POLYTECHNIC UNIVERSITY OF MARCHE

Research Doctorate in Life and Environmental Sciences

Curriculum Marine Biology and Ecology

Cycle XXXV

Environmental adaptation of ray-finned fish: new insights on the balance between transposable elements and their silencing mechanisms

Supervisors

Prof. Adriana Canapa

Prof. Maria Assunta Biscotti

PhD candidate

Elisa Carotti

Summary

1. Aim **3**

2. Introduction **5**

1.1 Genome composition 6

1.2 Transposable elements 8

1.2.1 History: Barbara McClintock and the "jumping gene" 8

1.2.2 Classification of TEs 10

1.2.3 Tools of TE analysis 13

1.3 Regulation of TE transcription 16

1.4 TE and environment: the influence of abiotic factors 20

1.5 Evolutionary success of vertebrates and their mobilome 23

1.5.1 Environmental changes and teleost fish species 29

3. Articles **35**

3.1 Shedding light upon the complex net of genome size, genome composition and environment in chordates 36

3.2 Mobile elements in ray-finned fish genomes 48

3.3 Investigation of the activity of transposable elements and genes involved in their silencing in the newt *Cynops orientalis*, a species with a giant genome 70

3.4 Transposable elements and teleost migratory behaviour 82

3.5 Transcriptional contribution of transposable elements in relation to salinity conditions in teleosts and silencing mechanisms involved 95

3.6 Transposable element tissue-specific response to temperature stress in the stenothermal fish *Puntius tetrazona* 110

4. Analyses on proteins involved in NuRD complex recruitment in fish **122**

5. Discussion and conclusions **130**

6. References **138**

1.Aim

The evolutionary success of vertebrates lies in the functional and structural complexity of their genomes. An increasing number of papers has provided support to the observation, initially made by Barbara McClintock (1956), that transposable elements (TEs) play a key role as drivers of genome diversity and hence of speciation. Indeed, TEs are known to influence the plasticity and structure of the genome thanks to their transposition and recombination ability. Vertebrate genomes are largely made up of TEs that contribute to the genome size and architecture. During evolution, vertebrates have faced significant environmental changes and transitions leading to extinction or adaptation and speciation. These events of environmental instability were accompanied by transposition bursts likely due to impaired TE silencing mechanisms.

In line with these premises, the aim of this PhD thesis was to investigate the impact of TEs in genome composition and how their activity and related silencing mechanisms respond to variation of abiotic factors. I focused my attention on ray-finned fish, adapted to different ecological niches and responsive to environmental changes. For comparison, basal sarcopterygian species, interesting for their genome size and phylogenetic position, were also investigated.

2.Introduction

1.1 Genome composition

In eukaryotes, the haploid DNA content (C-value) varies widely across lineages without an apparent correlation with the complexity of organisms. This incongruity has been called the C-value paradox and has been solved by demonstrating not all DNA is made up of genes but, on the contrary, most of it is constituted by non-coding DNA and often by repetitive DNA. This kind of DNA includes sequences present in multiple copies in the genome and can account for up to 90% of the genome size in some species (Biscotti et al. 2015; Lopez-Flores and Garrido-Ramos 2012). Repetitive DNA can be grouped into two main types: tandem repeats (TRs) and interspersed repeats that include transposable elements (TEs). Tandem repeats are sequences ranging from few to mega bases and are classified, based on the size of the repeat, in three major classes: microsatellites (short repeat units, usually <10 bp), minisatellites (longer units, between 10 and 100 bp), and satellites (units >100 bp). Mini and microsatellites have gained increasing attention over the past decade due to their contribution to intraspecies genetic diversity and use as genetic markers in population genetic studies. Satellite DNAs (satDNAs) have key roles in centromere function, heterochromatin formation and maintenance, and chromosomal pairing. Indeed, they can be found predominantly in pericentromeric, subtelomeric, and interstitial regions,

forming building blocks of constitutive heterochromatin that constitute the centromeres and telomeres. Moreover, satDNAs promote genome plasticity and modulate genomic architecture by promoting rearrangements (Plohl et al. 2014; Louzada et al. 2020; Heslop-Harrison and Schwarzacher 2013).

As regard the other class of repetitive elements, TEs are ubiquitous in both prokaryotes and eukaryotes. These genetic elements can mutate genomes by transposing to new locations or by facilitating homology-based recombination due to their abundance in the genome (Ahmed and Liang 2012). In vertebrates, the increasing number of sequenced genomes has shown that differences in genome size between lineages are ascribable to a variation in transposon content. These mobile elements, previously perceived as "junk DNA" or "selfish DNA", are now recognized as the major players in shaping genomes. During vertebrate evolution, TEs have been repeatedly co-opted and exapted to generate regulatory sequences, coding exons, or entirely new genes that lead to evolutionary advantages for the host. Moreover, TEs are also responsible for substantial rearrangements such as insertions, deletions, inversions, and duplications potentially associated with, or following, speciation events (Biscotti et al. 2019).

1.2 Transposable elements

Among repetitive sequences, TEs are genetic elements characterized by the ability to insert themselves in novel genome locations of the host and to increase in number by replication. Therefore, they are responsible for generating genetic variability on which natural selection acts and they can create pronounced differences in genome size (Garrido-Ramos 2017).

1.2.1 History: Barbara McClintock and the "jumping gene"

Some of the most famous genetic discoveries have been made using model organisms that are easily exploited by scientists due to their wide availability and ease of maintenance and proliferation. Around the 1950s Barbara McClintock, an American biologist, used the *Zea mays* (maize), a perfect organism for genetic analyses since each grain is an embryo produced by an individual fertilization, and thus hundreds of descendants can be marked on a single ear. This paved the way for the study of cytogenetics.

Together with Harriet Creighton, she provided the first experimental evidence that genes were physically located on chromosomes studying the genetic recombination. In particular, she described the phenomenon called crossing-over that is exchange and genetic recombination between parts of chromosomes during meiosis, which is fundamental for understanding genetic

variability. This was in line with the ideas proposed by Thomas Hunt Morgan who suggested the link between genetic traits and the exchange of genetic material between chromosomes. Thus, she discovered the existence of transposons, i.e. portions of DNA capable of moving from one chromosome to another.

McClintock's work was ground-breaking, suggesting the genome of organisms is subject to alteration and reorganisation. This concept was met with criticism but then in 1983 she was awarded the Nobel Prize for her study on the role of transposons and also for other contributions to the field of genetics.

Moreover, Barbara McClintock was the first scientist to investigate epigenetics or heritable changes in gene expression. In particular, she demonstrated that genes can be expressed and silenced during mitosis. This theory was proposed before understanding the molecular structure of DNA and the mechanisms of epigenetics, confirming herself as an innovator in this field (Pray and Zhaurova 2008).

1.2.2 Classification of TEs

TEs are DNA sequences capable of replicating, moving, and integrating into new regions of the genome. On the basis of their transposition mechanism,

TEs can be distinguished into Class I retroelements or retrotransposons and Class II DNA transposons (Goerner-Potvin and Bourque 2018; Bourque et al. 2018) (Figure 1).

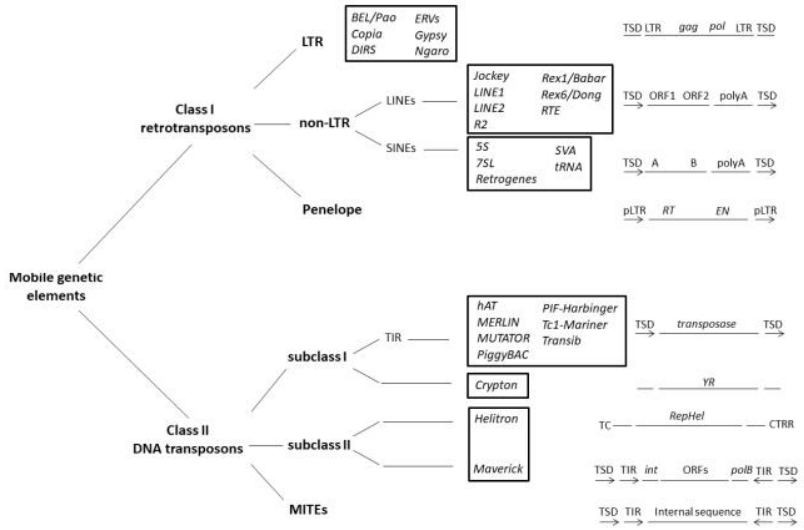

Figure 1. Mobile genetic element classification based on transposition mechanisms (Carducci et al. 2020).

Class I elements transpose via RNA intermediaries and are characterized by a copy and paste transposition mechanism. Their RNA intermediate is reverse-transcribed into its complementary DNA by a reverse transcriptase (RT) encoded by the mobile element. Reverse transcription is followed by reintegration of the new copy into the host genome. Through the copy and paste mechanism of transposition, Class I elements are the main source of increased repetitive fractions, thereby having a major impact in large genomes. Class I mobile elements are composed of long terminal repeat (LTR) and non-LTR subclasses. LTR retrotransposons are characterized by long terminal repeats that confer the ability to transpose. LTR

retrotransposons are structurally composed of *gag* and *pol* genes: the former encodes viral structural particles and the latter encodes the complete retrotranscription machinery (reverse transcriptase, ribonuclease H, and integrase). In contrast to LTR retrotransposons, exogenous retroviruses possess the *env* gene, which encodes the viral envelope. However, traces of the *env* gene have been found also in LTR retrotransposons. *Dictyostelium* Intermediate Repeat Sequence (DIRS), considered the most complex LTR retroelements, is structurally characterized by a tyrosine recombinase (YR) instead of an integrase and by inverted terminal repeats. non-LTR retroelements include Long Interspersed Nuclear Elements (LINEs) and Short Interspersed Nuclear Elements (SINEs). The former are autonomous retroelements constituted by two open reading frames (ORFs) and a poly A tail at the 3' end. Generally, ORF2 encodes a reverse transcriptase and an endonuclease protein. In contrast, SINEs are RT-lacking retroelements and need RT encoded by autonomous elements to transpose. Finally, another group of Class I elements, Penelope retroelements, must be considered separately due to their very large diversity in terms of structural features. The common components are pseudo-LTRs (pLTRs), a reverse transcriptase, and an endonuclease.

Class II mobile elements use a DNA intermediate to transpose their genomic DNA copies into a novel chromosomal position and can be divided into subclasses I and II. Subclass I consists of two main elements: TIR and Crypton. TIRs are autonomous elements characterized by terminal inverted repeats (TIRs) and a transposase through which transposition occurs via a cut and paste mechanism, in which both DNA strands are cleaved. The DNA transposons hAT, Merlin, Mutator, PiggyBac, PIF-Harbinger, Tc1-Mariner, and Transib can be found in this subclass. Crypton elements use a tyrosine recombinase (YR) in the transposition mechanism probably involving recombination between a circular intermediate and the DNA target. Helitrons and Maverick are the two major representative elements of subclass II. These DNA elements transpose via a copy and paste mechanism. Helitron DNA transposons replicate using a rolling-circle mechanism and encode for a replication initiation (Rep) and a DNA helicase (Hel), while Maverick transposons encode for an integrase, an ORF, and a polymerase B. For polymerase B, transposition involves a single-strand excision phase, extrachromosomal replication, and consequent reintegration into a new location. Miniature Inverted Transposable Elements (MITEs), also grouped in Class II, do not encode a transposase, therefore, they exploit transposases encoded by autonomous elements to move throughout the genome.

1.2.3 Tools of TE analysis

The advent of high-throughput sequencing technologies and bioinformatics has provided a wealth of genomic data that is extremely useful for obtaining more precise information about the genomes of model and non-model organisms.

The identification of TEs and their annotation are extremely important to characterize these elements in the genome. Various tools have been developed in relation also to the availability of the assembled genome. With an assembled genome, repository-based annotation can be performed, or alternatively, in absence of genome assembly, *de novo* annotation using raw reads can be used. While the repository-based RepeatMasker is currently the best choice for TE annotation, *de novo* approaches are excellent for identifying new TE families.

Given the importance of classifying items *de novo*, such tools can be used to achieve increasingly comprehensive and complete results. The performance of repository based annotation tools is linked to the quality and specificity of the databases used. The most widely used tool is RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) that screens DNA sequences for interspersed repeats and low complexity DNA sequences. The

output provided by this software is a detailed annotation of the repeats that are present in the query sequence and a modified version of the query sequence in which all the annotated repeats have been masked. Currently, over 56% of human genomic sequence is identified and masked by the program. Sequence comparisons in RepeatMasker are performed by one of several popular search engines such as nhmmer, cross_match, ABBlast/WUBlast, RMBlast and Decypher. RepeatMasker uses curated libraries of repeats and currently supports Dfam (profile HMM library derived from Repbase sequences) and Repbase, a service of the Genetic Information Research Institute.

To annotate newly sequenced genomes and identify missing TEs from RepeatMasker genome annotations, most popular tools for *de novo* annotation using assembled genomes are RECON (Bao and Eddy 2022), RepeatScout (Price et al. 2005), RepeatModeler (Smit and Hubley 2018) and phRAIDER (Schaeffer et al. 2016). Despite they might include false positives, a high percentage of TEs and novel families are annotated by these tools. For example, in the human genome an additional 10% or more is assigned to TEs.

De novo annotation tools use low coverage sequencing data to assemble TEs and other repetitive sequences directly from raw reads. One of the first tool was RepeatExplorer (Novák et al. 2013) initially published as algorithm to identify TEs from next generation sequencing (NGS) reads, then used as

cluster with RepeatMasker. In line with this principle, dnaPipeTE (Goubert et al. 2015) was designed to find, annotate, and quantify TEs in small samples of NGS datasets. It is very useful to quantify the proportion of TEs in newly sequenced genomes since it does not require genome assembly and works on small datasets. Finally, with REPdenovo (Chu et al. 2016) it is possible to assemble long repeats and identify missing reference TE. Recently, RepLong (Guo et al. 2018) was developed for long-read *de novo* TE annotation. This tool is the most sensitive by assigning 35% of the human genome to repeats.

Globally, TE assembly tools here described are excellent methods for annotating TEs in all those organisms for which there is no reliable reference genome. Tools used in repository-based annotation or *de novo* annotation of assembled genomes are more sensitive than these latter methods. However, they are able to assign a higher percentage of the genome to repeats including TEs that are absent in RepBase (Goerner-Potvin and Bourque 2018).

1.3 Regulation of TE transcription

TEs can be co-opted to benefit the host and be a source of genetic diversity and regulatory innovations. However, through insertions in coding-gene regions, TEs can have deleterious effects that can lead to decreased host fitness. To counteract the threat posed by TE transposition and new insertions,

different lineages have evolved protective mechanisms. For example, during the first days of embryogenesis TEs are repressed through RNA- and protein-based epigenetic mechanisms.

