9b</i>	<i>SOX9</i>	–	Y	Y	TS-3R WGD	Y	N	(Yan et al. 2005)
<i>oxr1a</i> and <i>oxr1b</i>	<i>OXR1</i>	86%	Y	Y	Segmental sequence duplication	Y	N	(Xu et al. 2020, 2021)
<i>Pax6a</i> and <i>pax6b</i>	<i>PAX6</i>	95%	Y	Y	TS-3R WGD	Y	N	(Kleinjan et al. 2008; Thumel et al. 2010)
<i>nrf2a</i> and <i>nrf2b</i>	<i>NRF2</i>	25%	Y	Y	TS-3R WGD	Y	N	(Sant et al. 2017; Timme-Laragy et al. 2012)
<i>tbx5a</i> and <i>tbx5b</i>	<i>TBX5</i>	83%	Y	N	TS-3R WGD	Y	N	(Anderson et al. 2022; Parrie et al. 2013)
<i>znf143a</i> and <i>znf143b</i>	<i>ZNF143</i>	65%	Y	N	TS-3R WGD	N	Y	(Huning & Kunkel 2020)
<i>Scn11aa</i> and <i>scn11ab</i>	<i>SCN1A</i>	67%	N	Y	–	Y	N	(Weuring et al. 2022)
<i>igf1a</i> , <i>igf1b</i> , and <i>igf2a</i> , <i>igf2b</i>	<i>IGF-1</i> and <i>IGF-2</i>	50% and 70%	Y	Y	TS-3R WGD	Y	N	(Zou et al. 2009)
<i>hoxa13a</i> and <i>hoxa13b</i>	<i>HOXA13</i>	–	Y	Y	TS-3R WGD	Y	–	(Crow et al. 2009)
<i>otos</i> and <i>otospiralin</i> like	<i>OTOS</i>	28%	Y	Y	–	Y	N	(Baanannou et al. 2020)
<i>dscama</i> and <i>dscamb</i>	<i>DSCAM</i>	83%	Y	Y	TS-3R WGD	–	–	(Galicia et al. 2018)
<i>rbp7a</i> and <i>rbp7b</i>	<i>RBP7</i>	80%	Y	Y	TS-3R WGD	–	–	(Belliveau et al. 2010)
<i>foxl2a</i> and <i>foxl2b</i>	<i>FOXL2</i>	64%	Y	Y	–	Y	–	(Yang et al. 2017)
<i>en2a</i> and <i>en2b</i>	<i>EN2</i>	79%	Y	Y	TS-3R WGD	Y	–	(Scholpp & Brand 2001)
<i>rh1</i> and <i>rh1-20/0/0000 0:00:00 AM</i>	<i>RH1</i>	75%	Y	Y	Zebrafish-specific duplication	–	–	(Morrow et al. 2011)
<i>dmrt2a</i> and <i>dmrt2b</i>	<i>DMRT2</i>	–	Y	Y	TS-3R WGD	Y	–	(Liu et al. 2009)
<i>atf5a</i> and <i>atf5b</i>	<i>AFT5</i>	40%	Y	Y	TS-3R WGD	–	N	(Rodríguez-Morales et al. 2020; Zhu et al. 2022)
<i>igfbp-1a</i> and <i>igfbp-1b</i>	<i>IGFBP1</i>	50%	Y	Y	TS-3R WGD	N	–	(Kamei et al. 2008)
<i>cryaa</i> and <i>cryabb</i>	<i>CRYAB</i>	50%	N	Y	TS-3R WGD	Y	–	(Smith et al. 2006)
<i>crabp1a</i> and <i>crabp1b</i>	<i>CRABP1</i>	88%	N	Y	TS-3R WGD	N	–	(R.-Z. Liu et al. 2005)
<i>atxn1a</i> and <i>atxn1b</i>	<i>ATXN1</i>	35%	Y	Y	TS-3R WGD	–	–	(Vauti et al. 2021)

Table 1 (continued)

Duplicate genes	Human ortholog	Amino acid similarity	Overlapping expression	Differential tissue distribution	Origin of the paralogs	Functional divergence	Compensatory expression	Reference
<i>dmbx1a</i> and <i>dmbx1b</i>	<i>DMBX1</i>	72%	Y	Y	TS-3R WGD	Y	N	(Chang et al. 2006; Wong et al. 2010)
<i>inpp5ka</i> and <i>inpp5kb</i>	<i>INPP5K</i>	56%	Y	Y	TS-3R WGD	Y	–	(Shukla et al. 2022)
<i>stat5.1</i> and <i>stat5.2</i>	<i>STAT5</i>	–	Y	–	Zebrafish specific duplication	N	N	(Lewis & Ward 2004)
<i>sox11a</i> and <i>sox11b</i>	<i>SOX11</i>	75%	Y	Y	WGD	N	N	(de Martino et al. 2000)
<i>mdka</i> and <i>mdkb</i>	<i>MDK</i>	72%	Y	Y	Segmental sequence duplication	Y	–	(Winkler et al. 2003)

A detailed comparison of various duplicated genes in relation to their human orthologs, amino acid similarity, expression patterns, tissue distribution, origin, functional divergence, compensatory expression, and literature references. All genes included in this table are derived from studies conducted on zebrafish. **Duplicate genes:** list pairs or groups of duplicated genes studied. **Human ortholog:** the corresponding human gene equivalent, if applicable. **Amino acid similarity:** percentage similarity between the amino acids of the duplicated genes in zebrafish compared to each other. **Expression:** indicates whether the genes share expression patterns in the same tissues. **Differential tissue distribution:** indicates if the genes are expressed in different tissues. **Origin of the paralogs:** describes the evolutionary origin of the gene duplication, such as the teleost-specific third round whole genome duplication (TS-3R WGD) or segmental sequence duplication (SSD). **Functional divergence:** indicates whether the duplicated genes have evolved to perform different functions post-duplication. **Compensatory expression:** indicates whether one gene compensates for the loss or reduction in function of its duplicate. **Reference:** citations of the studies or reviews from which the data were derived. A dash (“–”) indicates that this information was not provided.

Y yes, N no.

hippocampus, and cerebellum, reflecting its role in neuroprotection. It is also expressed in other tissues, including the heart, liver, and kidney, which aligns with its functions in antioxidant defenses and cellular stress responses (Oliver et al. 2011; Volkert & Crowley 2020). In zebrafish, *oxr1a* is maternally expressed and ubiquitous from the two-cell to the sphere stage, later localizing to the head, including the brain, olfactory bulbs, retina, and neurons by 30 h post-fertilization (hpf). In adulthood, *oxr1a* is highly expressed in the brain, eye, and testis, indicating roles in development, reproduction, and antioxidant defenses. *oxr1b* is strongly expressed during the cleavage period and is restricted to the head, specifically in the lateral line ganglia and spinal cord neurons during the pharyngula period (24–48 hpf). By 48–72 hpf, *oxr1b* is also present in the olfactory bulbs, eye, adenohypophysis, and neuromasts. In adulthood, *oxr1b* has higher expression levels compared to *oxr1a* in most tissues. Although both genes are expressed in the brain, olfactory bulbs, and eye, they exhibit distinct spatial domains: *oxr1a* is prominent in the retina and heart, while *oxr1b* is found in the adenohypophysis and neuromasts (Xu et al. 2020, 2021). These patterns indicate subfunctionalization in the zebrafish model, reflecting a division of the expression domain of human.

Another well-studied pair, *pax6a* and *pax6b*, is a noteworthy example of spatial subfunctionalization in the brain

and pancreas. Studying the non-coding regions that control them has led to the identification of distinct enhancer elements that regulate different aspects of *Pax6* expression (Delporte et al. 2008; Kleinjan et al. 2008; Thummel et al. 2010). Enhancers that drive expression in the pancreas versus the brain are different, and these can be pinpointed by looking at where and when each paralog is expressed and then comparing their respective cis-regulatory elements with pre-duplication outgroups. This knowledge can inform the understanding of human diseases related to *PAX6*, such as diabetes mellitus, which affects the pancreas, and aniridia, which presents with neurological issues (Gosmain et al. 2012; Landsend et al. 2021; Lim et al. 2017; Moosajee et al. 1993; Sekiya et al. 2022; Uttley et al. 2023). Furthermore, it may provide novel insights into the development of therapeutic approaches that target these enhancers to modulate *PAX6* expression in these and other disease conditions.

In addition to spatial divisions of labor, expression-based subfunctionalization can also act on temporal expression patterns. One such example is *NRF2* (nuclear factor erythroid 2-related factor 2). The distinct temporal expression patterns of the *nrf2a* and *nrf2b* paralogs are a critical factor in the zebrafish’s response to oxidative stress. Morpholino-induced knockdown experiments showed that while both *nrf2a* and *nrf2b* influence the glutathione redox state, their impact is not uniform throughout development. Notably,

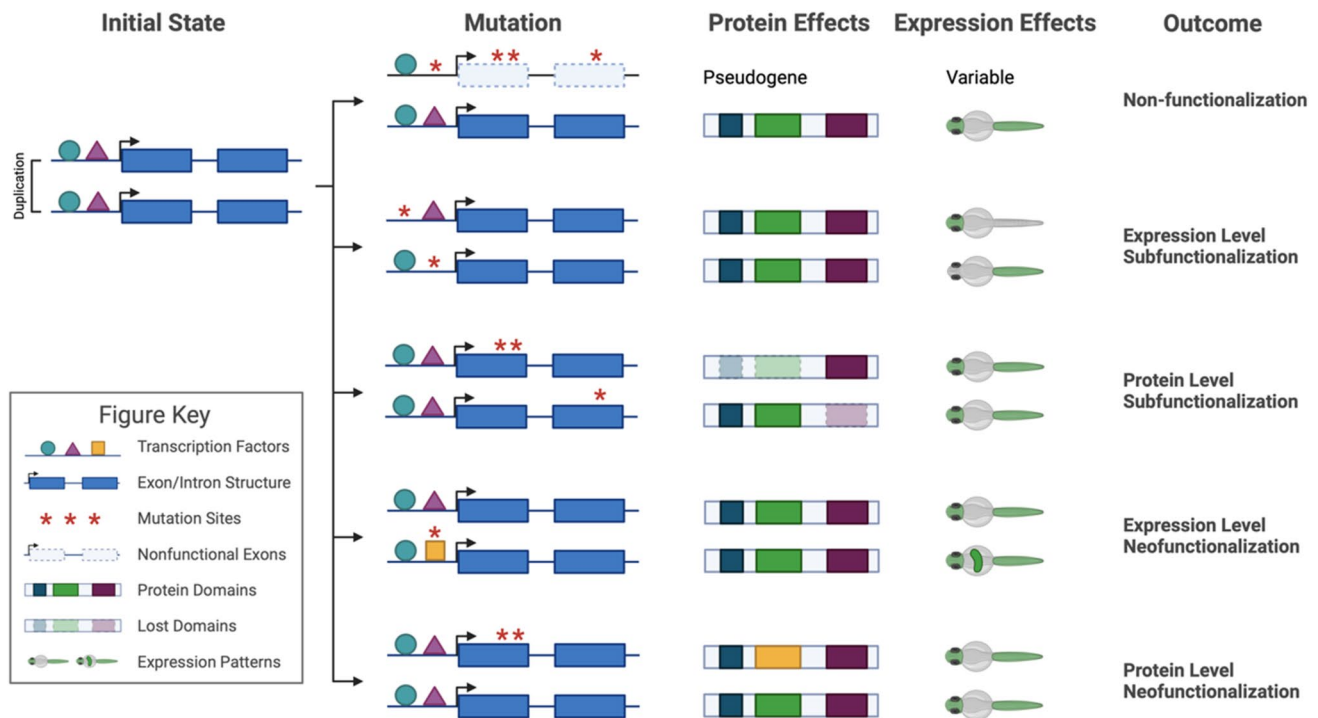


Fig. 3 Schematic representation of the different fates of duplications in zebrafish. Example protein domains are shown in blue, green, and dark purple. Hypothetical regulatory proteins are shown as a blue circle, purple triangle or yellow square. 1st outcome—non-functionalization: One of the gene copies acquires a degenerative mutation that turns it into a pseudogene (shown in gray boxes). This gene is no longer used, and only one functional copy remains. 2nd outcome—expression level subfunctionalization: Differential regulation leads to each gene copy to be expressed in specific patterns, allowing distinct

but complementary functionalities. 3rd outcome—protein level subfunctionalization: Mutations modify protein domains, changing their functions and allowing each gene copy to fulfill subsets of the original function. 4th outcome—expression level neofunctionalization: Novel expression patterns emerge in one of the gene copies, providing new functional capabilities. 5th outcome—protein level subfunctionalization: Unique mutations in one gene copy confer an entirely new protein function, enhancing biological diversity and adaptability. Created with [Biorender.com](https://www.biorender.com)

nrf2a morphants exhibited an oxidized glutathione redox potential and altered glutathione dynamics starting at 48 hpf, indicative of *nrf2a*'s active role during mid-embryogenesis in regulating the oxidative stress response. In contrast, *nrf2b* knockdown led to a reduction in the glutathione redox potential at 3 hpf, with later developmental stages resembling control embryos (Sant et al. 2017; Timme-Laragy et al. 2012).

Not all examples of subfunctionalization involve discernable differences in expression patterns. Subfunctionalization can result from protein domain modulation, where the duplicates divide the original function at the coding sequence level (Fig. 3: Protein level subfunctionalization). As a result, neither of the duplicates retains the complete set of functions of the original protein. This type of subfunctionalization has largely been dismissed as a rare exception, but gene expression analysis using modern transcriptomics techniques has shown that it may be more common than previously thought. For example, in a study by Hill et al., (2017), gene expression analysis in the zebrafish heart revealed that over 10% of duplicated genes retained similar temporal and spatial expression patterns.

Despite being assumed to be rare, a few examples of studies on protein-level subfunctionalization can be found. One well-characterized example of protein-level subfunctionalization was provided by Parrie et al. (2013), who demonstrated, despite co-expression in developing heart and limb, Tbx5a and Tbx5b display distinct amino acid sequences that confer unique functions. The T-box domain of Tbx5b shares only 83% sequence identity with its Tbx5a counterpart, which is significantly lower than the typical 95–99% sequence identity observed between paralogous T-box genes within the same subfamily. This divergence is reflected in their independent essential requirements for cardiac and fin development, as injection of mRNA from one *tbx5* paralog was unable to compensate for the loss of the other (Parrie et al. 2013). This domain-specific understanding is helpful because it allows for a granular view of protein functionality, which is essential for designing targeted therapies that can modulate specific protein activities without affecting others. Additionally, this knowledge can be pivotal in understanding the molecular basis of diseases caused by mutations in

specific protein domains of TBX5, paving the way for precision medicine approaches and improved genetic diagnostics.

In many cases, duplicated genes can be retained without discernable differences in expression pattern or gene function. This can sometimes be explained by changes in gene expression levels where each paralog is expressed at lower levels than the ancestral gene, thus requiring two copies to maintain sufficient protein levels in the cell. In other cases, an increase in the gene expression, or dosage, benefits the cell, leading to increased fitness. A study by Ihmels et al. (2007) observed that in yeast, certain duplicate genes, particularly those involved in dosage amplification like histone genes, maintain high and correlated expression levels post-duplication. This correlation in expression patterns is indicative of a co-regulation mechanism, suggesting that these gene duplicates are preserved at high abundance to meet cellular demands. Thus, cells may derive a selective advantage from such duplication due to the increased dosage of critical gene products, although the overall impact of duplicates on genetic robustness appears to be modest.

Studying subfunctionalization can provide insights into the dosage sensitivity of their human orthologs, especially those that exhibit haploinsufficiency phenotypes. Haploinsufficiency occurs when a single copy of a gene is not sufficient to maintain normal function. Although clear examples of haploinsufficient genes in humans being studied in zebrafish have been published, expression modulation to compensate for the loss of a duplicate is a well-characterized phenomenon in zebrafish (El-Brolosy et al. 2019). For example, Zinc Finger Protein 143 (ZNF143) is a sequence-specific transcriptional activator that plays a critical role in the regulation of both mRNA and small nuclear RNA gene promoters. In a study by Huning and Kunkel (2020), the knockout effects of the two paralogous genes, *znf143a* and *znf143b*, in zebrafish were investigated using CRISPR interference (CRISPRi). The study used in situ hybridization to analyze the expression of these two genes at 24 h post-fertilization (hpf). The results showed that both *znf143a* and *znf143b* mRNAs are strongly expressed in the brain regions, including the forebrain, midbrain, and hindbrain. Despite their similar spatial expression, there is a notable difference in their expression levels during early development. The knockdown results showed that knocking down either *znf143a* or *znf143b* led to similar phenotypic defects in zebrafish embryos. They also observed that knocking down *znf143a* led to a 1.5-fold increase in *znf143b* mRNA levels, indicating a compensatory mechanism that attempts to maintain functional Znf143 protein levels. This compensatory response was not reciprocated when *znf143b* was knocked down. The study concludes that the tight control of gene dosage is likely essential for maintaining developmental processes and that disruptions in this balance can lead to significant morphological and functional defects.

Abnormal expression levels of *znf143*, either too high or too low, could potentially lead to similar disruptions as seen in haploinsufficiency, affecting cell cycle regulation. In the case of WGD, organisms might develop mechanisms to balance the expression levels of all duplicated genes to prevent deleterious effects of over- or under-expression. On the other hand, tandem duplications, which occur when genes are duplicated within the same chromosome, often lead to dosage imbalances. This imbalance can be detrimental if the duplicated gene is dose-sensitive, as is the case with many genes involved in genetic disorders (Birchler & Yang 2022; Rice & McLysaght 2017). To summarize, the impact of dosage imbalance due to duplication can be either advantageous or detrimental, depending on the specific gene involved.

Similar research can also elucidate specific aspects of gene function that might be obscured in species where these functions are still combined in a single gene. For example, Dravet syndrome in humans is caused by heterozygous loss-of-function mutations in the *SCN1A* (sodium voltage-gated channel alpha subunit 1) gene. This protein is pivotal for initiating and propagating action potentials in neurons, with a pronounced expression in inhibitory interneurons of the central nervous system. The dysfunction of these channels, due to the mutations, compromises neuronal inhibition, leading to the neuronal hyperexcitability observed in seizure disorders (Martins Custodio et al. 2023). The phenotypic spectrum of Dravet syndrome encompasses seizures (generalized, focal, or unilateral), cognitive impairments (visual troubles, delayed motor skills, speech, and attention), and movement disorders (ataxia, choreoathetosis, and gait issues). This phenotypic spectrum is influenced by several factors related to the *SCN1A* gene mutations. The class of variant—whether missense, nonsense, or frameshift—has distinct effects on the protein function (Chen et al. 2022, b; Gonsales et al. 2019; Ma et al. 2022). Missense mutations may partially alter channel function, while nonsense or frameshift mutations could result in a complete loss of function. Although Dravet syndrome has been extensively studied in humans, the full phenotypic spectrum and the nuances of genotype–phenotype correlations are not yet fully understood.