Kruppel-associated zinc finger box proteins (KRAB-ZFPs) constitute a large family of transcription factors that first emerged more than 350 million years ago in the genomes of higher vertebrates. These are encoded by hundreds of genes and are involved in TE silencing, including endogenous retroviruses, LINE retroelements, and SINEs (Ecco et al. 2016). During the expansion of this gene family, they underwent strong positive selection at positions encoding for amino acids predicted to determine DNA-binding specificity, consistent with a role in countering rapidly mutating genetic invaders (Jacobs et al. 2014; Emerson and Thomas 2009). KRAB-ZFPs contain in the C-terminal an array of zinc fingers motifs to bind DNA and in the N-terminal the KRAB domain to recruit their co-repressor, belonging to the TRIM family, Tripartite Motif Containing 28 (TRIM28) protein also called KRAB associated protein 1 (KAP1). This protein includes the RING, B-box, coiled-coil (RBCC) motif consisting of a RING domain with ubiquitin E3 ligase activity, two B-box type zinc fingers, and a coiled-coil domain. The KAP1 binding surface of the KRAB domain is enriched in exposed hydrophobic amino acids and recognizes a hydrophobic patch in the core region of the

KAP1 coiled domain. The hydrophobic core is surrounded by a network of polar interactions and salt bridges. Recently, Stoll and colleagues (2022) have evidenced that the amino acids involved in the binding with KAP1 (Asp8, Val9, Ile11, Phe13, Glu17, Trp18, Leu21, Tyr29, Val 32, and Met33) are conserved and required for silencing. KRAB domain forms contacts with both the subunits of the KAP1 dimer by binding in different regions. The amino acid sequences of the KRAB-ZFP KRAB-A box are highly conserved in tetrapod vertebrates, with the exception of a small KRAB-ZFP subset, which appears to have acquired KAP1-independent functions (Helleboid et al. 2019).

The central part of KAP1 contains a PxVxL motif that recruits heterochromatin protein 1 (HP1), essential for transcriptional silencing, and the DNA Methyltransferases (DNMT1 and DNMT3A), which allow the deposition of transcriptional repressive marks. In the C-terminal region it is present a PHD bromodomain that recruits the SUMO E2 ligase Ubc9 to SUMOylate several lysines in the bromodomain. This post-translational modification is required for the recruitment and activation of the histone methyltransferase SETDB1, which adds the H3K9me3 mark, and the recruitment of the Nucleosome Remodelling and Deacetylase (NuRD) complex (Stoll et al. 2019). This latter is one of the major chromatin

remodelling complexes found in cells and contains the histone deacetylases 1 and 2 (HDAC1 and HDAC2), the chromatin helicase DNA binding protein 4 (CHD4), the Retinoblastoma-binding protein 4 and 7 (Rbbp4 and Rbbp7), either the zinc-finger proteins GATA Zinc Finger Domain Containing 2A or 2B (GATAD2a and GATAD2b), two Metastasis-associated Proteins (MTA1, MTA2, and/or MTA3), and the Methyl-CpG Binding Domain Protein 2 or 3 (MBD2 or MBD3). Although this mechanism was initially presumed to only act during early development, an increasing number of papers suggests that it may be also functional in adult tissues (Carotti et al. 2022; Carotti et al. 2023; Carducci et al. 2021; Pappalardo et al. 2021).

Another conserved mechanism of TE silencing is based on the action of Argonaute proteins (AGOs) that use small non-coding RNAs (ncRNAs), such as small interfering RNAs (siRNAs) and microRNAs (miRNAs), to degrade the mRNAs, to repress translation, and to form heterochromatin. These small RNAs are produced through the activity of DROSHA and DICER proteins. The former acts in the nucleus, loading double-stranded RNA substrates, helped by the DiGeorge Syndrome Critical Region Gene 8 (DGCR8), and modifying them to produce miRNA precursors. DICER cleaves these molecules as well as endogenous or exogenous siRNAs in the cytoplasm (Carducci et al. 2021). AGOs interact with trinucleotide repeat containing

adaptor 6 (TNRC6) called also multiple glycine/tryptophan repeats (GW or WG). This protein, together with Ago proteins, exerts its silencing activity in siRNA and miRNA pathways via multiple domains. Indeed, the Ago2-GW182 interaction is critical for the localization of Ago2 in the cytoplasmic foci and its repression function. Knockdown either GW182 or Ago2 causes loss of silencing activity correlating with the disassembly of cytoplasmic foci named GW bodies (GWBs) (Yao et al. 2013). GWBs are enriched in proteins that are involved in mRNA degradation, repressing translation, and enhancing mRNA turnover. Recent works have reported that exogenously introduced human Ago2 is also enriched in GW bodies, indicating that RNA interference function may be somehow linked to these structures (Braun et al. 2013; Jakymiw et al. 2005).

1.4 TE and environment: the influence of abiotic factors

Various interactions exist between genome size, transposons and base composition, as well as between these and the environment. In some cases, the environmental factors would seem to be the direct cause of changes in size and/or genome composition, while in other cases genome characters influence some cell structures and functions (Olmo 1983; Vinogradov 1999); thus, these relationships may not be always simple and linear.

The evolutionary success of species is strictly related to their ability to adapt to the environment and to cope with changes of environmental conditions. Phenotypic plasticity is the main mechanism used by organisms to response to rapid environmental perturbations. Indeed, it can produce well-adapted phenotypes that in turns might improve the fitness of organisms. Phenotypic plasticity is also the result of environmental sensitivity of genome that through modifications in the heterochromatin structure varies the expression not only of genes but also of TEs.

Recently, an increasing number of papers has reported the influence of abiotic factors such as temperature, salinity, and pH on TE activity, suggesting a role of mobile elements in the regulation of mechanisms responsible for environmental adaptation. Indeed, changes in the abiotic factors can cause variations in the epigenetic status leading to gene activation but also to the impairment of transposon silencing mechanisms inducing the activation of TEs with possible creation of genetic variability. Therefore, TEs represent a powerful adaptive response to environmental perturbations.

Recently, many researches on transposons have shown that their activity is stimulated by various environmental stressors (Canapa et al. 2015; Carducci et al. 2019b; Pappalardo et al. 2021), such as temperature, thermal shocks, radiations, chemicals, and viral infections (Capy et al. 2000; Fujino et al.

2021; Garcia Guerreiro 2012; Carducci et al. 2019b). In some cases, these stressors would act directly on transposons with a mechanism typical of defensive gene activation, since some transposons have sequences similar to the promoters of these genes (Takeda et al. 1999; Capy et al. 2000). However, in many other cases, the action would mainly depend on a temperature-dependent activation of transposase (Capy et al. 2000; Fujino et al. 2021; Carducci et al. 2019b). Besides to be involved in transposon activation, temperature is one of the most important elements of natural selection able to interact both with base composition and genome size. Studies conducted on base composition have clearly shown that body temperature affects GC percentage. This effect is evident in structural changes of macromolecules. In accordance with the "thermodynamic stability hypothesis", higher GC values are directly related to an increased DNA thermostability (Bernardi 2004) which in turn influences RNA composition and chemical and physical properties of proteins, which in species showing this genomic feature are richer in hydrophobic amino acids (Canapa et al. 2020).

In addition to temperature, salinity is an abiotic factor that influences the adaptation of marine and freshwater organisms. Variations in this parameter can represent a possible stress that threatens the survival of aquatic species. In teleosts, the physiology and morphology but also the genomic composition

are influenced by changes of environmental salinity due also to the migratory ability for some species (Palzenberger and Pohla 1992; Friedman et al. 2012; Evans et al. 2005). A recent analysis of five teleost species (i.e. zebrafish, medaka, stickleback, takifugu, and pufferfish) has showed that small differences of the average genome base composition hide great differences at the genome organization level. Furthermore, within the genome of the green spotted pufferfish the GC-rich genes showed higher transcriptional levels than GC-poor ones. Comparisons between migratory and non-migratory species have showed GC%, gill area, and metabolic rate significantly higher than non-migratory species (Tarallo et al. 2016).

The relationship between TE activation, environmental changes, and species adaptation is still poor understood and many questions remain unsolved. This topic is extremely interesting to be explored in fish species that are sensitive to variations in the environment even more in the context of climate changes.

1.5 Evolutionary success of vertebrates and their mobilome

Functional and structural complexity of the genome allowed the evolutionary success of vertebrates. One of the key components is represented by TEs, considered the major drivers of genomic and biological diversity. Their activity is responsible for genome expansion and chromosomal

rearrangements (Rebollo et al. 2010; Zhang et al. 2011; Auvinet et al. 2018), but also generates regulatory sequences, coding exons or entirely new genes that may represent evolutionary advantages for the host (Carducci et al. 2019a). It is important to underline that the transposon content varies considerably between lineages as I reviewed in the papers number 3.1 and 3.2 presented in the section Articles. Indeed, birds and pufferfishes present low amount of TEs while in elephant shark, lamprey, coelacanth, *Xenopus*, mammals, and non-bird reptile genomes mobile elements are widely represented (>20%). Moreover, TE composition varies among the genomes of vertebrates: DNA transposons are prevalent in *Xenopus* and some teleosts (tilapia, platy fish, medaka, cod, zebrafish, and European eel); LINE and SINE retroelements prevail in non-bony vertebrates, some actinopterygian fish (fugu and spotted gar), coelacanth, chicken, and mammals (Chalopin et al. 2015).

The differences in TE type and content in vertebrates might be the results of the significant transitions and severe ecosystem changes that have characterized the evolution of this taxon. One of the most important step was the transition from water to land. The terrestrial vertebrates would have derived from fossil forms of lobe-finned fishes as coelacanths and dipnoans, dated in the Devonian (400 Mya) (Cloutier and Ahlberg 1996). The first

living coelacanth (*Latimeria chalumnae*) was found off the estuary of river Chalumna, in South Africa and represents the only unique extant representative of the lineage from which dipnoans (lungfish) and tetrapods should have arisen (Amemiya et al. 2013; Pallavicini et al. 2013). This discovery was a great surprise since around the 1940s only two sarcopterygian taxa, dipnoans and tetrapods, were considered to have survived post-Devonian extinction.

With 26 TE superfamilies in its genome, coelacanths present the highest TE richness, in which LINE and SINE retroelements are dominant and have had at least one major middle-aged transposition burst and some recent LINE copies (Chalopin et al. 2015). While coelacanths have a genome size comparable to that of human (3.5 Gb), lungfish are characterized by very large genomes (40-130 Gb) (Figure 2) (Wang et al. 2021; Amemiya et al. 2013). Moreover, the activity of TEs in these two basal sarcopterygians showed clear differences. Indeed, in lungfish *Protopterus annectens* it has been recorded a lower TE activity than coelacanths, in line with the hypothesis that its genome is mainly made up of no active mobile elements (Biscotti et al. 2016; Biscotti et al. 2017; Metcalfe et al. 2012; Metcalfe and Casane 2013). This can be explained by the hypothesis of an ancient burst of

transposition followed by a long period of degeneration creating a “cemetery of TEs” (Wang et al. 2021; Sirijovsky et al. 2005).

Among vertebrates, also salamanders (Amphibia, Caudata) present large genomes ranging from 13.89 to 120.60 Gb. Indeed, Caudata shows a great variation in genome size also considering metamorphic and neotenic species, with the smallest and the largest genomes, respectively (Biscotti et al. 2020). In this taxon as in lungfish, the transition to land was accompanied by a high proliferation of TEs leading to organisms with giant genomes (Carducci et al. 2019a; Metcalfe et al. 2012; Sun et al. 2012b; Canapa et al. 2015; Biscotti et al. 2016). This feature in salamander genomes has been explained by observing a reduced DNA loss rate and the accumulation of TEs, in particular of LTR retrotransposons (Sun et al. 2012a). Moreover, the genomic expansion in lungfish lineage, dated in the Carboniferous (360-320 Mya), occurred independently (Meyer et al. 2021; Liedtke et al. 2018) and earlier than that of amphibian lineage, started in the Permian (280 Mya) up to the Middle Jurassic (208 Mya) (Laurin et al. 2016; Organ et al. 2016; Liedtke et al. 2016). This is in line with data obtained during my PhD project showing that, differently from lungfish, the mobilome of the newt *Cynops orientalis* presents younger TE copies responsible for the higher values of activity observed in this species (Biscotti et al. 2020). Moreover, in the transcriptome,

the major impact was due to non-LTR retroelements. Furthermore, by evaluating TE silencing pathways, we confirmed that the NuRD complex was active also in adults and that their expression was elevated in the gonads to preserve genome integrity. The findings obtained are presented in the paper number 3.3 in the section Articles.

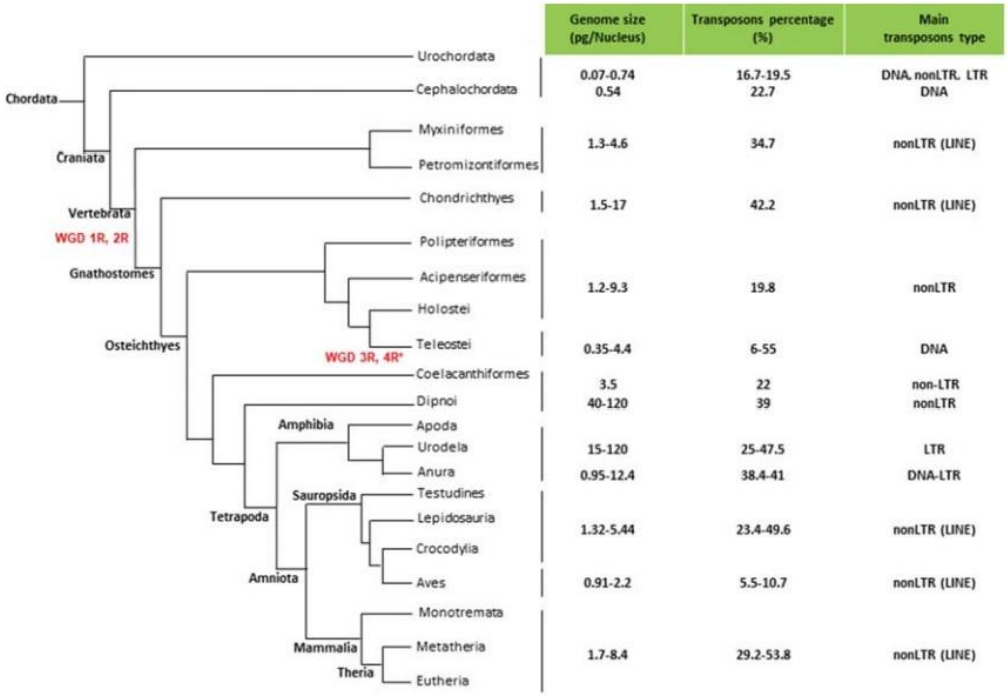

Figure 2. Cladogram showing evolutionary relationships between the main lineages of chordates. Whole Genome Duplication (WGD) events in vertebrate evolution are indicated in red: 1R and 2R occurred before the divergence of Vertebrata, 3R in teleosts and 4R* in salmonids, cyprinids, and catfishes (Canapa et al. 2020).

Among vertebrates, actinopterygians are one of the most successful and diverse groups with over 30,000 species (www.fishbase.org). Several molecular phylogenetic analyses have shown that polypterans are a basal lineage of ray-finned fishes. The order Acipenseriformes includes two

families: Acipenseridae (sturgeons) and Polyodontidae. Holostei and Teleostei are part of the Neopterygii, a group of fish that appeared in the late Permian and is characterized by improved swimming skills and feeding mechanisms that allowed the colonization of a wider range of habitats. Teleosts appeared in the late Triassic and their evolutionary radiation occurred during the Mesozoic and Cenozoic. As in other gnathostomes, their genome has undergone two rounds of whole genome duplication (WGD) (Figure 2) (Dehal and Boore 2005; Postlethwait et al. 2000; Vandepoele et al. 2004). In addition, teleosts have undergone a third specific event led to the duplication of genes that were probably responsible for their radiation clades (Vandepoele et al. 2004; Christoffels et al. 2004; Meyer and Van de Peer 2005). A further fourth round occurred independently in Neotropical Corydoradinae catfishes between 54-66 Mya (Marburger et al. 2018), in salmonids ~88-103 Mya (Alexandrou et al. 2013; Macqueen and Johnston 2014), and in cyprinids ~5.6-11.3 Mya (Wang et al. 2012). Teleosts are the most successful ray-finned fish with over 24,000 species. This might be linked to the high TE diversity that characterizes their genomes. Therefore, during my PhD thesis, I have analysed the TE genome contribution and landscape in 24 teleost fish species. Among these, some specie (*Anguilla* genus, *Salmo salar*, *Thunnus orientalis*) presented a balance between Class I

retroelements and Class II DNA transposons while in others, as *Cyprinus carpio*, *Danio rerio*, *Scartelaos histophorus*, *Neogobius melanostomus*, and *Astyanax mexicanus*, DNA transposons prevail. Moreover, the major impact of LTR retroelements were emerged in *Araipama gigas*, *Tenualosa ilisha* and *Lates calcarifer*. The findings obtained are presented in the paper number 3.4 in the section Articles.

1.5.1 Environmental changes and teleost fish species

Teleosts populate a wide range of habitats across the world, from tropical to polar regions of both sea water (SW) and freshwater (FW). However, some species are characterized by the ability to migrate, between these two environments, during their lifespan. Therefore, they are an excellent model to investigate the relationship between environmental parameters and TEs.

Recently, an increasing number of works has reported a link between the genomic abundance of TEs and genetic elements involved in mechanisms responsible for environmental adaptation. In general, fish and in particular those that migrate have to face changes in abiotic factors such as temperature, salinity, and pH (Casacuberta and González 2013; Makarevitch et al. 2015; Schmidt and Anderson 2006; Carducci et al. 2019b; Pappalardo et al. 2021).

Migration can occur within freshwater for potamodromous, marine water for oceanodromous, or between these two environments for diadromous. In the

latter case, the migratory behaviour is carried out for reproductive purposes and organisms have to face, in a defined stage of their life cycle, variations in salinity that require changes in the osmotic regulation. Among fish, this aspect is better tolerated by euryhaline species *sensu stricto* that have acquired the ability to adapt to various salinity conditions.

To cope with aquatic environments characterised by differences in salinity and temperature, migrating organisms have evolved an extraordinary physiological plasticity. Diadromous fish has always attracted scientific attention due to their spectacular migratory routes between SW and FW and some species are also of commercial importance for fishing activities. Three different types of diadromy have been recognised: anadromy and catadromy are related to spawning and differ for the passage from SW to FW and vice versa, respectively; a third typology, still debated among fish biologists, is amphidromy, for which migration to SW is not correlated with spawning necessity and occurs in a specific life stage (hatched larvae) for a restricted period of time (up to 200 days), after which juveniles return to FW, where they spend their adult life. Catadromous species belong to *Anguilla* genus that includes 19 species displaying a complex catadromous life cycle, with oceanic migrations ranging from a few hundred to thousands of kilometres depending on the species. Satellite systems, used to document the oceanic

migration route up to 1300 km off the European coasts, have shown that they perform daily vertical migrations between depths of 200 and 1000 m. They appear to swim into shallower, warmer waters during the night (average of 282 m and 11.7 °C), while at dawn they descend into deeper, colder waters (average of 564 m and 7-10 °C). The complexity of their life cycle includes major morphological changes. First, the larvae hatch as leptocephalus, which has a laterally compressed body and looks like a leaf with a small head. These first larvae are carried by ocean currents to continental coasts, where they transform into glass eels. At this stage, they show the eel-like shape but are thin, small, and non-pigmented. Subsequently, glass eels migrate to coastal waters and transform into pigmented eels, which subsequently migrate to inland waters and become yellow eels. In this stage, the eels undergo a phase of sedentary and freshwater feeding before entering the silver eel stage (called silvering). Silvering is an event related to puberty, marking the onset of sexual maturation, migration, and the reproductive stage. Silver eels are still sexually immature when they begin their spawning migration. Sexual maturation occurs during the period of migration to the spawning site in the ocean. On the contrary, anadromous fish species as salmons spend much of their life in salt water and go up rivers to spawn (Lee et al. 2020; Beechie et al. 2021) (Figure 3).

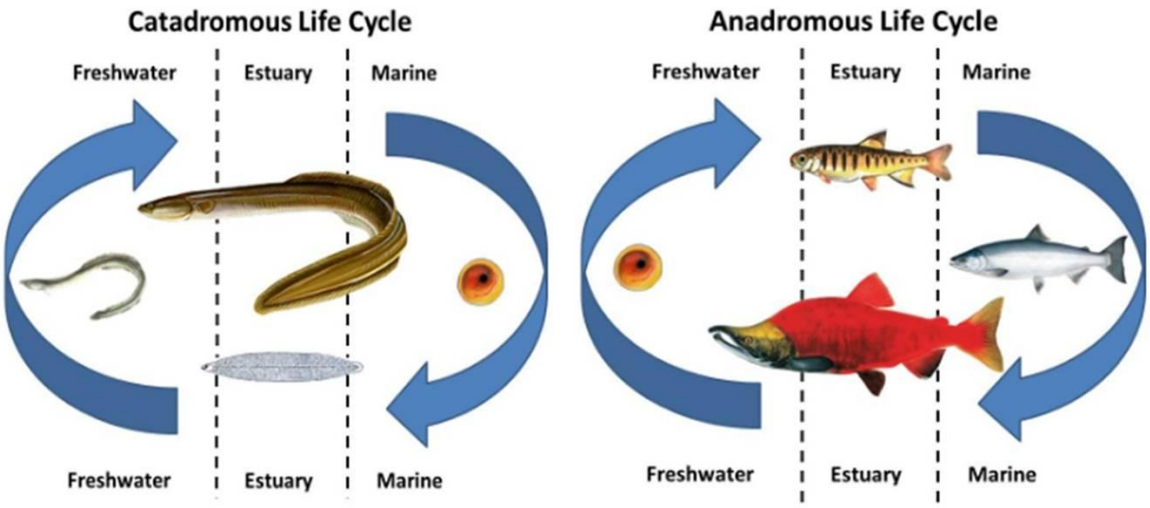

Figure 3. Catadromous and anadromous life cycles (https://thefisheriesblog.com/2013/05/20/anadromous-catadromous-amphidromous-oceanodromous-or-potamodromous/).

The migration is a common behaviour among salmonids, although the seawater acclimation period varies from species to species. For example, differently from other salmonids, the chum salmon fry shows a remarkable seawater adaptability prior to seawater entry (Lee et al. 2020). The analyses on 15 migrating species allowed to evidence a remarkable relationship between the content of SINE retroelements and the migratory behaviour of catadromous species. The results obtained of this topic are reported in the paper number 3.4 in the section Articles.

Moreover, diadromous species as eel and salmon migrate between SW and FW and thus they have to face changes in salinity in a defined stage of their life cycle. It is known that salinity is an abiotic factor that strongly influences the adaptation of marine and freshwater fish, and its changes can also represent a possible stress that threatens their survival. In line with this topic,

data obtained during my PhD evidenced a variation in TE transcriptional contribution in the juvenile eels, commonly adapted to salty water, when exposed to brackish and freshwater conditions. This increase was followed by a higher transcription of genes involved in TE silencing mechanisms in gill cells. The results obtained are reported in the paper number 3.5 in the section Articles.

Another important abiotic factor for fish species is the environmental temperature. This is a major stimulus affecting physiological and metabolic activities, especially for poikilothermic species (Dos Santos et al. 2013). Tolerance to this abiotic factor and the mechanisms activated by temperature stress have been extensively explored in eurythermal and extreme stenothermal fish (Logan and Buckley 2015; Gracey et al. 2004). To support this concept, interesting data have been reported for damselfish (*Pomacentrus moluccensis*) (Kassahn et al. 2007), barramundi (*Lates calcarifer*) (Newton et al. 2013), and tiger barb *Puntius tetrazona* (Liu et al. 2020; Carotti et al. 2023). This latter is a popular aquarium trade fish worldwide (Ng and Tan 1997; Kirankumar and Pandian 2003; Leknes 2014; Chapman 1997; Russo et al. 2007) belonging to the family Cyprinidae, living in freshwater fish at 21-28 °C. In my PhD project, I have considered this species to clarify the response of TEs to temperature changes in tissues of interest as gills, brain,

and liver. The data obtained evidenced a clear response of TEs and related silencing mechanisms in the hepatic tissue suggesting this organ as target for this king of stress. Indeed, TE activation might regulate the expression of stress-induced genes leading this organism to adapt to changes in temperature. Once again, these findings support the idea of a potential role of TEs in the rapid adaptation, thus representing a promising molecular tool for species resilience. The results obtained on this topic are reported in the paper number 3.6 in the section Articles.

3.Articles

3.1 Shedding light upon the complex net of genome size, genome composition and environment in chordates

The European Zoological Journal, 2020, 192–202
Vol. 87, No. 1, https://doi.org/10.1080/24750263.2020.1747558

Shedding light upon the complex net of genome size, genome composition and environment in chordates

A. CANAPA *, M. A. BISCOTTI, M. BARUCCA, F. CARDUCCI, E. CAROTTI, & E. OLMO

Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche, Ancona, Italy

(Received 20 February 2020; accepted 23 March 2020)

Abstract
The nucleotype theory has been advanced on the basis of studies regarding genome size and composition in various plant and animal species, i.e. the influence that genome can have on the phenotype independently of the informational content of DNA. It has also been noted that during evolution various interactions between different environmental factors and genome structural and functional parameters would have occurred. In this review, changes in genome size, transposon content, and base composition occurred during the evolution of chordates were examined. Many environmental stresses, such as temperature, can act on transposons and through these on genome size. Temperature is also one of the most important elements of natural selection able to interact both with base composition and genome size. It has been evidenced that temperature exerts a direct influence on base composition and its increase would have led to an higher content of genome GC-rich components during the evolution of chordates, in particular in endotherms. Temperature would have controlled the rate of biosynthesis in G1 phase and consequently the cell cycle duration which in turn would have interacted with genome size. The combined action of temperature, base composition, and genome size would also have been very important in controlling the metabolic rate. Finally, another important aspect of the nucleotypic effect is the influence that genome size and cell cycle duration, in correlation with environmental temperature, would have exert on embryo and larval development, very important for environmental adaptation. In conclusion, studies here reviewed to confirm the existence in chordates of a mutual influence between environment and genome non-coding components that would have played an important role in the evolution of these animals especially in environmental adaptation processes.

Keywords: *Genome evolution, transposons, GC composition, methylation*

Introduction

The idea that DNA plays a quantitative role, independently from its sequences, besides that one of protein coding, was first proposed by Commoner (1964). This was explicitly formulated by Bennett (1971, 1972) with the *nucleotype* theory according to which "that condition of the nucleus (most notably the DNA content) that affects the phenotype independently of the informational content of the DNA". This theory was based on the relationships evidenced initially for many plant species and subsequently observed also in various animal species, especially chordates (Gregory 2005). Indeed, in these latter, associations between genome size and several structural and functional cell parameters such as size and metabolism, and also parallelisms between quantitative genome variations and environmental conditions have been evidenced. Possible interactions with environment have been hypothesized also for other genome components such as transposons (McClintock 1984) and base composition (Bernardi 2004). Since genome size, transposon percentage and base composition are known for many species of chordates, this phylum is particularly interesting to understand the meaning of interactions between genome and environment. Indeed, they are one of the most common models to study the evolutionary processes and mechanisms. In this regard, it is important to examine the variations that the

*Correspondence: A. Canapa, Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche, Ancona, Italy. Email: a.canapa@univpm.it

aforementioned genome parameters have undergone during the evolution of chordates and potential interactions with the environment, assessing if they are random or an important constitutive phenomenon for some evolutionary processes such as, for example, speciation and adaptation. Other important aspects that have to be clarified are if variations in genome features are the consequence of environmental changes or if at the contrary, they are the drive for adaption of organisms to new habitats. Moreover, it is interesting to understand if genome–environment interactions are the same for variations in nuclear DNA content, in the transposon percentage and in base composition and if any reciprocal influences can be identified between these three parameters. In this review, we firstly examined changes that genome and environment have undergone in chordates evolution, since it is known that transposons and genome size are closely related.

Genome size and transposons

Data on genome sizes derive from Gregory (2019), while those related to TE percentage are referred to Canapa et al. (2015) (see Figure 1).

The trend of variations in genome size and in transposon percentage leads to both phylogenetic considerations and evaluations of the meaning of correlations between the aforementioned genome characters and various structural and functional parameters of cell and organism in adaptive processes.

It is assumed that the ancestral values of genome size in chordates were low and later they increased during evolution; however, the situation is more complex because during evolution of this phylum, increases and decreases in genome size occurred and a similar trend is also found within each class. Furthermore, even if it is ascertained that the main cause of genome size increase is due to TE (transposons) amplification (Canapa et al. 2015), other mechanisms such as whole genome duplication and polyploidy have been suggested (Gregory 2005).