Dravet syndrome represents a clear use-case where duplications in zebrafish present a valuable model for study. The zebrafish paralogs *scn1laa* and *scn1lab* not only have different spatial expression patterns but also exhibit distinct phenotypic traits (Weuring et al. 2022). Significant attention has been directed towards the *scn1lab* gene in zebrafish, as it is the more conserved ortholog of the human *SCN1A* gene, and mutations in *scn1lab* have been consistently linked to epilepsy phenotypes in zebrafish, echoing the pathological features observed in Dravet syndrome (Griffin et al. 2017; Schoonheim et al. 2010; Sourbron et al. 2016; Weuring et al. 2020). However, the *scn1laa* gene, despite being less

studied, has been shown to influence neuronal excitability as well. Mutant phenotypes of *scn11aa* under standard conditions present with altered brain and forebrain transmission of nerve impulses, supporting *scn11aa*'s involvement in the neurological pathways relevant to Dravet syndrome (Griffin et al. 2017; Weuring et al. 2022).

The idea that research should focus solely on *scn11ab*, as suggested by some in the field (Weuring et al. 2022), overlooks the complexity of genetic interactions and the potential insights offered by studying both paralogs. In zebrafish, *scn11aa* and *scn11ab* likely have undergone subfunctionalization or neofunctionalization as mutants display partially overlapping phenotypes. By studying both paralogs, we can uncover potential compensatory mechanisms that one gene may exert in the presence of mutations in the other. Such compensatory interactions can have critical implications for understanding the variability and penetrance of *SCN1A*-related disorders. Exploring differential drug responses that might not be apparent when examining a single paralog may also pave the way for more personalized approaches to treatment.

Another example is Insulin-like Growth Factor (IGF). Zebrafish possess four distinct IGF genes: *igf-1a*, *igf-1b*, *igf-2a*, and *igf-2b* (Zou et al. 2009). This duplication has permitted the subfunctionalization of IGF proteins, with each evolving distinct expression patterns and physiological roles. In zebrafish, the IGF genes *igf-2a* and *igf-2b* exhibit high sequence identity with human *IGF-2* and with each other. However, *igf-2b* mRNA is specifically expressed in the liver, while *igf-2a* mRNA is widely expressed across tissues, likely reflecting the partitioning of function among these genes. The presence of two distinct 5'-UTR sequences in both zebrafish *igf-2a* and *igf-1b*, resulting in different transcription initiation sites and signal peptides, further underscores the evolutionary pressure to maintain distinct physiological roles for these paralogs (Zou et al. 2009). These divergent expression profiles suggest that studying these genes in zebrafish can provide deeper insight into their individual contributions to tissue development and homeostasis. In humans, dysregulation of IGF signaling is implicated in diseases ranging from growth disorders to cancer (Murrell et al. 2004). By dissecting the distinct roles of *igf-2a* and *igf-2b* in zebrafish, we can gain valuable insights into the tissue-specific functions and regulatory mechanisms of IGF signaling, potentially leading to targeted therapies in human diseases where IGF is dysregulated.

In summary, the study of subfunctionalized genes in zebrafish enhances our understanding of gene function by allowing us to see how individual aspects of a gene's role evolve and operate in a biological context. This information can be crucial for understanding similar processes in humans, where such functional distinctions might not be as easily observable.

Neofunctionalization

Neofunctionalization is the process through which functional divergence occurs when one of the duplicated genes obtains a novel function advantageous to the organism, and the gene is retained (Hurles 2004) (Fig. 3: neofunctionalization). It is difficult to study due to the low probability of a gene acquiring a new function, leading to a natural paucity of cases. In addition, pinpointing the original gene function is challenging, making it hard to prove changes in protein function. However, recent advances in computational methods for ancestral gene reconstruction and protein modeling combined with the rapid proliferation of reference genomes may make neofunctionalization easier to identify across evolutionary timelines (Cai et al. 2004; Finnigan et al. 2012; Joy et al. 2016; Nucedal & Laub 2022; Ogawa et al. 2013; Scossa & Fernie 2021; Voordeckers et al. 2012). Neofunctionalization typically co-occurs with subfunctionalization, with one gene adopting new roles while preserving a subset of ancestral functions.

Although rare, a few confirmed cases of neofunctionalization exist in the literature. One well-documented instance of this is observed in the gene duplication of *hoxa13a* and *hoxa13b* in zebrafish. The gene *hoxa13a* exhibits hypermutability, which has allowed it to accumulate a higher number of mutations when compared to its paralog *hoxa13b*. Such hypermutability has been linked to the development of the median fin fold (MFF) in zebrafish, a specialized fin structure observed in the early development of many teleost fish. This suggests that the function associated with *hoxa13a* and its hypermutability is a novel role rather than one retained from the ancestral gene function before duplication (Crow et al. 2009). This type of mutation-driven divergence is crucial in the evolution of species as it provides the genetic variability on which natural selection can act, giving rise to new phenotypes that can be subject to evolutionary pressures.

Complex evolutionary fates

In many cases, two or more of the evolutionary fates discussed above happen together. For example, duplicates may exhibit a combination of the expression pattern and protein modulation forms of subfunctionalization. In other instances, duplicate genes have similar and unique functions that are required in tandem with one another but at different times in development. Dosage compensation is often accompanied by various forms of subfunctionalization (Hultman et al. 2007; Lagman et al. 2015; Leerberg et al. 2019; Sedletcaia & Evans 2011). For example, *hoxb1a* and *hoxb1b* in zebrafish have significant functional redundancy as each of them is capable of promoting Mauthner neuron differentiation and rescuing the defects caused by knock-down of *hoxb1b* while also performing unique functions

by themselves (McClintock et al. 2002). In this particular situation, the “Piggyback hypothesis,” which posits that the preservation of redundant parts of the gene is influenced by structural constraints within the genome, provides further insight (Qian et al. 2010). This preservation is thought to be influenced by the proximity of the gene segment that codes for a unique function to the redundant part, which is why the redundant function remains in the genome. Other combinations where the duplicated genes both subfunctionalize and neofunctionalize are possible, although cases have only recently begun to be identified.

Impact on human health

Comparative studies that analyze and compare both duplicate genes in zebrafish can inform research into human biology in at least four important ways. First, comparisons of the structural and functional differences between pairs can be used to identify the roles of specific domains and enhancers. Paralogs can evolve to possess distinct cis-regulatory elements that drive their expression in different tissues or developmental stages, leading to functional diversification. For instance, the zebrafish *sox9a* and *sox9b* genes, which arose from the teleost-specific genome duplication, have been shown to possess distinct functions in craniofacial and pectoral fin development due to their distinct expression patterns; *sox9a* is expressed predominantly in the somites and pharyngeal arches while *sox9b* is more prominent in the eye and otic vesicle (ear) (Yan et al. 2005). Consistently, loss of *sox9b* leads to ear defects and reduced craniofacial cartilage. Conversely, the lack of *sox9a* affects chondrocyte stacking in cartilage, which could be due to alterations in domains responsible for chondrogenic differentiation. Thus, the specific functions of *sox9* in each of these tissues can be more easily parsed out in the zebrafish than it can in mammalian models where a single gene is involved in both processes.

The second way that duplicate gene studies can inform our understanding of human biology is by comparative analysis of gene clusters. For instance, the Hox gene clusters, which are fundamental for the development of the body plan, have been refined through multiple whole-genome duplication events. In early vertebrates, two rounds of genome duplication events (2R) produced the four paralogous Hox clusters (HoxA, HoxB, HoxC, and HoxD) (Fig. 2) widely conserved across most species (Ozernyuk & Schepetov 2022; Singh & Krumlauf 2022; Soshnikova et al. 2013; G. P. Wagner et al. 2003). Teleosts, however, have seven or eight Hox clusters due to the TS-3R WGD and subsequent gene loss in some species (Amores et al. 1998; Málaga-Trillo & Meyer 2001; Pascual-Anaya et al. 2013; Yamada et al. 2021). By comparing the differences and similarities in how these genes function in zebrafish and humans, researchers

can gain insights into their role in development and disease. For example, Dietrich et al. (2021) compared Hox gene expression patterns between zebrafish in the developing limb and highlighted how both the similarities and differences between genes and expression patterns can be used to model human skeletal diseases like osteogenesis imperfecta and osteopetrosis. This comparative genetic approach underscores the importance of *Hox* genes in vertebrate skeletal formation, disease phenotypes and potential therapies.

Another example of gene cluster evolution is the duplication of the Major Histocompatibility Complex (MHC) genes, which are crucial for the immune response to pathogens (Cruz-Tapias et al. 2013). *MHC* genes have undergone several rounds of duplication in both humans and zebrafish (Sambrook et al. 2005). In humans, *MHC* gene expansion is primarily due to tandem duplications (Traherne 2008). Zebrafish also share this mechanism, but uniquely, they have further diversified their MHC genes through the TS-3R WGD (Bingulac-Popovic et al. 1997; Dirscherl et al. 2014). This has led to a wide variety of MHC class I genes in zebrafish, categorized into U, Z, and L lineages, dispersed across multiple chromosomes (Dirscherl et al. 2014). By comparing the evolution of the more extensive and more complex MHC repertoire in zebrafish, we can uncover how the interplay between unique immunological threats and gene duplication have driven the evolution of our immune system (Bingulac-Popovic et al. 1997; Dirscherl et al. 2014).

Third, gene duplication in zebrafish can also serve as a model for human diseases caused by gene duplication or amplification. For instance, gene duplication events are associated with certain cancers in humans where the extra copies of genes lead to overexpression and tumorigenesis (Baines et al. 2022; Glenfield & Innan 2021). A prime example is the amplification of the *HER2* gene, which occurs in about 15–20% of breast cancers. Increased copy numbers of *HER2* result in continuous growth signals that contribute to the uncontrolled proliferation characteristic of cancer. In this context, zebrafish offer a powerful model for studying the implications of gene duplication, as *her2* is duplicated in the zebrafish genome but without the typical oncogenic effects (Cappuzzo et al. 2005; Chen et al. 2022, b).