In the primitive chordates, genome size is very low and ranges from 0.07 pg in *Oikopleura* to 0.74 pg in *Botryllus*. The transposons range from 16.7% in *Ciona* to 22% in the amphioxus and in the three species studied are DNA, LTR, and nonLTR.

A first substantial increase would have occurred at the origin of vertebrates and it is assumed that one

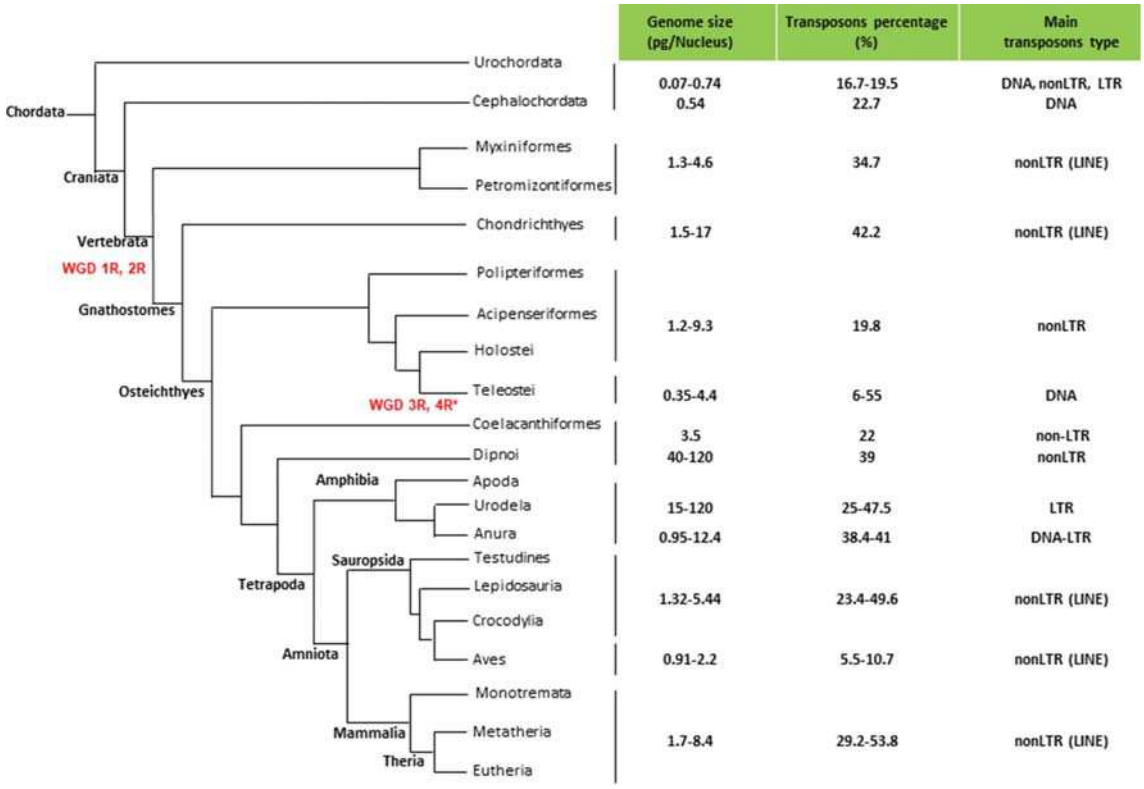

Figure 1. Cladogram showing evolutionary relationships between the main lineages of chordates. Whole Genome Duplication (WGD) events in vertebrate evolution are indicated in red: 1R and 2R occurred before the divergence of Vertebrata, 3R in Teleost and 4R* in salmonids.

of the causes was the first whole genome duplication (Meyer & Schartl 1999).

In cyclostomes, the genome size ranges from 1.3 pg to 2.5 pg in Petromyzontiformes and from 2.3 pg to 4.6 pg in Mixyniformes and it has been hypothesized that this difference is linked to the complexity of the development since Mixyniformes have a direct development while Petromyzontiformes present a metamorphosing stage (Hardie & Hebert 2003). The transposons have been studied only in *Petromyzon* where they represent 34.7% and are mainly nonLTR retrotransposons.

The genome size of Chondrichthyes ranges from 1.51 pg in the holocephalus *Hydrolagus* to over 17 pg in the elasmobranch *Rhinobatus*. In this subclass, the average genome size is significantly different in rays (6.7 pg) and sharks (5 pg). Furthermore, studies on the reassociation kinetics have highlighted cases of cryptic polyploidy in selachians (Olmo et al. 1982).

In Chondrichthyes, there is a positive correlation between genome size and nuclear and cell volume. The highest values were found in species that live in extreme conditions of temperature, light, pressure, and scarce food availability, such as deep and/or cold sea (Stingo et al. 1980; Hardie & Hebert 2003). The differences in the genome size can also be correlated with development complexity since clear differences between viviparous species and those ovoviviparous and oviparous have been evidenced (Hardie & Hebert 2004). Transposons have been studied only in the elepha0nt shark *Callorhinchus milii* where they represent 42.2% and are nonLTR retrotransposons.

The genome size of the primitive Osteichthyes ranges from 1.15 pg in *Amia* to 7.25 pg in *Polypterus*. Also in this group, there is a positive correlation between genome size and nuclear and cell volume. Transposons have been studied in *Lepisosteus* in which are mainly of the nonLTR type and represent 19.77% of its genome.

In vertebrates, the third whole genome duplication occurred at the origin of Teleostei, but this did not lead to genome size values higher than the average values of most living vertebrates (Meyer & Schartl 1999). The genome size varies from 0.35 pg in tetraodontids that have the smallest and most compact genome of all chordates to a maximum of 4.4 pg. However, if polyploidy cases are excluded, the maximum values rarely exceed 2.5 pg. The genome size is positively correlated with nuclear and cell volume (Hardie & Hebert 2003).

There are several parallelisms between genome size and environment: freshwater and anadromous species have larger genomes, while marine and catadromous species have smaller genomes (Hardie & Hebert 2004). This is in agreement with the observations made by Ebeling et al. (1971) according to which the genomes of eurobiotic (eurihalyne) species are larger than those of stenobiotic (stenohalyne) species. Moreover, in the primitive Protoacanthopterygii and Paraacanthopterygii, species that live in the ocean depths and in cold environments, poorly lighted and with scarce nutrients, have larger genomes than those of the related species living in coastal waters (Ebeling et al. 1971). In addition to climatic characteristics, the stability of the environment, in which certain teleost species live, would be also very important. Indeed, it has been observed that polar teleosts and those of poor variable tropical environments, have small genomes and this could be related to a specific reproductive strategy (Hardie & Hebert 2003, 2004).

The percentage of transposons varies from only 6% in tetraodontids to 55% in zebrafish and, unlike most vertebrates, are mainly DNA transposons (Chalopin et al. 2015). On average, teleosts exhibit the widest diversity in transposons reaching 27 superfamilies in zebrafish (Sotero-Caio et al. 2017).

A case of correlation between transposons and environment was observed in a study on the identification and characterization of the transposable elements *Rex3* in teleosts. In this paper, Carducci et al. (2019a) have demonstrated for the first time, through phylogenetic analyses, the correlation between these transposable elements and environmental temperature. In particular, the results highlighted a clear sequence distinction of *Rex3* elements belonging to fish living in cold waters compared to those of fish living in temperate waters, regardless of the evolutionary and taxonomic relationships of the analysed species.

The most significant increase in genome size and in the transposon percentage would have occurred during the transition from aquatic to terrestrial environment, that is, in a crucial step of evolution, and this led to the large genomes that are observed in lungfish and amphibians (Organ et al. 2015). It is difficult to establish whether this increase is directly related to the conquest of the land, but it is interesting that a similar situation has also been noted in the transition from aquatic to terrestrial gastropods (Vinogradov 2000).

Primitive sarcopterygians occupy a very important position in the evolution of vertebrates because tetrapods originated from them. The genome size is very different in crossopterygians and lungfish.

The only two living species of *Latimeria* have a genome of 3.5–3.6 pg and the TE percentage (mainly SINE and LINE Chalopin et al. 2015) is 22% (Amemiya et al. 2013).

In lungfish, the genome size is the highest of all vertebrates and ranges from 50 pg to over 120 pg. A polyploid karyotype has been noted in the species *Protopterus dolloi*. The living lungfish have several characteristics that make them paedomorphic organisms and it has been speculated that this would be related to the large genome size (Joss 2006; Joss & Johanson 2007). In primitive lungfish, genome size would have been similar to that of *Latimeria* (Organ et al. 2015) but from the Carboniferous a great expansion would have started which would have led to the current huge genomes and it would have been accompanied by an evolutionary decline with great reduction in the number of genera and species (Thomson 1972; Thomson & Muraszko 1978). This decline contrasts with the hypothesis that genome size increase may promote evolutionary diversification (Kraaijeveld 2010). Transposons have only been studied in *Neoceratodus* in which they are mainly nonLTR retrotransposons and make up 39% of its genome (Metcalfe et al. 2012).

In amphibians, genome size shows very different values in the three orders. In anurans it ranges from 0.95 pg to 12.4 pg; in gymnophions, in which only three species have been studied, it ranges from 3.7 pg to 13.95 pg; in urodeles it ranges from a minimum of 15 pg, higher than the maximum value observed in the other two orders, to over 120 pg in *Necturus*, a value similar to that recorded in lungfish. Cases of polyploidy are known in both anurans and urodeles (Schmid et al. 2015).

The transposon percentage in the only two anuran species analysed is around 40% and are both DNA and LTR elements; in urodeles it ranges from 25% to 47.5%, and, unlike the majority of vertebrates, they are almost exclusively LTR retroelements (Canapa et al. 2015; Chalopin et al. 2015). In both the most primitive and most recent species, the majority of transposons belongs to a few families and is mainly represented by the LTR Gypsy elements constituting more than 25% in *Cryptobranchus alleganiensis* (Sun & Mueller 2014). Several works have shown that transposons have been the main cause of genome expansion in these amphibians.

Fossil amphibians (rhipidistians, labyrinthodonts and anthracosaurs) had low values similar to the current ones of bufonids and ranids (5–10 pg) (Thomson & Muraszko 1978; Organ et al. 2011).

The expansion of the lissamphibian genome would have started in the Permian before the separation of urodeles from anurans and the expansion of the urodeles genome, due to a few LTR families, would have continued in the early and Middle Triassic (Laurin et al. 2015; Christoph-Liedtke 2016). The first phase of the expansion took place gradually and led to larger genomes in anuran; the next phase would have occurred because of the LTR retrotransposon saltatory proliferation (Sun & Mueller 2014) and would have led to very large genomes typical of current urodeles already in the Middle Jurassic (Laurin et al. 2015; Organ et al. 2015; Christoph-Liedtke et al. 2018).

An interesting aspect, common to the great expansion of genome in lungfish and amphibians, especially urodeles, is that it would have occurred in periods of drastic environmental changes such as the beginning of Carboniferous and the transition from Permian to Triassic and this is in agreement with the observation that transposon activity is often the consequence of environmental stresses (McClintock 1984; Capy et al. 2000).

A positive relationship between genome size and nuclear and cell volume has been recorded in anurans and urodeles.

Even in amphibians, there are correlations between the genome size and the environment. In general, species that live in cold environments have more DNA and a longer development; in particular, in urodeles there is a direct correlation between genome size and latitude (Litvinchuk et al. 2007).

In amphibians, variations in genome size show interesting relationships with reproductive strategies and embryonic development. A positive correlation is observed between genome size and cell cycle duration during the synchronous divisions of segmentation (Vinogradov 1999). In anurans, genome size is correlated with cell proliferation rate and with developmental rates in different embryonic stages (Chipman et al. 2001); in urodeles, on the other hand, there is a negative relationship between genome size and regeneration and growth rates during embryogenesis and differentiation (Sessions & Larson 1987). In Plethodontid urodeles genome size is positively correlated to development duration (Jockusch 1997).

A peculiar feature of urodeles is the relationship between genome size, paedomorphosis and neoteny: species that have a biphasic life cycle have the smallest genomes; those with large genomes are often optional neotenic and those with giant genomes are obliged neotenic (Morescalchi 1991; Gregory 2005).

Also the direct correlation between genome size and cell cycle duration (Vinogradov 1999) and the inverse correlation between genome size and metabolic rate observed in amphibians (Monnickendam & Balls 1973; Licht & Lowcock 1991) are interesting from an evolutionary point of view.

The genome size in reptiles ranges from 1.32 pg to 5.44 pg. The average higher values are found in cheloniids and in Tuatara. There is an inverse relationship between genome size and chromosome changing rate

that is higher in Squamata and is directly related to the number of living species (Olmo 2005). Furthermore, genome size is also directly correlated with nucleus and cell size and is inversely correlated to the metabolic rate (Olmo 2003).

A difference between cheloniids/crocodiles and squamates is the presence in the latter of two types of chromosomes: microchromosomes, richer in genes, and macrochromosomes (Kasai et al. 2019).

The percentage of transposons ranges from 23.4% to 49.6% and are mainly LINE. Unlike what is known in other amniotes, in the Squamata, despite a little genome size variations, there is a surprisingly large variability in the amount of repetitive sequences especially transposons and in snakes there is the highest content of microsatellites originating from transposons (Pasquesi et al. 2018).

Birds have very low genome size ranging from 0.91 pg to 2.16 pg, with the highest value in ostrich. The percentage of transposons is the lowest among vertebrates and ranges from 5.5% to 10.7% and are mainly LINE.

Even in birds, there are microchromosomes richer in genes and macrochromosomes. Originally the genome of birds would have had a similar size to that of current mammals (3–3.7 pg) and would have undergone a reduction in Mesozoic and Caenozoic eras that would have led to current values (Organ et al. 2007). A similar situation is assumed to have previously characterized the pterosaurs (Organ & Shedlock 2009). There is an inverse relationship between genome size and metabolic rate (Vinogradov 1997; Gregory 2001) leading to hypothesize that small genomes and corresponding high metabolism would have been the premise for the development of flight (Hughes & Hughes 1995).

In mammals, the nuclear DNA content ranges from 2.9 pg to 3.6 pg in monotremes, from 2.99 pg to 5.02 pg in marsupials, and from 1.7 pg to 8.4 pg in placentals.

The percentage of transposons is 44.6% in platypus, from 52% to 53.8% in marsupials and from 29.2% to 46% in placental and are nonLTR retrotransposons.

An inverse relationship between genome size and metabolic rate has also been noted in mammals (Vinogradov 1995) and it has been hypothesized that there is also a correlation between genome size, metabolism and flight adaptation in Chiroptera, even if data are contradictory (Smith et al. 2013).

Base composition and methylation

The research groups of Bernardi and Vinogradov have been the main conductors of studies concerning the base composition. They have highlighted variations of this parameter during the evolution of the chordates.

Vinogradov (1998a) evidenced a general positive but not linear relationship between GC content and genome size; the GC percentage shows great variability at lower values while it tends to stabilize at the level of about 46% with the increase of genome size. The regression line of reptiles and birds follows the general trend but has a steeper slope due to a greater increase in GC composition compared to the increase in nuclear DNA content, while in teleosts an inverse relationship has been noted. The higher percentage of GC is positively correlated with bendability, thermostability and the ability to change the DNA from B to Z form, property linked to chromatin opening and transcription activation, and inversely related to curvature that favours chromatin condensation. Bernardi and colleagues' studies have shown GC% values in line with those reported by Vinogradov, although they have noted a negative correlation with genome size in most vertebrate classes (Bernardi & Bernardi 1990; Jabbari et al. 1997; Bernardi 2004). These researchers have shown a compartimentalization in eukaryotic genome, i.e. made up of a series of distinct fractions, each characterized by a specific and different GC average composition, the so-called isochores. Changes in base composition would have mainly depended on body temperature. In fact, during vertebrate evolution, a clear variation in GC percentage has been observed in isochore distribution, only in the transition from cold-blooded to warm-blooded. In the former, similarly to invertebrates, genome is constituted almost exclusively of AT-rich isochores, while in birds and mammals there are GC-rich isochores. An intermediate situation has been reported for reptiles, where in Squamata the isochore pattern is similar to that one observed in fish and amphibians, while in turtles and crocodiles it is more similar to that of birds and mammals (Hughes et al. 2002). The transition in base composition from ectothermic to endothermic vertebrates has been explained with the "thermodynamic stability hypothesis" according to which the increase of GC-rich components in endotherms would have led to a greater thermodynamic stability of DNA, RNA, and proteins. This is of great importance to cope with the impact that the increase in body temperature has on various structural and functional aspects of cell and genome (Bernardi 2004). The percentage in base composition has shown a positive correlation with methylation level which, however, is different between fishes and amphibians on the one hand,

and birds and mammals on the other, where two-fold values have been evidenced in the former group (Bernardi 2004). Moreover, for reptiles, an intermediate situation has been described (Varriale & Bernardi 2006). In all vertebrates, methylation is inversely related to the genome size and methylation level is inversely correlated with body temperature in fish (Bernardi 2004; Varriale & Bernardi 2006). A positive correlation has been noted between the GC content and the metabolic level (Uliano et al. 2010; Bernà et al. 2012; Varriale 2014; Tarallo et al. 2016). In teleosts, it has been observed that the GC percentage increases from freshwater to marine species and within these from non-migratory to migratory species. This trend would be related to both salinity changes and higher costs in terms of energy for migratory species (Tarallo et al. 2016).

Relationship of genome characters with the environment

As described in the previous sections, various interactions exist between genome size, transposons and base composition, and between these and the environment. Some of these interactions concern all the aforementioned genome characters, while others are referred to some of these features. However, these relationships may not be always simple and linear, since in some cases the environmental factors would seem to be the direct cause of the changes in the size and/or genome composition, while in other cases genome characters influencing some cell structures and functions, such as cell size (Olmo 1983) and cell cycle (Vinogradov 1999) would have favored adaptation to new environmental conditions. However, not all these interactions seem to be supported by experimental data.

Recently, many researches on transposons have shown that they are among the main causes of genome expansion and that their activity is stimulated by various environmental stressors (Canapa et al. 2015; Carducci et al. 2019b), such as temperature, thermal shocks, radiations, chemicals, and viral infections (Capy et al. 2000; Fujino et al. 2011; Garcia Guerreiro 2012; Carducci et al. 2019a, b). In some cases, these stressors would act directly on transposons with a mechanism typical of defensive gene activation, since some transposons have sequences similar to the promoters of these genes (Takeda et al. 1999; Capy et al. 2000); however, in many other cases, the action would mainly depend on a temperature-dependent activation of transposase (Capy et al. 2000; Fujino et al. 2011; Carducci et al. 2019a).

Besides to be involved in transposon activation, temperature is one of the most important elements of natural selection able to interact both with base composition and genome size (see Figure 2). Studies conducted on base composition have clearly shown that body temperature affects GC percentage. This effect is evident in structural changes of macromolecules. In accordance with the "thermodynamic stability hypothesis", GC higher values are directly related to an increased DNA thermostability (Bernardi 2004) which in turn influences RNA composition and chemical and physical properties of proteins, which in species showing a higher GC percentage are richer in hydrophobic amino acids. Thus, macromolecules of endothermic organisms

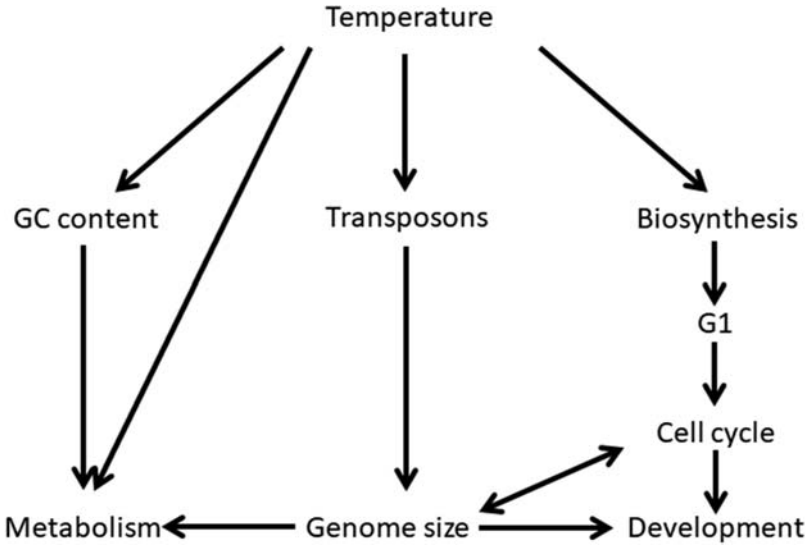

Figure 2. Schematic representation of interactions between genome features and parameters.

are thermodynamically more stable and can function optimally even in the case of an increase in body temperature (Bernardi 2004).

As previously described, an increase in GC content confers to DNA not only a greater thermostability, but also facilitates chromatin opening and the consequent transcriptional activation; on the contrary, a higher GC richness reflects a lower DNA curvature which favors chromatin condensation (Vinogradov 1998a, 2003, 2005). The GC-rich DNA, with its higher physical stability would compensate for higher gene mutation rate observed in larger genomes and would also protect against damages due to chemical mutagens, many of which have a higher affinity for GC-rich DNA and then would be more easily absorbed and neutralized by the higher GC-rich DNA regions (Vinogradov 1994; Bernardi 2004).

However, the influence of temperature would not be the direct cause of the GC content increase but it would be one of the main selective factors responsible for structural changes in macromolecules, allowing birds and mammals to cope with the increase of body temperature (Bernardi 2004).

Moreover, the direct relationship between GC percentage and methylation increase demonstrates that temperature, *via* base composition, has been able to influence an important epigenetic mechanism involved in transposon inactivation (Canapa et al. 2015) and in the control of different cellular functions, such as tissue-specific gene expression, cell differentiation, and development (Varriale 2014).

In addition, the influence of temperature on genome size could act indirectly through the control of the cell cycle duration.

This cycle mainly depends on the DNA amount that has to be duplicated. However, differences have been observed in two phases of the cell cycle: S phase is strictly related to the DNA amount, while the G1 phase duration is influenced both by DNA content and biosynthesis rate. Therefore, a higher biosynthesis rate corresponds to a shorter G1 phase and thus to a shorter cell cycle.

Since the biosynthesis rate is influenced by temperature, the G1 phase and the cell cycle can be controlled by temperature variations. Xia (1995) has reported an inverse relationship between biosynthesis rate and genome size. This would explain why species, living in warm environments (especially poikilothermic species), have smaller genomes, a higher biosynthesis rate, and a shorter cell cycle compared to related species living in colder environments.

According to this author, smaller changes in G1 phase duration (and in biosynthesis rate) at specific changes in body temperature occur in species with large genomes compared to those with smaller genomes. Therefore, a large genome could be used by poikilothermic species to buffer potential damages caused by climate changes (Xia 1995).

The relationship between genome parameters and temperature-related metabolism (Kleiber 1932) is more complex.

An inverse relationship between genome size and metabolic rate has been observed in all vertebrates (Licht & Lowcock 1991; Gregory 2005; Vinogradov & Anatskaya 2006) in which a different trend has been evidenced in reptiles-birds compared to amphibians-mammals: indeed, an increase in genome size corresponds to a greater reduction in the metabolic rate in the former than the latter (Vinogradov & Anatskaya 2006); these two trends are similar to those concerning the ratio between genome size and GC content, i.e. in reptiles-birds at the same genome size the GC percentage is higher than that of amphibians-mammals (Vinogradov 1998a).

The GC content is proportional to chromatin condensation, thus nuclei with a greater GC quantity are larger for the same DNA amount. Therefore, it has been hypothesized that metabolic rate is influenced not by the genome size but mainly by the nuclear volume as determined by the interaction between genome size and GC content (Vinogradov & Anatskaya 2006).

It is assumed that metabolism is conditioned by both respiratory and nutritive exchanges which depend on surface–volume ratio of nucleus and cell (Olmo 1983; Vinogradov 1995). Consequently, genome size determining these two-dimensional cell parameters is probably one of the most important factors regulating cell metabolism (Olmo 1983; Hardie & Hebert 2003).

Relationships observed between metabolic rate and base composition are less clear. In teleosts, contradictory aspects have been evidenced in the relationships between these two parameters and the temperature (Uliano et al. 2010). Indeed, the GC composition has a positive correlation with metabolic rate, but both these parameters have a negative correlation with temperature. This trend cannot be generalized since a similar low metabolic rate has been found in tropical species and in organisms living in deep sea, characterized by lower temperature.

In this sub-class, genome parameters would also be influenced by salinity, especially by the ability to cope variations in saline concentration: it has been observed

that GC content increases from non-migratory freshwater species to migratory marine species, similarly the genome size is higher in eurobiotic species than in stenobiotic ones (Ebeling et al. 1971; Hardie & Hebert 2003, 2004; Tarallo et al. 2016).

Tarallo et al. (2016) have hypothesized that this situation is linked both to metabolic rate and to a higher necessity of oxygen for migratory species. Another possible explanation has been proposed by Vinogradov (1998b, 1998c) which hypothesizes that non-coding DNA performs a buffering function on the intracellular concentration of solutes through non-specific bonds with proteins. Cells with a major amount of repetitive DNA, and therefore constituted by larger genomes, have a higher ability to regulate the intracellular composition of solutes balancing the variations in the composition of the extracellular solutes.

Finally, another interesting aspect of the relationship between genome size and processes involved in speciation and environmental adaptation concerns both time and mode of embryo and larval development.

It is known that the cell cycle length influences the duration of embryo and larval development (Vinogradov 1999; Gregory 2005). A positive relationship between genome size, cell cycle, and developmental duration has been detected in many classes of vertebrates, especially anamniotes, while no relationship has been reported in amniotes (Gregory 2001; Olmo 1983) concerning these aspects. These correlations have been studied particularly in amphibians. In anurans genome size is directly related to cell proliferation rate in segmentation and developmental rates in different stages (Chipman et al. 2001). In urodeles, besides a positive relationship between nuclear DNA content and developmental duration (Jockusch 1997), there is a negative relationship with the regeneration, the differentiation, and the growth rate during embryogenesis (Sessions & Larson 1987).

The relationship between genome size and the related developmental duration can represent a limiting factor in environment adaptation, especially concerning temperature and water availability. Indeed it has been observed that in amphibians, especially in anurans, species that live in environments characterized by scarcity of water have a small genome and a rapid development, while species that live in cold environments with water availability have larger genomes and a slow development (Goin et al. 1968; McMenamin & Hadly 2010); similarly in urodeles species living in cold environments have a larger genome and a longer duration of development compared to species living in warmer environments (Litvinchuk et al. 2006; Lertzman-Leofsky et al. 2019). Similar correlations have also been reported for some fishes (Hardie & Hebert 2003, 2004). In addition to the developmental duration, variations in the DNA content also seem to affect the developmental complexity (Gregory 2001, Gregory 2005). Genome size shows relationships with different aspects of development such as viviparity, direct or indirect development, and neoteny. In Chondrichthyes viviparous species have larger genomes (Hardie & Hebert 2004). A trend reported in several animals concerns the increase in genome size in relation to the transition from metamorphosis to the direct development, and facultative or obligatory neoteny. In cyclostomes, anurans, and urodeles, species having a complete metamorphosis present smaller genomes than those with direct development. Moreover, urodeles that have large genomes are facultative or obligatory neotenic species.

This trend could be common to all metazoans since also in insects it has been observed an increase in DNA content from holometabolous, to incomplete metamorphosing species (hemimetabolous) to those with direct development (ametabolous) (Gregory 2005). This aspect would be linked to heterochrony, a mechanism that involves changes in developmental rate and duration, and from which paedomorphosis can result, i.e. the mantainment of immature characters in adults, when development slows down or interrupts prematurely (Morescalchi 1991). In amphibians, one of the main causes of paedomorphosis would be an increase in genome size since this increase would reduce the growth rate, differentiation, and cell migration during embryogenesis by preventing the formation of late-developing traits (Roth et al. 1997; Womack et al. 2019).

Paedomorphosis played a very important role in the evolution of amphibians, since paedomorphic forms were already present in the labyrinthodonts and in the dyssorophoids which are the ancestors of the current lissamphibians (Morescalchi 1991). Paedomorphosis would have played an important role also in the Dipnoi, which have large genomes (similar to those of the facultative or obligatory neotenic urodeles) and show larval characters (Bemis 1984; Joss 2006; Joss & Johanson 2007).

Conclusions

This analysis shows the existence in chordates of a mutual influence between environment and the

non-coding components of the genome and this supports the hypothesis that the nucleotype would have played an important role in the evolution of this phylum, mainly in environmental adaptation processes.

The parameter that seems to be most relevant in genome–environment relationships is certainly the temperature variation (Figure 2) that directly or indirectly, e.g. through metabolism, has an effect on transposons, genome size, and base composition. The action of genome–temperature interactions was one of the most relevant phenomena, mainly for the evolution of the poikilothermic species. However, temperature-dependent enrichment of the GC-rich isochores that has occurred in the transition from ectothermic to endothermic vertebrates was very important for the latter since it has contributed to making these vertebrates less environment-dependent.