Finally, studying gene duplication can also contribute to understanding chemotherapy resistance, as duplication in cancer cells can result in overexpression and, consequently, the development of resistance to certain drugs (Glenfield & Innan 2021; Wu et al. 2015). An early illustration of the critical impact of gene duplication in cancer resistance was the discovery of the dihydrofolate reductase (*DHFR*) gene’s duplication in 1978 (Alt et al. 1978). The DHFR enzyme, which is targeted by the drug methotrexate, is vital for DNA synthesis and cellular proliferation (Askari & Krajcinovic 2010). However, duplication of the DHFR gene results in increased enzyme production, enabling cancer cells to

overcome the effects of methotrexate (Alt et al. 1978; Glenfield & Innan 2021; Turner et al. 2017). Future studies on the duplication and diversification of these and other genes involved in cancer will increase our understanding of how gene overexpression following duplication promotes oncogenesis, potentially paving the way for novel therapeutic approaches in treating human cancers (Kalkat et al. 2017).

Conclusion

Studying how genes duplicate and evolve is crucial for understanding their structure, function, and regulation and how changes in these factors can lead to evolution and disease. Post-duplication, these genes acquire mutations more freely, fostering new functions and aiding adaptation. Regardless of whether duplicated genes remain unaltered, become pseudogenes, or evolve novel functions, examining each outcome can shed light on evolutionary innovations and identify factors for genetic disease modeling and therapeutic development. In addition, analysis of pseudogene formation provides new perspectives on gene expression regulation and genome stability relevant to cancer genomics and the identification of genetic disease markers.

However, in zebrafish research, certain paralogous genes have generally been studied more extensively compared to their paralogs. This pattern demonstrates the tendency in genetic research to concentrate on genes that exhibit more pronounced or observable phenotypic effects or are more closely related to their human counterparts, potentially overshadowing the significant roles of their paralogs. In such cases, valuable information about how evolution drives functional innovation and how that innovation may affect disease and even potential treatments may be missed. More importantly, the results from genetic manipulations of one paralog in zebrafish may not accurately represent the full repertoire of gene functions for its ortholog in humans, as gene expression patterns and protein domain functions may be partitioned between the two paralogs. These factors cannot be ignored if we want to maximize the utility of zebrafish as a model for human disease.

A well-designed study for the duplicates should include at least two components: (1) Identification of the expression domain of the paralogs and (2) functional analysis, such as generating single and double knockouts, to identify individual functions and check for redundancy. Additional supportive information can be obtained through rescue experiments to check for compensation, analysis of the phylogenetic relationship with the ancestral gene, and examination of expression patterns across different developmental stages. However, employing heterologous expression systems comes with challenges, such as potential discrepancies in post-translational modifications, differences in

cellular environments, and the intricacies of gene regulation across species. Despite these obstacles, the strategic use of domain deletions or switches between duplicated genes in such systems can elucidate protein domains' roles in mediating specific processes and phenotypes, thereby confirming the molecular basis for particular functions.

Although the specific applications of these methods will vary on a case-by-case basis, the key is that both paralogs are studied together. Only by comparing paralogs can we understand how evolutionary pressures shape gene expression and protein function. Only by comparing paralogs can we accurately model human disease in the zebrafish. And only by expanding our view to embrace the unique information embedded in the zebrafish genome can we fully utilize this vast genetic resource to advance our understanding of nature and disease.

Acknowledgements We extend our deepest gratitude to Logan Melting, Sophia Schroeder, and Hannah Abernathy for their invaluable contributions to the research presented in this paper. Their dedication and meticulous attention to detail have been indispensable to the success of this work. We are also profoundly thankful to Dr. Jeffrey Barrow for his thorough review and insightful feedback on our manuscript.

Author contribution M.T. and J.H. helped research and write the manuscript. P. W. Created and edited the figures and helped edit the paper.

Funding M.T. and J.T.H. are supported by a grant from the NICHD (R15HD098969).

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adhish M, Manjubala I (2023) Effectiveness of zebrafish models in understanding human diseases—a review of models. *Heliyon* 9(3):e14557. <https://doi.org/10.1016/j.heliyon.2023.e14557>
- Alt FW, Kellems RE, Bertino JR, Schimke RT (1978) Selective multiplication of dihydrofolate reductase genes in

- methotrexate-resistant variants of cultured murine cells. *J Biol Chem* 253(5):1357–1370
- Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH (1998) Zebrafish hox clusters and vertebrate genome evolution. *Science* 282(5394):1711–1714. <https://doi.org/10.1126/science.282.5394.1711>
- Anatskaya OV, Vinogradov AE (2022) Polyploidy as a fundamental phenomenon in evolution, development, adaptation and diseases. *Int J Mol Sci* 23(7):3542. <https://doi.org/10.3390/ijms23073542>
- Anderson EB, Mao Q, Ho RK (2022) Tbx5a and Tbx5b paralogues act in combination to control separate vectors of migration in the fin field of zebrafish. *Dev Biol* 481:201–214. <https://doi.org/10.1016/j.ydbio.2021.10.008>
- Arnegard ME, Zwickl DJ, Lu Y, Zakon HH (2010) Old gene duplication facilitates origin and diversification of an innovative communication system—Twice. *Proc Natl Acad Sci* 107(51):22172–22177. <https://doi.org/10.1073/pnas.1011803107>
- Askari BS, Krajcinovic M (2010) Dihydrofolate reductase gene variations in susceptibility to disease and treatment outcomes. *Curr Genomics* 11(8):578–583. <https://doi.org/10.2174/138920210793360925>
- Ata H, Ekstrom TL, Martínez-Gálvez G, Mann CM, Dvornikov AV, Schaeffbauer KJ, Ma AC, Dobbs D, Clark KJ, Ekker SC (2018) Robust activation of microhomology-mediated end joining for precision gene editing applications. *PLoS Genet* 14(9):e1007652. <https://doi.org/10.1371/journal.pgen.1007652>
- Auer TO, Duroure K, Cian AD, Concordet J-P, Bene FD (2014) Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res* 24(1):142–153. <https://doi.org/10.1101/gr.161638.113>
- Baanannou A, Rastegar S, Bouzid A, Takamiya M, Gerber V, Souissi A, Beil T, Jrad O, Strähle U, Masmoudi S (2020) Gene duplication and functional divergence of the zebrafish otospiralin genes. *Dev Genes Evol* 230(1):27–36. <https://doi.org/10.1007/S00427-019-00642-8/TABLES/3>
- Baines C, Meitern R, Kreitsberg R, Sepp T (2022) Comparative study of the evolution of cancer gene duplications across fish. *Evol Appl* 15(11):1834–1845. <https://doi.org/10.1111/eva.13481>
- Barut BA, Zon LI (2000) Realizing the potential of zebrafish as a model for human disease. *Physiol Genomics* 2(2):49–51. <https://doi.org/10.1152/physiolgenomics.2000.2.2.49>
- Belliveau DJ, Venkatachalam AB, Thisse C, Thisse B, Ma H, Wright JM (2010) The duplicated retinol-binding protein 7 (rbp7) genes are differentially transcribed in embryos and adult zebrafish (*Danio rerio*). *Gene Expr Patterns: GEP* 10(4–5):167–176. <https://doi.org/10.1016/j.gep.2010.04.003>
- Bingulac-Popovic J, Figueroa F, Sato A, Talbot WS, Johnson SL, Gates M, Postlethwait JH, Klein J (1997) Mapping of Mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. *Immunogenetics* 46(2):129–134. <https://doi.org/10.1007/s002510050251>
- Birchler JA, Yang H (2022) The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell* 34(7):2466–2474. <https://doi.org/10.1093/plcell/koac076>
- Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution[W]. *Plant Cell* 16(7):1679–1691. <https://doi.org/10.1105/tpc.021410>
- Cai W, Pei J, Grishin NV (2004) Reconstruction of ancestral protein sequences and its applications. *BMC Evol Biol* 4(1):33. <https://doi.org/10.1186/1471-2148-4-33>
- Cappuzzo F, Varella-Garcia M, Shigematsu H, Domenichini I, Bartolini S, Ceresoli GL, Rossi E, Ludovini V, Gregorc V, Toschi L, Franklin WA, Crino L, Gazdar AF, Bunn PA, Hirsch FR (2005) Increased HER2 gene copy number is associated with response to gefitinib therapy in epidermal growth factor receptor-positive non-small-cell lung cancer patients. *J Clin Oncol: Off J Am Soc Clin Oncol* 23(22):5007–5018. <https://doi.org/10.1200/JCO.2005.09.111>
- Chang L, Khoo B, Wong L, Tropepe V (2006) Genomic sequence and spatiotemporal expression comparison of zebrafish mbx1 and its paralog, mbx2. *Dev Genes Evol* 216(10):647–654. <https://doi.org/10.1007/s00427-006-0082-7>
- Chen C, Fang F, Wang X et al (2022). Phenotypic and genotypic characteristics of SCN1A associated seizure diseases. *Front Mol Neurosci* 15. <https://doi.org/10.3389/fnmol.2022.821012>
- Chen X, Lin Y, Jiang Z, Li Y, Zhang Y, Wang Y, Yu F, Guo W, Chen L, Chen M, Zhang W, Wang C, Fu F (2022b) HER2 copy number quantification in primary tumor and cell-free DNA provides additional prognostic information in HER2 positive early breast cancer. *The Breast: Off J Eur Soc Mastology* 62:114–122. <https://doi.org/10.1016/j.breast.2022.02.002>
- Chen X, Wan L, Wang W et al (2020) Re-recognition of pseudogenes: from molecular to clinical applications 10(4):1479–1499. <https://doi.org/10.7150/thno.40659>
- Chia K, Klingseisen A, Sieger D et al (2022). Zebrafish as a model organism for neurodegenerative disease. *Front Mol Neurosci* 15. <https://www.frontiersin.org/articles/10.3389/fnmol.2022.940484>
- Choi T-Y, Choi T-I, Lee Y-R, Choe S-K, Kim C-H (2021) Zebrafish as an animal model for biomedical research. *Exp Mol Med* 53(3):310–317. <https://doi.org/10.1038/s12276-021-00571-5>
- Christiaens JF, Van Mulders SE, Duitama J, Brown CA, Ghequire MG, De Meester L, Michiels J, Wenseleers T, Voordeckers K, Verstrepen KJ (2012) Functional divergence of gene duplicates through ectopic recombination. *EMBO Rep* 13(12):1145–1151. <https://doi.org/10.1038/embor.2012.157>
- Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet* 9(12):938–950. <https://doi.org/10.1038/nrg2482>
- Copley SD (2020) Evolution of new enzymes by gene duplication and divergence. *FEBS J* 287(7):1262–1283. <https://doi.org/10.1111/febs.15299>
- Crow KD, Amemiya CT, Roth J, Wagner GP (2009) Hypermutability of HoxA13A and functional divergence from its paralog are associated with the origin of a novel developmental feature in zebrafish and related taxa (Cypriniformes). *Evol Int J Org Evol* 63(6):1574–1592. <https://doi.org/10.1111/j.1558-5646.2009.00657.x>
- Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP (2006) The “Fish-Specific” hox cluster duplication is coincident with the origin of teleosts. *Mol Biol Evol* 23(1):121–136. <https://doi.org/10.1093/molbev/msj020>
- Crow KD, Wagner GP (2006) What is the role of genome duplication in the evolution of complexity and diversity? *Mol Biol Evol* 23(5):887–892. <https://doi.org/10.1093/molbev/msj083>
- Cruz-Tapias P, Castiblanco J, Anaya J.-M (2013). Major histocompatibility complex: antigen processing and presentation. In *Autoimmunity: From Bench to Bedside* [Internet]. El Rosario University Press. <https://www.ncbi.nlm.nih.gov/books/NBK459467/>. Accessed 23 Nov 2022
- de Martino S, Yan YL, Jowett T, Postlethwait JH, Varga ZM, Ashworth A, Austin CA (2000) Expression of sox11 gene duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development. *Dev Dynamics: Off Publ Am Assoc Anat* 217(3):279–292. [https://doi.org/10.1002/\(SICI\)1097-0177\(200003\)217:3<279::AID-DVDY6%3E3.0.CO;2-S](https://doi.org/10.1002/(SICI)1097-0177(200003)217:3<279::AID-DVDY6%3E3.0.CO;2-S)
- Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 3(10):e314. <https://doi.org/10.1371/journal.pbio.0030314>