Overall, it is the environment that exerts its effect on genome, but in certain cases the variations in the genome components, which influence the structural and functional characteristics of cell, regulate evolutionary processes, such as the environmental adaptation.

In all cases, the interactions between environmental factors and genome components are part of the typical processes of natural selection and the variations affecting genome size, the transposon percentage, and base composition represent preadaptations on which environmental variations operate.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

A. Canapa http://orcid.org/0000-0001-8087-9331
M. A. Biscotti http://orcid.org/0000-0003-3114-240X
M. Barucca http://orcid.org/0000-0002-0869-3214
F. Carducci http://orcid.org/0000-0002-3985-7998
E. Carotti http://orcid.org/0000-0003-0592-7457

References

Amemiya CT, Alfoldi J, Lee AP, Fan S, Philippe H, MacCallum I, Braasch I, Manousaki T, Schneider I, Rohner N, Organ C, Chalopin D, Smith JJ, Robinson M, Dorrington RA, Gerdol M, Aken B, Biscotti MA, Barucca M, Baurain D, Berlin AM, Blatch GL, Buonocore F, Burmester T, Campbell MS, Canapa A, Cannon JP, Christoffels, De Moro G, Edkins AL, Fan L, Fausto AM, Feiner N, Forconi M, Gamieldien J, Gnerre S, Gnirke A, Goldstone JV, Haerty W, Hahn ME, Hesse U, Hoffmann S, Johnson J, Karchner SI, Kuraku S, Lara M, Levin JZ, Litman GW, Mauceli E, Miyake T, Mueller MG, Nelson DR, Nitsche A, Olmo E, Ota T, Pallavicini A, Sumir P, Picone B, Ponting CP, Prohaska SJ, Przybylski D, Saha NR, Ravi V, Ribeiro FJ, Sauka-Spengler T, Scapigliati G, Searke SMJ, Sharpe T, Simakov O, Stadler PF, Stegeman JJ, Sumiyama K, Tabbaa D, Tafer H, Turner-Maier J, van Heusden P, White S, Williams L, Yandell M, Brinkmann H, Volff JN, Tabin CJ, Shubin N, Schartl M, Jaffe DB, Postlethwait JH, Venkatesh B, Di Palma F, Lander ES, Meyer A, Lindblad-Toh K. 2013. The african coelacanth genome provides insight into tetrapod evolution. Nature 496:311–316. DOI: 10.1038/nature12027.

Bemis WE. 1984. Paedomorphosis and the evolution of the Dipnoi. Paleobiology 10:293–307. DOI: 10.1017/S0094837300008277.

Bennett MD. 1971. The duration of meiosis. Proceedings of the Royal Society of London B 178:277–299.

Bennett MD. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. Proceedings of the Royal Society of London B 181:109–135.

Bernà L, Chaurasia A, Angelini C, Federico C, Saccone S, D'Onofrio G. 2012. The footprint of metabolism in the organization of mammalian genomes. BMC Genomics 13:174. DOI: 10.1186/1471-2164-13-174.

Bernardi G. 2004. Structural and evolutionary genomics. Lausanne, New York, Oxford, Shannon, Singapore, Tokyo: Elsevier Amsterdam. pp. 434+XV.

Bernardi G, Bernardi G. 1990. Compositional patterns in the nuclear genome of cold-blooded verterbrates. The Journal of Molecular Evolution 31:265–281. DOI: 10.1007/BF02101122.

Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E. 2015. Transposons, Genome size, and evolutionary insights in animals. Cytogenetic and Genome Research 147:217–239. DOI: 10.1159/000444429.

Capy P, Gasperi G, Biémont C, Bazin C. 2000. Stress and transposable elements: Co-evolution or useful parasites? Heredity 85:101–106. DOI: 10.1046/j.1365-2540.2000.00751.x.

Carducci F, Biscotti MA, Barucca M, Canapa A. 2019b. Transposable elements in vertebrates: Species evolution and environmental adaptation. The European Zoological Journal 86:497–503. DOI: 10.1080/24750263.2019.1695967.

Carducci F, Biscotti MA, Forconi M, Barucca M, Forconi M. 2019a. An intriguing relationship between teleost *Rex3* retroelement and environmental temperature. Biology Letters 15:20190279. DOI: 10.1098/rsbl.2019.0279.

Chalopin D, Naville M, Plard F, Galiana D, Volff JN. 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution invertebrates. Genome Biology and Evolution 7:567–580. DOI: 10.1093/gbe/evv005.

Chipman AD, Khaner O, Haas A, Tchernov E. 2001. The evolution of genome size: What can be learned from anuran development? Journal of Experimental Zoology Part B (Molecular and Developmental Evolution) 291:365–374. DOI: 10.1002/jez.1135.

Commoner B. 1964. The structure and function of chromatin. Nature 202:960–968. DOI: 10.1038/202960a0.

Ebeling AW, Atkin NB, Setzer PY. 1971. Genome sizes of teleostean fishes: Increase in some deep-sea species. The American Naturalist 105:549–562. DOI: 10.1086/282744.

Fujino K, Hashida SN, Ogawa T, Natsume T, Uchiyama T, Mikami T, Kishima Y. 2011. Temperature controls nuclear import of *Tam3* transposase in *Antirrhinum*. Plant Journal 65:146–155. DOI: 10.1111/j.1365-313X.2010.04405.x.

Garcia Guerreiro MP. 2012. What makes transposable elements move in the *Drosophila* genome? Heredity 108:461–468. DOI: 10.1038/hdy.2011.89.

Goin OB, Goin CJ, Bachmann K. 1968. DNA and Amphibian life history. Copeia 1968:532–540. DOI: 10.2307/1442021.

Gregory TR. 2001. Genome size and developmental parameters in the homeothermic vertebrates. Genome 45:833–838. DOI: 10.1139/g02-050.

Gregory TR. 2005. Genome size evolution in animals. In: Gregory TR, editor. The evolution of the genome. Boston, Heidelberg, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo: Elsevier Amsterdam. pp. 3–87.

Gregory TR. 2019. Animal genome size database. Available: http:www.genomesize.com.

Hardie DC, Hebert PDN. 2003. The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. Genome 46:683–706. DOI: 10.1139/g03-040.

Hardie DC, Hebert PDN. 2004. Genome size evolution in fishes. Canadian Journal of Fisheries and Aquatic Science 61:1636–1646. DOI: 10.1139/f04-106.

Hughes AL, Hughes MK. 1995. Small genomes for better flyers. Nature 377:391. DOI: 10.1038/377391a0.

Hughes S, Clay O, Bernardi G. 2002. Compositional patterns in reptilian genomes. Gene 295:323–329. DOI: 10.1016/S0378-1119(02)00732-1.

Jabbari K, Cacciò S, Païs de Barros JP, Desgrès J, Bernardi G. 1997. Evolutionary changes in CpG and methylation levels in the genome of vertebrates. Gene 205:109–118. DOI: 10.1016/S0378-1119(97)00475-7.

Jockusch EL. 1997. An evolutionary correlate of genome size change in plethodontid salamanders. Proceedings of the Royal Society of London B 264:597–604. DOI: 10.1098/rspb.1997.0085.

Joss J, Johanson Z. 2007. Is *Palaeospondylus gunni* a fossil larval lungfish? Insights from *Neoceratodus forsteri* development. Journal of the Experimental Zoology Part B Molecular and Developmental Evolution 308:163–171. DOI: 10.1002/jez.b.21125.

Joss JM. 2006. Lungfish evolution and development. General and Comparative Endocrinology 148:285–289. DOI: 10.1016/j.ygcen.2005.10.010.

Kasai F, O'Brien PCM, Ferguson-Smith MA. 2019. Squamate chromosome size and GC content assessed by flow karyotyping. Cytogenetic and Genome Research 157:46–52. DOI: 10.1159/000497265.

Kleiber M. 1932. Body size and metabolism. Hilgardia 6:315–353. DOI: 10.3733/hilg.v06n11p315.

Kraaijeveld K. 2010. Genome size and species diversification. Evolutionary Biology 37:227–233. DOI: 10.1007/s11692-010-9093-4.

Laurin M, Canoville A, Struble M, Organ C, de Buffrénil V. 2015. Early genome size increase in urodeles. Comptes Rendus Palevol 15:74–82. DOI: 10.1016/j.crpv.2014.12.006.

Lertzman-Leofsky G, Mooers AO, Greenberg DA. 2019. Ecological constraints associated with genome size across salamander lineages. Proceedings of the Royal Society of London B 286(1911):20191780. DOI: 10.1098/rspb.2019.1780.

Licht LE, Lowcock LA. 1991. Genome size and metabolic rate in salamanders. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 100:83–92. DOI: 10.1016/0305-0491(91)90089-V.

Liedtke CH, Gower DJ, Wilkinson M, Gomez-Mestre I. 2018. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nature Ecology and Evolution 2:1792–1799. DOI: 10.1038/s41559-018-0674-4.

Liedtke HC, Müller H, Rödel MO, Menegon M, Gonwouo LN, Barej MF, Gvoždík V, Schmitz A, Channing A, Nagel P, Loader SP. 2016. No ecological opportunity signal on a continental scale? diversification and life-history evolution of african true toads (anura: bufonidae). Evolution 70:1717–1733.

Litvinchuk SN, Rosanov JM, Borkin L. 2006. Correlations of geographic distribution and temperature of embryonic development with the nuclear DNA content in the salamandridae (Urodela, Amphibia). Genome 50:333–342.

Litvinchuk SN, Rosanov JM, Borkin L. 2007. Correlations of geographic distribution and temperature of embryonic development with the nuclear DNA content in the Salamandridae (Urodela, Amphibia). Genome 50:333–342. DOI: 10.1139/G07-010.

McClintock B. 1984. The significance of responses of the genome to challenge. Science 226:792–801. DOI: 10.1126/science.15739260.

McMenamin SK, Hadly EA. 2010. Developmental dynamics of *Ambystoma tigrinum* in a changing landscape. BMC Ecology 10(1):10. DOI: 10.1186/1472-6785-10-10.

Metcalfe CJ, Filée J, Germon J, Joss J, Casane D. 2012. Evolution of the Australian lungfish (*Neoceratodus forsteri*) genome: A major role for CR1 e L2 line elements. Molecular Biology and Evolution 29:3529–3539. DOI: 10.1093/molbev/mss159.

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. Current Opinion in Cell Biology 11:699–704. DOI: 10.1016/S0955-0674(99)00039-3.

Monnickendam MA, Balls M. 1973. The relationship between cell sizes respiration rates and survival od amphibian tissues in long term organ culture. Comparative Biochemistry and Physiology Part A: Physiology 44:871–880. DOI: 10.1016/0300-9629(73)90150-3.

Morescalchi A. 1991. Genome and the rise of terrestrial vertebrates. In: Ghiara G, editor. Symposium on the evolution of terrestrial vertebrates. Modena: Mucchi. pp. 31–50.

Olmo E. 1983. Nucleotype and cell size in vertebrates: A review. Basic and Applied Histochemistry 27:227–256.

Olmo E. 2003. Reptiles: A group of transition in the evolution of genome size and of the nucelotypic effect. Cytogenetic and Genome Research 101:166–171. DOI: 10.1159/000074174.

Olmo E. 2005. Rate and chromosome changes and speciation in reptiles. Genetica 125:185–203. DOI: 10.1007/s10709-005-8008-2.

Olmo E, Stingo V, Cobror O, Capriglione T, Odierna G. 1982. Repetitive DNA and polyploidy in selachians. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 73:739–745. DOI: 10.1016/0305-0491(82)90311-X.

Organ CL, Canoville A, Reisz RR, Laurin M. 2011. Paleogenomic data suggest mammal-like genome size in the ancestral amniote and derived large genome size in amphibians. Journal of Evolutionary Biology 24:372–380. DOI: 10.1111/j.1420-9101.2010.02176.x.

Organ CL, Shedlock AM. 2009. Paleogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. Biology Letters 5:47–50. DOI: 10.1098/rsbl.2008.0491.

Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV. 2007. Origin of avian genome size and structure in non-avian dinosaurs. Nature 446:180–184. DOI: 10.1038/nature05621.

Organ CL, Struble M, Canoville A, de Buffrénil, Laurin M. 2015. Macroevolution of genome size in sarcopterygians during the water-land transition. Comptes Rendus Palevol. DOI: 10.1016/j.crpv.2015.09.003.

Pasquesi GIM, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, Reyes-Velasco J, Ruggiero RP, Vandewege MW, Shortt JA, Castoe TA. 2018. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. Nature Communications 9:2774. DOI: 10.1038/s41467-018-05279-1.

Roth G, Nishikawa KC, Wake DB. 1997. Genome size secondary simplification and the evolution of the brain in salamanders. Brain, Behavior and Evolution 50:50–59. DOI: 10.1159/000113321.

Schmid M, Evans BJ, Bogart JP. 2015. Polyploidy in Amphibia. Cytogenetic and Genome Research 145:315–330. DOI: 10.1159/000431388.

Sessions SK, Larson A. 1987. Developmental correlates of genome size in plethodontid salamanders and their implications for genome evolution. Evolution 41:1239–1251. DOI: 10.1111/j.1558-5646.1987.tb02463.x.

Smith JD, Bickham JW, Gregory TR. 2013. Patterns of genome size diversity in bats (order Chiroptera). Genome 56:457–472. DOI: 10.1139/gen-2013-0046.

Sotero-Caio CG, Platt 2nd RN, Suh A, Ray DA. 2017. Evolution and diversity of transposable elements in vertebrate genomes. Genome Biology and Evolution 9:161–177. DOI: 10.1093/gbe/evw264.

Stingo V, Du Buit MH, Odierna G. 1980. Genome size in some selachian fishes. Italian Journal of Zoology 47:129–137.

Sun C, Mueller RL. 2014. Hellbender genome sequences shed light on genome expansion at the base of crown salamanders. Genome Biology and Evolution 6:1818–1829. DOI: 10.1093/gbe/evu143.

Takeda S, Sugimoto K, Otsuki H, Hirochika H. 1999. A 13-pb *cis*-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant Journal 18:383–393. DOI: 10.1046/j.1365-313X.1999.00460.x.

Tarallo A, Angelini C, Sanges R, Yagi M, Agnisola C, D'Onofrio G. 2016. On the genome base composition of teleosts: The effect of environment and lifestyle. BMC Genomics 17:173. DOI: 10.1186/s12864-016-2537-1.

Thomson KS. 1972. An attempt to reconstruct evolutionary changes in the cellular DNA content of lungfish. Journal of Experimental Zoology 180(3):363–371. DOI: 10.1002/jez.1401800307.