- del Pozo JC, Ramirez-Parra E (2015) Whole genome duplications in plants: an overview from Arabidopsis. *J Exp Bot* 66(22):6991–7003. <https://doi.org/10.1093/jxb/erv432>
- Delporte FM, Pasque V, Devos N, Manfroid I, Voz ML, Motte P, Bie-mar F, Martial JA, Peers B (2008) Expression of zebrafish pax6b in pancreas is regulated by two enhancers containing highly conserved cis-elements bound by PDX1, PBX and PREP factors. *BMC Dev Biol* 8:53. <https://doi.org/10.1186/1471-213X-8-53>
- Dietrich K, Fiedler IA, Kurzyukova A, López-Delgado AC, McGowan LM, Geurtzen K, Hammond CL, Busse B, Knopf F (2021) Skeletal biology and disease modeling in zebrafish. *J Bone Miner Res* 36(3):436–458. <https://doi.org/10.1002/jbmr.4256>
- Dirscherl H, McConnell SC, Yoder JA, de Jong JLO (2014) The MHC class I genes of zebrafish. *Dev Comp Immunol* 46(1):11–23. <https://doi.org/10.1016/j.dci.2014.02.018>
- Doyle JM, Croll RP (2022) A critical review of zebrafish models of Parkinson's disease. *Front Pharmacol* 13:835827. <https://doi.org/10.3389/fphar.2022.835827>
- El-Brolosy MA, Kontarakis Z, Rossi A et al (2019) Genetic compensation triggered by mutant mRNA degradation. *Nature* 568:193–197. <https://doi.org/10.1038/s41586-019-1064-z>
- Evans-Yamamoto D, Dubé AK, Saha G et al (2023). Parallel nonfunctionalization of CK1 δ/ϵ kinase ohnologs following a whole-genome duplication event. *bioRxiv*. <https://doi.org/10.1101/2023.10.02.560513>
- Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW (2012) Evolution of increased complexity in a molecular machine. *Nature* 481(7381):360–364. <https://doi.org/10.1038/nature10724>
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151(4):1531–1545. <https://doi.org/10.1093/genetics/151.4.1531>
- Galicia CA, Sukeena JM, Stenkamp DL, Fuerst PG (2018) Expression patterns of dscam and sdk gene paralogs in developing zebrafish retina. *Mol vis* 24:443–458
- Glasauer SMK, Neuhauss SCF (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics* 289(6):1045–1060. <https://doi.org/10.1007/s00438-014-0889-2>
- Glenfield C, Innan H (2021) Gene duplication and gene fusion are important drivers of tumorigenesis during cancer evolution. *Genes* 12(9):1376. <https://doi.org/10.3390/genes12091376>
- Glenfield C, McLysaght A (2018) Pseudogenes provide evolutionary evidence for the competitive endogenous RNA hypothesis. *Mol Biol Evol* 35(12):2886–2899. <https://doi.org/10.1093/molbev/msy183>
- Goldsmith JR, Jobin C (2012) Think small: zebrafish as a model system of human pathology. *J Biomed Biotechnol* 2012:817341. <https://doi.org/10.1155/2012/817341>
- Gonsales MC, Montenegro MA, Preto P, Guerreiro MM, Coan AC, Quast MP, Carvalho BS, Lopes-Cendes I (2019) Multimodal analysis of SCN1A missense variants improves interpretation of clinically relevant variants in Dravet syndrome. *Front Neurol* 10:289. <https://doi.org/10.3389/fneur.2019.00289>
- Gosmain Y, Katz LS, Masson MH et al (2012) Pax6 is crucial for β -cell function, insulin biosynthesis, and glucose-induced insulin secretion. *Mol Endocrinol* (Baltimore, Md) 26(4):696–709. <https://doi.org/10.1210/me.2011-1256>
- Griffin A, Hamling KR, Knupp K, Hong S, Lee LP, Baraban SC (2017) Clemizole and modulators of serotonin signalling suppress seizures in Dravet syndrome. *Brain* 140(3):669–683. <https://doi.org/10.1093/brain/aww342>
- Groen JN, Capraro D, Morris KV (2014) The emerging role of pseudogene expressed non-coding RNAs in cellular functions. *Int J Biochem Cell Biol* 0:350–355. <https://doi.org/10.1016/j.biocel.2014.05.008>
- Hakes L, Pinney JW, Lovell SC, Oliver SG, Robertson DL (2007) All duplicates are not equal: The difference between small-scale and genome duplication. *Genome Biol* 8(10):R209. <https://doi.org/10.1186/gb-2007-8-10-r209>
- Hallin J, Landry CR (2019) Regulation plays a multifaceted role in the retention of gene duplicates. *PLoS Biol* 17(11):e3000519. <https://doi.org/10.1371/journal.pbio.3000519>
- Hardison RC (2012) Evolution of hemoglobin and its genes. *Cold Spring Harb Perspect Med* 2(12):a011627. <https://doi.org/10.1101/cshperspect.a011627>
- Hill JT, Demarest B, Gorski B, Smith M, Yost HJ (2017) Heart morphogenesis gene regulatory networks revealed by temporal expression analysis. *Development* (Cambridge) 144(19):3487–3498. <https://doi.org/10.1242/dev.154146>
- Hoffmann FG, Storz JF, Kuraku S et al (2021) Whole-genome duplications and the diversification of the Globin-X genes of vertebrates. *Genome Biol Evol* 13(10):evab205. <https://doi.org/10.1093/gbe/evab205>
- Holland LZ, Ocampo Daza D (2018) A new look at an old question: when did the second whole genome duplication occur in vertebrate evolution? *Genome Biol* 19(1):2–5. <https://doi.org/10.1186/s13059-018-1592-0>
- Howe K, Clark MD, Torroja CF et al (2013). The zebrafish reference genome sequence and its relationship to the human genome. <https://doi.org/10.1038/nature12111>
- Huang X, Chen Y, Li W-B, Cohen SN, Liao F-F, Li L, Xu H, Zhang Y (2010) The Rps23rg gene family originated through retroposition of the ribosomal protein s23 mRNA and encodes proteins that decrease Alzheimer's β -amyloid level and tau phosphorylation. *Hum Mol Genet* 19(19):3835–3843. <https://doi.org/10.1093/hmg/ddq302>
- Hultman KA, Bahary N, Zon LI, Johnson SL (2007) Gene Duplication of the zebrafish kit ligand and partitioning of melanocyte development functions to kit ligand a. *PLoS Genet* 3(1):e17. <https://doi.org/10.1371/journal.pgen.0030017>
- Huning L, Kunkel GR (2020) Two paralogous znf143 genes in zebrafish encode transcriptional activator proteins with similar functions but expressed at different levels during early development. *BMC Mol Cell Biol* 21(1):3. <https://doi.org/10.1186/s12860-020-0247-7>
- Hurles M (2004) Gene Duplication: The Genomic Trade in Spare Parts 2. <https://doi.org/10.1371/journal.pbio.0020206>
- Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh J-RJ, Joung JK (2013) Efficient in vivo genome editing using RNA-guided nucleases. *Nat Biotechnol* 31(3):227–229. <https://doi.org/10.1038/nbt.2501>
- Ihmels J, Collins SR, Schuldiner M et al (2007). Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. *Mol Syst Biol* 3. <https://doi.org/10.1038/MSB4100127>
- Innan H, Kondrashov F (2010) The evolution of gene duplications: Classifying and distinguishing between models. *Nat Rev Genet* 11(2):97–108. <https://doi.org/10.1038/nrg2689>
- Inoue J, Sato Y, Sinclair R, Tsukamoto K, Nishida M (2015) Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. *Proc Natl Acad Sci USA* 112(48):14918–14923. <https://doi.org/10.1073/pnas.1507669112>
- Joy JB, Liang RH, McCloskey RM, Nguyen T, Poon AFY (2016) Ancestral Reconstruction. *PLOS Comput Biol* 12(7):e1004763. <https://doi.org/10.1371/journal.pcbi.1004763>
- Kalkat M, De Melo J, Hickman KA, Lourenco C, Redel C, Resettec D, Tamachi A, Tu WB, Penn LZ (2017) MYC deregulation in primary human cancers. *Genes* 8(6):151. <https://doi.org/10.3390/genes8060151>