Thomson KS, Muraszko K. 1978. Estimation of cell size and DNA content in fossil fishes and amphibians. Journal of Experimental Zoology 205:315–320. DOI: 10.1002/jez.1402050216.

Uliano E, Chaurasia A, Bernà L, Agnisola C, D'Onofrio G. 2010. Metabolic rate and genome GC. What we can learn from teleost fish. Marine Genomics 3:29–34. DOI: 10.1016/j.margen.2010.02.001.

Varriale A. 2014. DNA methylation, epigenetics, and evolution in vertebrates: Facts and challenges. International Journal of Evolutionary Biology 2014:475981. DOI: 10.1155/2014/475981.

Varriale A, Bernardi G. 2006. DNA methylation in reptiles. Gene 385:122–127. DOI: 10.1016/j.gene.2006.05.034.

Vinogradov AE. 1994. Measurement by flow cytometry of genome AT/GC ratio and genome size. Cytometry 16:34–40. DOI: 10.1002/cyto.990160106.

Vinogradov AE. 1995. Nucleotypic effect in homeotherms body-mass-corrected basal metabolic rate of mammals is related to genome size. Evolution 49:1249–1259. DOI: 10.1111/j.1558-5646.1995.tb04451.x.

Vinogradov AE. 1997. Nucleotypic effect in homeotherms: Body-mass independent resting metabolic rate of passerine birds is related to genome size. Evolution 51:220–225. DOI: 10.1111/j.1558-5646.1997.tb02403.x.

Vinogradov AE. 1998a. Genome size and GC-percent in vertebrate as determined by flow cytometry: The triangular relationship. Cytometry 31:100–109. DOI: 10.1002/(SICI)1097-0320-(19980201)31:2<100::AID-CYTO5>3.0.CO;2-Q.

Vinogradov AE. 1998b. Buffering: A possible passive-homeostasis role for redundant DNA. Journal of Theoretical Biology 193:197–198. DOI: 10.1006/jtbi.1997.0629.

Vinogradov AE. 1998c. Variation in ligand-accessible genome size and its ecomorphological correlates in a pond snail. Hereditas 128:59–65. DOI: 10.1111/j.1601-5223.1998.00059.x.

Vinogradov AE. 1999. Genome in toto. Genome 42:361–362. DOI: 10.1139/g98-117.

Vinogradov AE. 2000. Larger genomes for molluskan land pioneers. Genome 43:211–212. DOI: 10.1139/g99-063.

Vinogradov AE. 2003. DNA helix: The importance of being GC-rich. Nucleic Acid Research 31:1838–1844. DOI: 10.1093/nar/gkg296.

Vinogradov AE. 2005. Genome size and chromatin condensation in vertebrates. Chromosoma 113:362–369. DOI: 10.1007/s00412-004-0323-3.

Vinogradov AE, Anatskaya OV. 2006. Genome size and metabolic intensity in tetrapods: A tale of two lines. Proceedings of the Royal Society B: Biological Science 273:27–32. DOI: 10.1098/rspb.2005.3266.

Womack MC, Metz MJ, Hoke KL. 2019. Larger genomes linked to slower development and loss of late-developing traits. The American Natural 194:654–664. DOI: 10.1086/705897.

Xia X. 1995. Body temperature, rate of biosynthesis, and evolution of genome size. Molecular Biology and Evolution 12:834–842. DOI: 10.1093/oxfordjournals.molbev.a040260.

3.2 Mobile elements in ray-finned fish genomes

life

MDPI

Review

Mobile Elements in Ray-Finned Fish Genomes

Federica Carducci, Marco Barucca, Adriana Canapa, Elisa Carotti and Maria Assunta Biscotti *

Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche, 60131 Ancona, Italy; f.carducci@univpm.it (F.C.); m.barucca@univpm.it (M.B.); a.canapa@univpm.it (A.C.); e.carotti@pm.univpm.it (E.C.)

* Correspondence: m.a.biscotti@univpm.it; Tel.: +39-071-2204625

Received: 29 July 2020; Accepted: 22 September 2020; Published: 25 September 2020

Abstract: Ray-finned fishes (Actinopterygii) are a very diverse group of vertebrates, encompassing species adapted to live in freshwater and marine environments, from the deep sea to high mountain streams. Genome sequencing offers a genetic resource for investigating the molecular bases of this phenotypic diversity and these adaptations to various habitats. The wide range of genome sizes observed in fishes is due to the role of transposable elements (TEs), which are powerful drivers of species diversity. Analyses performed to date provide evidence that class II DNA transposons are the most abundant component in most fish genomes and that compared to other vertebrate genomes, many TE superfamilies are present in actinopterygians. Moreover, specific TEs have been reported in ray-finned fishes as a possible result of an intricate relationship between TE evolution and the environment. The data summarized here underline the biological interest in Actinopterygii as a model group to investigate the mechanisms responsible for the high biodiversity observed in this taxon.

Keywords: ray-finned fishes; transposable elements; genome evolution

1. Introduction

Actinopterygii is one of the most diverse groups of vertebrates, with over 30,000 species [1]. This taxonomic group includes Polypteriformes, Acipenseriformes, Holostei, and Teleostei. The order Polypteriformes comprises a unique family, Polypteridae, with two extant genera, *Erpetoichthys* (including only one species) and *Polypterus* (including eleven species). These fish appeared during the Cretaceous and inhabit equatorial and subequatorial freshwater areas in Africa. Several molecular phylogenetic analyses have shown that bichirs are a basal lineage of the ray-finned fishes [2–5]. Data about genome size are available for only four species, with values ranging from 3.69 to 7.25 pg/N [6] (Figure 1), while entire genomic DNA sequencing has been performed for only *Erpetoichthys calabaricus* (Ensembl release 99—January 2020). The order Acipenseriformes includes two families: Acipenseridae (sturgeons), with four extant genera (*Acipenser*, *Huso*, *Scaphirhynchus*, and *Pseudoscaphirhynchus*), contains 24 species, and Polyodontidae (paddlefish), with two extant genera (*Polyodon* and *Psephurus*), contains three species. These fish are widely distributed in the rivers, lakes, and seas of Northern Hemisphere countries. Data from 17 species reveal that these taxa have a very wide range of genome sizes, spanning from 1.22 to 9.32 pg/N [6]. Moreover, the high value (9.32 pg/N) recorded in *Acipenser brevirostrum* is the highest value among ray-finned fishes (Figure 1).

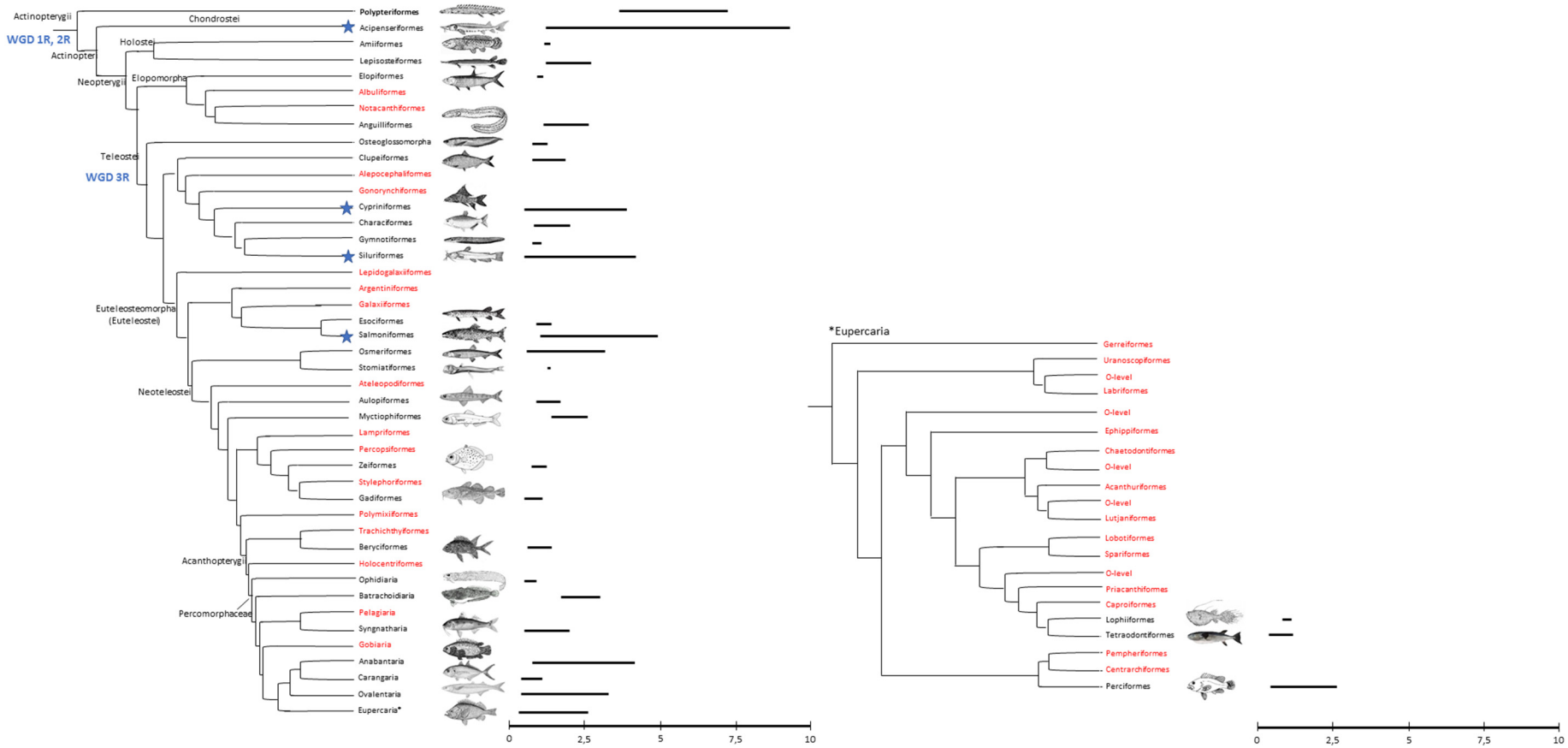

Figure 1. On the left: cladogram showing the relationships of bony fishes (modified from Betancur-R et al. [7]), with the relative DNA content from the genome size database [6]. The teleost orders lacking genome size information are shown in red. Whole-genome duplications (WGDs) are shown in blue: WGD 1R and 2R occurred in the common ancestor of vertebrates; WGD 3R occurred in the common ancestor of teleosts. The blue stars indicate taxa that underwent further independent WGD events. On the right: a separated cladogram of the Eupercaria clade with its relative nuclear DNA content. Order-level incertae sedis (O level in this figure) includes families awaiting evidence to clarify their phylogenetic status [7].

Holostei and Teleostei are part of the Neopterygii, a group of fishes that appeared in the late Permian and were characterized by better swimming capabilities and feeding mechanisms that allowed them to colonize a wider range of habitats. The infraclass Holostei comprises two orders: Amiiformes (bowfins) has a single family, Amiidae, which includes only one living species, *Amia calva*; Lepisosteiformes also has a single family, Lepisosteidae (gars), including seven species subdivided into two genera, *Atractosteus* (three species) and *Lepisosteus* (four species). *Amia calva* is found in freshwater rivers and lakes of North America, while gars are distributed in fresh, brackish, and, sometimes, marine waters of North and Central America. Genome sizes are available for only three species of Holostei, with an average of 1.29 pg/N [6], and only the genome of *Lepisosteus oculatus* has been fully sequenced (Ensembl release 99—January 2020). Teleostei is the most successful group of ray-finned fishes, with more than 24,000 species subdivided into 72 orders [7]. Teleosts appeared in the Late Triassic, and their evolutionary radiation occurred during the Mesozoic and Cenozoic. They have adapted to aquatic environments worldwide, from salt to fresh waters, from cold to warm seas, and from high-elevation mountain lakes to extreme sea depths [8]. In the genome size database, more than 1800 teleost species are listed, with values ranging from 0.34 pg/N (the most compact genome among vertebrates, recorded in a species of the Tetraodontidae family) to 4.90 pg/N (in a species belonging to the Salmonidae) [6], and the genomes of more than 300 species have been sequenced (NCBI genome database).

Knowledge of genome composition and architecture is fundamental in the comprehension of the evolutionary processes responsible for fish radiation. The advent of high-throughput sequencing technologies and bioinformatics has provided a great amount of genomic data, which has been extremely useful for obtaining insights into the evolution of ray-finned fish genomes. Indeed, the identification of gene loss and duplication events, genomic arrangements, variation in base composition, and different selection pressures on specific genomic regions have been highlighted through comparative genomics [8]. Moreover, it is known that, as in other vertebrates, the genome of actinopterygians underwent two rounds of whole-genome duplication (WGD) [9–11]. A third event occurred in teleosts approximately 226–350 Mya, leading to duplicated genes that were probably responsible for the radiation of these clades [11–13]. A fourth round of whole-genome duplication occurred independently in Neotropical Corydoradinae catfishes between 54 and 66 Mya [14], in salmonids approximately 88–103 Mya [15,16] and in cyprinids, approximately 5.6–11.3 Mya [17]. Another interesting aspect of the fish genome is the GC content. On the basis of GC percentage, genomes can be divided into five isochores, i.e., regions longer than 300 kb, with a high degree of uniformity in guanine and cytosine: two light isochores, L1 with 34–36% and L2 with 37–40%, and three heavy isochores, H1 with 41–45%, H2 with 46–52%, and H3 with more than 53%. While the genomes of mammals and birds contain heavy isochores, resulting in GC heterogeneity, fish and amphibian genomes have isochores with low CG content and typically have two light isochores [18]. Excluding salmonids, a negative correlation between genome size and genomic GC% in fish has been reported [19]. In contrast, nonteleost gars possess an AT/GC compartmentalized genome [20], and their closest living relatives, the bowfin *Amia calva*, have a typical teleost-like AT/GC homogenous genome, despite being nonteleosts [19].

Finally, the presence of mobile elements in fish genomes most likely contribute to shaping their genomes, providing advantageous features that have allowed them to successfully adapt to different environments [21–23]. Given the ability to move throughout the genome, the impact of mobile elements on genome evolution is higher than commonly supposed, and several papers have recognized the role of these elements as one of the most powerful evolutionary tools [23]. On the basis of these premises, this review is focused on the importance of mobile genetic elements for the genomes of actinopterygian, one of the most diverse vertebrate groups (Table 1).

Table 1. Table summarizing works published to date about transposable elements in teleost orders.