- Kamei H, Lu L, Jiao S, Li Y, Gyurup C, Laursen LS, Oxvig C, Zhou J, Duan C (2008) Duplication and diversification of the hypoxia-inducible IGFBP-1 gene in zebrafish. *PLoS ONE* 3(8):e3091. <https://doi.org/10.1371/journal.pone.0003091>
- Kikuta H, Laplante M, Navratilova P et al (2007) Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res* 17(5):545–555. <https://doi.org/10.1101/gr.6086307>
- Kleinjan DA, Bancewicz RM, Gautier P, Dahm R, Schonhaler HB, Damante G, Seawright A, Hever AM, Yeyati PL, van Heyningen V, Coutinho P (2008) Subfunctionalization of duplicated zebrafish pax6 genes by cis-regulatory divergence. *PLoS Genet* 4(2):e29. <https://doi.org/10.1371/journal.pgen.0040029>
- Kondrashov FA (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc R Soc B Biol Sci* 279(1749):5048–5057. <https://doi.org/10.1098/rspb.2012.1108>
- Koo D-H, Jugulam M, Putta K, Cuvaca IB, Peterson DE, Currie RS, Friebe B, Gill BS (2018) Gene duplication and aneuploidy trigger rapid evolution of herbicide resistance in common waterhemp1[OPEN]. *Plant Physiol* 176(3):1932–1938. <https://doi.org/10.1104/pp.17.01668>
- Kuzmin E, Taylor JS, Boone C (2022) Retention of duplicated genes in evolution. *Trends Genet: TIG* 38(1):59–72. <https://doi.org/10.1016/j.tig.2021.06.016>
- Lagman D, Callado-Pérez A, Franzén IE, Larhammar D, Abalo XM (2015) Transducin duplicates in the zebrafish retina and pineal complex: differential specialisation after the teleost tetraploidisation. *PLoS ONE* 10(3):e0121330. <https://doi.org/10.1371/journal.pone.0121330>
- Landsend ECS, Lagali N, Utheim TP (2021) Congenital aniridia – a comprehensive review of clinical features and therapeutic approaches. *Surv Ophthalmol* 66(6):1031–1050. <https://doi.org/10.1016/j.survophthal.2021.02.011>
- Leerberg DM, Hopton RE, Draper BW (2019) Fibroblast growth factor receptors function redundantly during zebrafish embryonic development. *Genetics* 212(4):1301–1319. <https://doi.org/10.1534/genetics.119.302345>
- Lewis RS, Ward AC (2004) Conservation, duplication and divergence of the zebrafish stat5 genes. *Gene* 338(1):65–74. <https://doi.org/10.1016/j.gene.2004.05.012>
- Lim HT, Kim DH, Kim H (2017) PAX6 aniridia syndrome: clinics, genetics, and therapeutics. *Curr Opin Ophthalmol* 28(5):436–447. <https://doi.org/10.1097/ICU.0000000000000405>
- Liu R-Z, Sharma MK, Sun Q, Thisse C, Thisse B, Denovan-Wright EM, Wright JM (2005) Retention of the duplicated cellular retinoic acid-binding protein 1 genes (crabp1a and crabp1b) in the zebrafish genome by subfunctionalization of tissue-specific expression. *FEBS J* 272(14):3561–3571. <https://doi.org/10.1111/j.1742-4658.2005.04775.x>
- Liu S, Li Z, Gui J-F (2009). Fish-specific duplicated dmrt2b contributes to a divergent function through hedgehog pathway and maintains left-right asymmetry establishment function. <https://doi.org/10.1371/journal.pone.0007261>
- Lu J, Peatman E, Tang H et al (2012). Profiling of gene duplication patterns of sequenced teleost genomes: evidence for rapid lineage-specific genome expansion mediated by recent tandem duplications. *BMC Genomics* 13(1). <https://doi.org/10.1186/1471-2164-13-246>
- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science (New York, N.Y)* 290(5494):1151–1155. <https://doi.org/10.1126/science.290.5494.1151>
- Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154(1):459–473
- Ma R, Duan Y, Zhang L, Qi X, Zhang L, Pan S, Gao L, Wang C, Wang Y (2022) SCN1A-related epilepsy: novel mutations and rare phenotypes. *Front Mol Neurosci* 15:826183. <https://doi.org/10.3389/fnmol.2022.826183>
- Ma Y, Liu S, Gao J et al (2021) Genome-wide analysis of pseudogenes reveals HBBP1's human-specific essentiality in erythropoiesis and implication in β -thalassemia. *Dev Cell* 56(4):478–49.e11. <https://doi.org/10.1016/j.devcel.2020.12.019>
- Magadum S, Banerjee U, Murugan P et al (2013). Gene duplication as a major force in evolution. In *J Genet* 92:155–161. <https://doi.org/10.1007/s12041-013-0212-8>
- Málaga-Trillo E, Meyer A (2001) Genome duplications and accelerated evolution of hox genes and cluster architecture in teleost fishes I. *Am Zool* 41(3):676–686. <https://doi.org/10.1093/icb/41.3.676>
- Marlétaz F, Firbas PN, Maeso I et al (2018) Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* 564(7734):64–70. <https://doi.org/10.1038/s41586-018-0734-6>
- Martins Custodio H, Clayton LM, Bellampalli R, Pagni S, Silvennoinen K, Caswell R, Brunklaus A, Guerrini R, Koeleman BPC, Lemke JR, Møller RS, Scheffer IE, Weckhuysen S, Zara F, Zuberi S, Kuchenbaecker K, Balestrini S, Mills JD, Sisodiya SM (2023) Widespread genomic influences on phenotype in Dravet syndrome, a 'monogenic' condition. *Brain* 146(9):3885–3897. <https://doi.org/10.1093/brain/awad111>
- McClintock JM, Kheirbek MA, Prince VE (2002) Knockdown of duplicated zebrafish hoxb1 genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. *Development (Cambridge, England)* 129(10):2339–2354. <https://doi.org/10.1242/dev.129.10.2339>
- Meyer A, Schartl M (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol* 11(6):699–704. [https://doi.org/10.1016/S0955-0674\(99\)00039-3](https://doi.org/10.1016/S0955-0674(99)00039-3)
- Molinari NA, Petrov DA, Price HJ, Smith JD, Gold JR, Vassiliadis C, Dudley JW, Biradar DP, Devos KM, Bennetzen JL, Le QH, Schoen DJ, Bureau TE, Blumenstiel JP, Hartl DL, Conery JS, Macpherson JM, Petrov DA, Lisch DR, ... Gaut BS(2008). Synteny and collinearity in plantgenomes. *Science*, pp 486–489. <https://doi.org/10.1126/science.1153917>
- Moosajee M, Hingorani M, Moore AT (1993) PAX6-Related Aniridia. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Bean LJ, Gripp KW, Amemiya A (Eds.) *GeneReviews*®. University of Washington Seattle. <http://www.ncbi.nlm.nih.gov/books/NBK1360/>. Accessed 11 Dec 2022
- Moriyama Y, Ito F, Takeda H et al (2016). Evolution of the fish heart by sub/neofunctionalization of an elastin gene. *Nat Commun* 7. <https://doi.org/10.1038/ncomms10397>
- Moriyama Y, Koshiba-Takeuchi K (2018) Significance of whole-genome duplications on the emergence of evolutionary novelties. *Brief Funct Genomics* 17(5):329–338. <https://doi.org/10.1093/bfgp/ely007>
- Morrow JM, Lasic S, Chang BSW (2011) A novel rhodopsin-like gene expressed in zebrafish retina. *Vis Neurosci* 28(4):325–335. <https://doi.org/10.1017/S0952523811000010>
- Murrell A, Heeson S, Cooper WN, Douglas E, Apostolidou S, Moore GE, Maher ER, Reik W (2004) An association between variants in the IGF2 gene and Beckwith-Wiedemann syndrome: interaction between genotype and epigenotype. *Hum Mol Genet* 13(2):247–255. <https://doi.org/10.1093/hmg/ddh013>
- Nakamura-García AK, Espinal-Enríquez J (2023) Pseudogenes in Cancer: State of the Art. *Cancers* 15(16):4024. <https://doi.org/10.3390/cancers15164024>
- Nocedal I, Laub MT (2022) Ancestral reconstruction of duplicated signaling proteins reveals the evolution of signaling specificity. *eLife* 11 e77346. <https://doi.org/10.7554/eLife.77346>
- Ogawa T, Shirai T, Ogawa T et al (2013). Experimental molecular archeology: reconstruction of ancestral mutants and evolutionary history of proteins as a new approach in protein engineering. In

- Protein Engineering—Technology and Application. <https://doi.org/10.5772/56255>
- Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag, New York
- Oliver PL, Finelli MJ, Edwards B, Bitoun E, Butts DL, Becker EBE, Cheeseman MT, Davies B, Davies KE (2011) *Oxr1* is essential for protection against oxidative stress-induced neurodegeneration. *PLoS Genet* 7(10):e1002338. <https://doi.org/10.1371/journal.pgen.1002338>
- Opazo JC, Butts GT, Nery MF, Storz JF, Hoffmann FG (2013) Whole-genome duplication and the functional diversification of teleost fish hemoglobins. *Mol Biol Evol* 30(1):140–153. <https://doi.org/10.1093/molbev/mss212>
- Otis JP, Zeituni EM, Thierer JH, Anderson JL, Brown AC, Boehm ED, Cerchione DM, Ceasrine AM, Avraham-Davidi I, Tempelhof H, Yaniv K, Farber SA (2015) Zebrafish as a model for apolipoprotein biology: comprehensive expression analysis and a role for ApoA-IV in regulating food intake. *Dis Model Mech* 8(3):295–309. <https://doi.org/10.1242/dmm.018754>
- Ozernyuk N, Schepetov D (2022) HOX-Gene Cluster organization and genome duplications in fishes and mammals: transcript variant distribution along the anterior–posterior axis. *Int J Mol Sci* 23(7):17. <https://doi.org/10.3390/ijms23179990>
- Panchy N, Lehti-Shiu M, Shiu SH (2016) Evolution of gene duplication in plants. *Plant Physiol* 171(4):2294–2316. <https://doi.org/10.1104/pp.16.00523>
- Panopoulou G, Poustka AJ (2005) Timing and mechanism of ancient vertebrate genome duplications—the adventure of a hypothesis. *Trends Genet* 21(10):559–567. <https://doi.org/10.1016/j.tig.2005.08.004>
- Parrie LE, Renfrew EM, Wal AV, Mueller RL, Garrity DM (2013) Zebrafish *tbx5* paralogs demonstrate independent essential requirements in cardiac and pectoral fin development. *Dev Dyn* 242(5):485–402. <https://doi.org/10.1002/dvdy.23953>
- Pascual-Anaya J, D'Aniello S, Kuratani S, Garcia-Fernández J (2013) Evolution of Hoxgene clusters in deuterostomes. *BMC Dev Biol* 13(1):26. <https://doi.org/10.1186/1471-213X-13-26>
- Patton EE, Zon LI, Langenau DM (2021) Zebrafish disease models in drug discovery: from preclinical modelling to clinical trials. *Nat Rev Drug Discovery* 20(8):611–628. <https://doi.org/10.1038/s41573-021-00210-8>
- Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Francisco Carter DR (2011) Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA* 17(5):792–798. <https://doi.org/10.1261/rna.2658311>
- Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* 10(12):1890–1902. <https://doi.org/10.1101/gr.164800>
- Qian SH, Chen L, Xiong Y-L, Chen Z-X (2022) Evolution and function of developmentally dynamic pseudogenes in mammals. *Genome Biol* 23:235. <https://doi.org/10.1186/s13059-022-02802-y>
- Qian W, Liao B-Y, Chang AY-F, Zhang J (2010) Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends Genet* 26(10):425–430. <https://doi.org/10.1016/j.tig.2010.07.002>
- Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH (2019) Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol* 20(1):38. <https://doi.org/10.1186/s13059-019-1650-2>
- Qiu Y, Tay YV, Ruan Y, Adams KL (2020) Divergence of duplicated genes by repeated partitioning of splice forms and subcellular localization. *New Phytol* 225(2):1011–1022. <https://doi.org/10.1111/nph.16148>
- Ranawakage DC, Okada K, Sugio K et al (2021). Efficient CRISPR-Cas9-mediated knock-in of composite tags in zebrafish using long ssDNA as a donor. *Front Cell Dev Biol* 8. <https://doi.org/10.3389/fcell.2020.598634>
- Rastogi S, Liberles DA (2005) Subfunctionalization of duplicated genes as a transition state to neofunctionalization. <https://doi.org/10.1186/1471-2148-5-28>
- Rice AM, McLysaght A (2017) Dosage-sensitive genes in evolution and disease. *BMC Biol* 15(1):78. <https://doi.org/10.1186/s12915-017-0418-y>
- Roberts TC, Morris KV (2013) Not so pseudo anymore: pseudogenes as therapeutic targets. *Pharmacogenomics* 14(16):2023–2034. <https://doi.org/10.2217/pgs.13.172>
- Rochette C, Gilbert N, Simard L (2001) SMN gene duplication and the emergence of the SMN2 gene occurred in distinct hominids: SMN2 is unique to *Homo sapiens*. *Hum Genet* 108(3):255–266. <https://doi.org/10.1007/s004390100473>
- Rodríguez-Marí A, Cañestro C, BreMiller RA, Catchen JM, Yan Y-L, Postlethwait JH (2013) Retinoic acid metabolic genes, meiosis, and gonadal sex differentiation in zebrafish. *PLoS ONE* 8(9):e73951. <https://doi.org/10.1371/journal.pone.0073951>
- Rodríguez-Morales R, Vélez-Negrón V, Torrado-Tapias A et al (2020). Expression patterns of activating transcription factor 5 (*atf5a* and *atf5b*) in zebrafish. *Gene Expr Patterns* 37:119126. <https://doi.org/10.1016/j.gep.2020.119126>
- Roth C, Rastogi S, Arvestad L, Dittmar K, Light S, Ekman D, Liberles DA (2007) Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. *J Exp Zool B Mol Dev Evol* 308B(1):58–73. <https://doi.org/10.1002/jez.b.21124>
- Sambrook JG, Figueroa F, Beck S (2005) A genome-wide survey of major histocompatibility complex (MHC) genes and their paralogues in zebrafish. *BMC Genomics* 6(1):152. <https://doi.org/10.1186/1471-2164-6-152>
- Sant KE, Hansen JM, Williams LM, Tran NL, Goldstone JV, Stegeman JJ, Hahn ME, Timme-Laragy A (2017) The role of *Nrf1* and *Nrf2* in the regulation of glutathione and redox dynamics in the developing zebrafish embryo. *Redox Biol* 13:207–218. <https://doi.org/10.1016/j.redox.2017.05.023>
- Santoriello C, Zon LI (2012) Science in medicine hooked! Modeling human disease in zebrafish. *Sci Med* 122(7):2337–2343. <https://doi.org/10.1172/JCI60434.combines>
- Scholpp S, Brand M (2001) Morpholino-induced knockdown of zebrafish engrailed genes *eng2* and *eng3* reveals redundant and unique functions in midbrain–hindbrain boundary development. *Genesis* 30(3):129–133. <https://doi.org/10.1002/gene.1047>
- Schoonheim PJ, Arrenberg AB, Bene FD, Baier H (2010) Optogenetic localization and genetic perturbation of saccade-generating neurons in zebrafish. *J Neurosci* 30(20):7111–7120. <https://doi.org/10.1523/JNEUROSCI.5193-09.2010>
- Scossa F, Fernie AR (2021) Ancestral sequence reconstruction—an underused approach to understand the evolution of gene function in plants? *Comput Struct Biotechnol J* 19:1579–1594. <https://doi.org/10.1016/j.csbj.2021.03.008>
- Sedletcaia A, Evans T (2011) Heart chamber size in zebrafish is regulated redundantly by duplicated *tbx2* genes. *Dev Dyn* 240(6):1548–1557. <https://doi.org/10.1002/dvdy.22622>
- Sekiya M, Yuhara M, Murayama Y, Ohyama Osawa M, Nakajima R, Ohuchi N, Matsumoto N, Yamazaki D, Mori S, Matsuda T, Sugano Y, Osaki Y, Iwasaki H, Suzuki H, Shimano H (2022) A case of early-onset diabetes with impaired insulin secretion carrying a PAX6 gene *Gln135** mutation. *Endocrinol Diabetes Metab Case Rep* 2022:22–0271. <https://doi.org/10.1530/EDM-22-0271>
- Shehwana H, Konu O (2019) Comparative transcriptomics between zebrafish and mammals: a roadmap for discovery of conserved and unique signaling pathways in physiology and disease. *Front*

- Cell Dev Biol 7. <https://www.frontiersin.org/articles/10.3389/fcell.2019.00005>
- Shukla D, Gural B, Cauley E et al (2022). Duplicated zebrafish (*Danio rerio*) inositol phosphatases *inpp5ka* and *inpp5kb* diverged in expression pattern and function. <https://doi.org/10.1101/2022.08.31.506059>
- Singh NP, Krumlauf R (2022) Diversification and functional evolution of HOX proteins. *Front Cell Dev Biol* 10:798812. <https://doi.org/10.3389/fcell.2022.798812>
- Smith AA, Wyatt K, Vacha J, Vihtelic TS, Jr SZ, J., Wistow, G. J., & Posner, M. (2006) Gene duplication and separation of functions in α B-crystallin from zebrafish (*Danio rerio*). *FEBS J* 273(3):481–490. <https://doi.org/10.1111/j.1742-4658.2005.05080.x>
- Soshnikova N, Dewaele R, Janvier P, Krumlauf R, Duboule D (2013) Duplications of hox gene clusters and the emergence of vertebrates. *Dev Biol* 378(2):194–199. <https://doi.org/10.1016/j.ydbio.2013.03.004>
- Sourbron J, Schneider H, Kecskés A, Liu Y, Buening EM, Lagae L, Smolders I, de Witte P (2016) Serotonergic modulation as effective treatment for dravet syndrome in a zebrafish mutant model. *ACS Chem Neurosci* 7(5):588–598. <https://doi.org/10.1021/acschemneuro.5b00342>
- Storz JF, Natarajan C, Grouleff MK et al (2020) Oxygenation properties of hemoglobin and the evolutionary origins of isoform multiplicity in an amphibious air-breathing fish, the blue-spotted mudskipper (*Boleophthalmus pectinirostris*). *J Exp Biol* 223(Pt 2):jeb217307. <https://doi.org/10.1242/jeb.217307>
- Storz JF, Opazo JC, Hoffmann FG (2013) Gene duplication, genome duplication, and the functional diversification of vertebrate globins. *Mol Phylogenet Evol* 66(2):469–478. <https://doi.org/10.1016/j.ympev.2012.07.013>
- Streisinger G, Walker C, Dower N, Knauber D, Singer F (1981) Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature* 291(5813):293–296. <https://doi.org/10.1038/291293a0>
- Taylor JS, Van de Peer Y, Braasch I, Meyer A (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc B: Biol Sci* 356(1414):1661–1679. <https://doi.org/10.1098/rstb.2001.0975>
- Teame T, Zhang Z, Ran C, Zhang H, Yang Y, Ding Q, Xie M, Gao C, Ye Y, Duan M, Zhou Z (2019) The use of zebrafish (*Danio rerio*) as biomedical models. *Anim Front: Rev Mag Anim Agric* 9(3):68–77. <https://doi.org/10.1093/af/vfz020>
- Thummel R, Enright JM, Kassen SC, Montgomery JE, Bailey TJ, Hyde DR (2010) *Pax6a* and *Pax6b* are required at different points in neuronal progenitor cell proliferation during zebrafish photoreceptor regeneration. *Exp Eye Res* 90(5):572–582. <https://doi.org/10.1016/J.EXER.2010.02.001>
- Tiedke J, Gerlach F, Mitz SA, Hankeln T, Burmester T (2011) Ontogeny of globin expression in zebrafish (*Danio rerio*). *J Comp Physiol [b]* 181(8):1011–1021. <https://doi.org/10.1007/S00360-011-0588-9/METRICS>
- Timme-Laragy AR, Karchner SI, Franks DG, Jenny MJ, Harbeitner RC, Goldstone JV, McArthur AG, Hahn ME (2012) *Nrf2b*, novel zebrafish paralog of oxidant-responsive transcription factor NF-E2-related factor 2 (NRF2). *J Biol Chem* 287(7):4609–4627. <https://doi.org/10.1074/jbc.M111.260125>
- Traherne JA (2008) Human MHC architecture and evolution: implications for disease association studies. *Int J Immunogenet* 35(3):179–192. <https://doi.org/10.1111/j.1744-313X.2008.00765.x>
- Turner KM, Deshpande V, Beyter D et al (2017) Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* 543(7643):122–125. <https://doi.org/10.1038/nature21356>
- Uttley K, Papanastasiou AS, Lahne M et al (2023). Unique activities of two overlapping PAX6 retinal enhancers. *Life Sci Alliance* 6(11). <https://doi.org/10.26508/lsa.202302126>
- Vandepoel K, De Vos W, Taylor JS, Meyer A, Van De Peer Y (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 USA* 101(6):1638–1643. <https://doi.org/10.1073/pnas.0307968100>
- Vauti F, Vögele V, Deppe I, Hahnenstein ST, Köster RW (2021) Structural analysis and spatiotemporal expression of *Atxn1* genes in zebrafish embryos and larvae. *Int J Mol Sci* 22(21):11348. <https://doi.org/10.3390/ijms222111348>
- Vaz R, Hofmeister W, Lindstrand A (2019) Zebrafish models of neurodevelopmental disorders: limitations and benefits of current tools and techniques. *Int J Mol Sci* 20(6) Article 6. <https://doi.org/10.3390/ijms20061296>
- Viguera E, Canceill D, Ehrlich SD (2001) Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J* 20(10):2587–2595. <https://doi.org/10.1093/emboj/20.10.2587>
- Volkert MR, Crowley DJ (2020) Preventing neurodegeneration by controlling oxidative stress: the role of OXR1. *Front Neurosci* 14:611904. <https://doi.org/10.3389/fnins.2020.611904>
- Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ (2012) Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol* 10(12):e1001446. <https://doi.org/10.1371/journal.pbio.1001446>
- Wagner A (2008) Gene duplications, robustness and evolutionary innovations. *BioEssays: News Rev Mol Cell Dev Biol* 30(4):367–373. <https://doi.org/10.1002/bies.20728>
- Wagner GP, Amemiya C, Ruddle F (2003) Hox cluster duplications and the opportunity for evolutionary novelties. *Proc Natl Acad Sci USA* 100(25):14603–14606. <https://doi.org/10.1073/pnas.2536656100>
- Welch JD, Baran-Gale J, Perou CM, Sethupathy P, Prins JF (2015) Pseudogenes transcribed in breast invasive carcinoma show subtype-specific expression and ceRNA potential. *BMC Genomics* 16(1):113. <https://doi.org/10.1186/s12864-015-1227-8>
- Weuring WJ, Hoekman JW, Braun KPJ et al (2022). Genetic and functional differences between duplicated zebrafish genes for human SCN1A. *Cells* 11(3) Article 3. <https://doi.org/10.3390/cells11030454>
- Weuring WJ, Singh S, Volkens L et al (2020). *NaV1.1* and *NaV1.6* selective compounds reduce the behavior phenotype and epileptiform activity in a novel zebrafish model for Dravet Syndrome. *PLOS ONE* 15(3) e0219106. <https://doi.org/10.1371/journal.pone.0219106>
- Winkler C, Schäfer M, Duschl J, Scharl M, Volff J-N (2003) Functional Divergence of two zebrafish *Midline* growth factors following fish-specific gene duplication. *Genome Res* 13(6a):1067–1081. <https://doi.org/10.1101/gr.1097503>
- Wong L, Weadick CJ, Kuo C, Chang BS, Tropepe V (2010) Duplicate *dmx1* genes regulate progenitor cell cycle and differentiation during zebrafish midbrain and retinal development. *BMC Dev Biol* 10:100. <https://doi.org/10.1186/1471-213X-10-100>
- Wu A, Zhang Q, Lambert G, Khin Z, Gatenby RA, Kim HJ, Pourmand N, Bussey K, Davies PCW, Sturm JC, Austin RH (2015) Ancient hot and cold genes and chemotherapy resistance emergence. *Proc Natl Acad Sci USA* 112(33):10467–10472. <https://doi.org/10.1073/pnas.1512396112>
- Xu AF, Molinuevo R, Fazzari E et al (2023). Subfunctionalized expression drives evolutionary retention of ribosomal protein paralogs *Rps27* and *Rps27l* in vertebrates. *eLife* 12 e78695. <https://doi.org/10.7554/eLife.78695>
- Xu D-M, Chai F-R, Liang X-F, Lu K (2024) Knockout of *lws1* in zebrafish (*Danio rerio*) reveals its role in regulating feeding and

- vision-guided behavior. *Funct Integr Genomics* 24(2):62. <https://doi.org/10.1007/s10142-024-01333-y>
- Xu H, Jiang Y, Li S, Xie L, Tao Y-X, Li Y (2020) Zebrafish *Oxr1a* knockout reveals its role in regulating antioxidant defenses and aging. *Genes* 11(10):1118. <https://doi.org/10.3390/genes11101118>
- Xu H, Wang G, Chi Y-Y et al (2021). Expression profiling and functional characterization of the duplicated *Oxr1b* gene in zebrafish. *Comparative Biochemistry and Physiology. Part D Genomics Proteomics* 39 100857. <https://doi.org/10.1016/j.cbd.2021.100857>
- Yamada K, Maeno A, Araki S et al (2021) An atlas of seven zebrafish *hox* cluster mutants provides insights into sub/neofunctionalization of vertebrate *Hox* clusters. *Development* 148(11):dev198325. <https://doi.org/10.1242/dev.198325>
- Yan Y-L, Willoughby J, Liu D, Crump JG, Wilson C, Miller CT, Singer A, Kimmel C, Westerfield M, Postlethwait JH (2005) A pair of *Sox*: distinct and overlapping functions of zebrafish *sox9* co-orthologs in craniofacial and pectoral fin development. *Development* 132(5):1069–1083. <https://doi.org/10.1242/dev.01674>
- Yang Y-J, Wang Y, Li Z, Zhou L, Gui J-F (2017) Sequential, divergent, and cooperative requirements of *Foxl2a* and *Foxl2b* in ovary development and maintenance of zebrafish. *Genetics* 205(4):1551–1572. <https://doi.org/10.1534/genetics.116.199133>
- Zhang J (2003) Evolution by gene duplication: An update. *Trends Ecol Evol* 18(6):292–298. [https://doi.org/10.1016/S0169-5347\(03\)00033-8](https://doi.org/10.1016/S0169-5347(03)00033-8)
- Zhu H, Zhang X, Xu S, Wu J, Hou M, Zhao H, Zhou Q, Zhong X (2022) Gene duplication, conservation, and divergence of activating transcription factor 5 gene in zebrafish. *J Exp Zool B Mol Dev Evol* 338(5):301–313. <https://doi.org/10.1002/jez.b.23124>
- Zou S, Kamei H, Modi Z, Duan C (2009) Zebrafish IGF genes: gene duplication, conservation and divergence, and novel roles in midline and notochord development. *PLoS ONE* 4(9):e7026. <https://doi.org/10.1371/journal.pone.0007026>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.