Teleost Orders	References
Polypteriformes	
Acipenseriformes	[24,25]
Amiiformes	
Lepisosteiformes	[19,20,26–35]
Elopiformes	[36]
Albuliformes	
Notacanthiformes	
Anguilliformes	[24,32,36–62]
Osteoglossomorpha	[63,64]
Clupeiformes	[65,66]
Alepocephaliformes	
Gonorynchiformes	[67]
Cypriniformes	[14,68–76]
Characiformes	[77–91]
Gymnotiformes	[92–94]
Siluriformes	[14,31,95–100]
Lepidogalaxiiformes	
Argentiniformes	[101]
Galaxiiformes	[102]
Esociformes	[103]
Salmoniformes	[104–107]
Osmeriformes	[108,109]
Stomiatiformes	[109]
Ateleopodiformes	[109]
Aulopiformes	[109,110]
Myctiophiformes	[109]
Lampriformes	[109]
Percopsiformes	[109]
Zeiformes	[109]
Stylephoriformes	[109]
Gadiformes	[109,111,112]
Polymixiiformes	[109]
Trachicthtyiformes	
Beryciformes	[34,109,113–116]
Holocentriformes	[109]
Ophidiaria	[109,111,117]

Table 1. *Cont.*

Teleost Orders	References
Batrachoidiaria	[109]
Pelagiaria	[109]
Synghatharia	[118]
Gobiaria	[109]
Anabantaria	[109,119]
Carangaria	[109,120–122]
Ovalentaria	[109,123–126]
Eupercaria	[30,31,109,119,127–152]

2. Mobile Elements

Mobile elements are genetic elements capable of moving throughout the genome by a transposition mechanism. The effect of their movement can be deleterious for the host genome if they interrupt genes; in contrast, it can lead to advantageous innovations, creating new genes or regulatory sequences through a process called molecular domestication.

On the basis of their transposition intermediate, either RNA or DNA, mobile elements can be distinguished into two main classes, according to the classification proposed by Wicker and colleagues [153] (Figure 1).

Class I elements transpose via RNA intermediaries and are characterized by a *copy and paste* transposition mechanism. Their RNA intermediate is reverse-transcribed into its complementary DNA by a reverse transcriptase (RT) encoded by the mobile element. Reverse transcription is followed by reintegration into the host genome. Through the *copy and paste* mechanism of transposition, Class I elements are the main source of increased repetitive fractions, thereby having a major impact in large genomes [153–155]. Class I mobile elements are composed of long terminal repeat (LTR) and non-LTR subclasses. LTR retrotransposons are characterized by long terminal repeats that confer the ability to transpose. For exogenous retroviruses, LTR retrotransposons are structurally composed of *gag* and *pol* genes; *gag* genes encode viral structural particles and *pol* genes encode the whole retrotranscription machinery (*reverse transcriptase*, *ribonuclease H*, and *integrase*; Figure 2). In contrast to LTR retrotransposons, exogenous retroviruses possess the *env* gene, which encodes the viral envelope. However, traces of the *env* gene have been found in LTR retrotransposons [156]. *DIRS*, considered more complex LTR retroelements [37], are structurally characterized by a *tyrosine recombinase* (*YR*) instead of an *integrase* and by inverted terminal repeats. Long and short interspersed nuclear elements (LINEs and SINEs) are non-LTR retrotransposons. Of these, LINEs are autonomous retroelements constituted by two open reading frames (ORFs) and a poly A tail at the 3′ end. Generally, ORF2 encodes a reverse transcriptase and an endonuclease protein [153]. In contrast, SINEs are *RT*-lacking retroelements, and they need RT encoded by autonomous elements to transpose [157]. Finally, another group of Class I elements, Penelope retroelements, must be considered separately due to their very large diversity in terms of structural features. The common components are pseudo-LTRs (pLTRs), a *reverse transcriptase*, and an *endonuclease* [158,159].

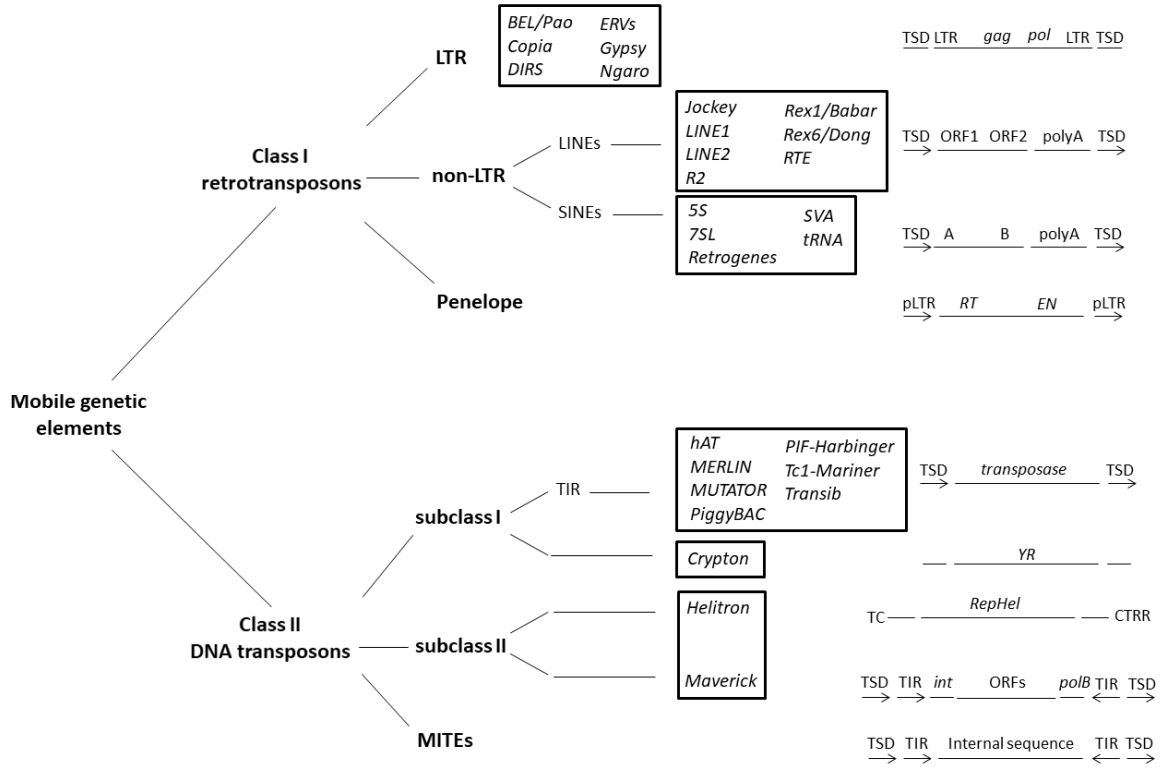

Figure 2. Mobile genetic element classification based on transposition mechanisms (according to Wicker et al. [153]). The main structural components of the elements are reported on the right (modified from Makalowsky et al. [155]).

Class II mobile elements use a DNA intermediate to transpose their genomic DNA copies into a novel chromosomal position [160,161] and can be divided into subclasses I and II. Subclass I consists of two main elements: TIR and *Crypton*. *TIRs* are autonomous elements characterized by terminal inverted repeats (TIRs) and a transposase through which transposition occurs via a *cut and paste* mechanism, in which both DNA strands are cleaved. The DNA transposons *hAT*, *Merlin*, *Mutator*, *PiggyBac*, *PIF-Harbinger*, *Tc1-Mariner*, and *Transib* can be found in this subclass. *Crypton* elements use a tyrosine recombinase (YR) in a transposition mechanism, probably involving recombination between a circular intermediate and the DNA target [37]. *Helitrons* and *Maverick* are the two major representative elements of subclass II. These DNA elements transpose via a *copy and paste* mechanism [153]. *Helitron* DNA transposons replicate using a rolling-circle mechanism and encode for replication initiation (Rep) and a DNA helicase (Hel) [162], while *Maverick* transposons encode for an integrase, an ORF, and polymerase B. For polymerase B, transposition involves a single-strand excision phase, extrachromosomal replication, and consequent reintegration into a new location [163]. Miniature Inverted Transposable Elements (MITEs), also grouped in Class II, do not encode a transposase; therefore, they exploit transposases encoded by autonomous elements to move throughout the genome [164].

3. Transposable Elements in Actinopterygians

The evolutionary dynamics of TEs are different in several lineages, which strongly support their pivotal role in genome evolution. The evaluation of mobile element impact on the actinopterygian genome is a fundamental step toward understanding the biodiversity of this taxon. With increasing genomic resources, a clear positive correlation between genome size and the percentage of TEs has been found in ray-finned fish [21,37,68,164,165]. Moreover, a wide range of TE amounts has been recorded

in this taxon, with only 6% in the compact pufferfish genome and 55% in the zebrafish genome [165] (Figure 1).

Data published to date suggest that compared to other vertebrate genomes, class II DNA transposons are the most abundant component in most fish genomes [32,165]. Most TE superfamilies (i.e., *Gypsy*, *BEL/Pao*, ERV, *DIRS*, *Penelope*, *Rex6/Dong*, *R2*, *L1*, *RTE*, *L2*, *Rex1/Babar*, *Jockey*, *Helitron*, *Maverick*, *Zisupton*, *Tcl-Mariner*, *hAT*, *PIF-Harbinger*, *PiggyBac*, and *EnSpm*) are present in the actinopterygian genome, evidencing a higher diversity than that in other vertebrates [165]. Among them, *Tc/mariner*, *hAT*, *L1*, *L2*, and *Gypsy* are the most widespread and predominant TE superfamilies in fish genomes [31,68]. Comparing the distribution of the transposon superfamilies among the actinopterygians, the Cyprinidae family presents the highest level of TE diversity [165]. However, some organisms present a predominance of specific TE superfamilies, such as *Gypsy* in *Boleophthalmus pectinirostris*, *L2* and *RTE* in *Nothobranchius furzeri*, *Tc/mariner* in *Astyanax mexicanus*, and *hAT* in *Danio rerio* [165]. These elements have been preserved in the genomes of these organisms, and, thus, they could have had a pivotal role in their evolution. Shao and colleagues [165] proposed that the interaction between TEs and host genomes is comparable to that between organisms and their environments, explained by the Red Queen paradigm: harmful TEs are eliminated by host genomes, while beneficial TEs are instead preserved. Moreover, a critical role of *CR1* in vertebrate evolution has been reported by the same authors. The low copy number of *CR1* elements found in teleosts, contrary to primitive fishes and sarcopterygians, suggests the preservation and proliferation of these elements during the transition from water to land in tetrapods [165].

In the deeply branched nonteleost ray-finned fishes, the mobilome has been inferred from the genomes of the sturgeon *Acipenser ruthenus* and the spotted gar, *L. oculatus*. The former has a similar pattern to that observed in teleosts [166], while the latter shows a predominance of non-LTR retrotransposons [32,37]. The condition observed in spotted gar is also common to the elephant shark, *Callorhinchus milii*, and the lamprey, *Petromyzon marinus* [32]. The amount of non-LTR in bony fishes might be due to the presence of mechanisms restricting the invasion of retroelements in their genomes [31].

Another interesting feature of the ray-finned fish mobilome is the presence of more recent TE copies than those found in other vertebrate lineages. In particular, cod, stickleback, and fugu have very recent TE copies, and differences in TE activity can also be observed between species closely related to medaka and platyfish [37]. Kimura distance-based copy divergence analysis performed on 35 actinopterygians shows one or, at most, two TE amplification bursts [32,37,68,165]. These events were preceded by periods in which new elements arose through genetic mutations or where TEs invaded the host genome through horizontal transfer. Subsequently, natural selection and defense mechanisms of the host genome select beneficial mobile elements, and a period of coexistence between TEs and the host genome begins. These steps, which occurred during the history of TE activity, are associated with species radiation [148,167–170], suggesting that TEs are responsible for important evolutionary events.

L. oculatus is a nonteleost ray-finned fish that has not undergone further WGDs after those that have occurred at the base of vertebrates (1R and 2R WGDs). The quantitative analysis of TEs showed no differences among teleosts. This finding does not support any link between ancestral genome duplication and TE expansion in the teleost lineage [32]. The analysis of the *Salmo salar* genome revealed an expansion of DNA transposons, with a return to the diploid state after the 4R WGD [106]. The rediploidization is also achieved through the contraction of the genome associated with TE loss. This could explain the loss of *Rex3*, a teleost-specific non-LTR retroelement, absent in salmonids [61].

A positive correlation has been reported between the GC content of TEs and genomes [19]. Analyzing the GC% in the main TE groups, Class I retrotransposons, with 45.6%, are more GC-rich than Class II DNA transposons, with 40.1%; DIRS are the TEs with the highest GC content (53.8%), while the *CMC* transposons are the mobile elements with the lowest GC content (35.8%) in fish genomes. The GC-poor DNA transposons seem to be responsible for the overall GC homogenization of fish genomes.

4. Rex Retroelements

Rex retroelements are repeated elements that are widely distributed among teleost genomes and were deeply active during the evolution of this lineage [59–61]. Published in 1999 by Volff and his research team [59], the first report of three *reverse transcriptase* (*RT*)-carrying retrotransposons in the model fish *Xiphophorus maculatus* is attributable to the origin of the name *Rex* for this class of fish-specific retroelements.

A sequence derived from the Y chromosome of *X. maculatus* of the Rio Jamapa allowed Volff and colleagues firstly to isolate a truncated copy *Rex1-Ximj*, and then to evidence many other copies of this non-LTR retrotransposon in different teleost species, defining a second class of *Rex* retroelements named *Rex1*. Concerning their main structural features, *Rex1* non-LTRs are characterized by an *RT*, an apurinic/apyrimidinic (A/P*) site that can be located upstream or downstream of the *RT*-encoding region, and a 3′-UTR region. On the other hand, *Rex3* and *Rex6* retroelements harbor a gene encoding an endonuclease (*EN*) in addition to *RT*.

A high copy number of a novel class of *Rex* retroelements, the so-called *Rex6* elements, was further evidenced by Volff and colleagues in 2001 in the genomes of several teleosts [61]. *Rex6* is a member of the *R4* family [171] of non-LTR retrotransposons and it encodes a specific type of endonuclease, closely related to the type IIS restriction enzymes isolated from trypanosomes, nematodes, and arthropods [172]. *Rex6* elements have been found in many teleost orders: Anguilliformes [55,59], Beloniformes [60,61], Carangiformes [129], Centrarchiformes [59], Characiformes [79,82,149,173–175], Cichliformes [60,132, 133,136,176–180], Cypriniformes [59], Cyprinodontiformes [59], Esociformes [60], Perciformes [60,181], Salmoniformes [60], Siluriformes [174,182–184], and Tetraodontiformes [60,61,134,142,185].

Although no evolutionary relationship among *Rex1*, *Rex3*, and *Rex6* has emerged to date, they are usually considered together in fluorescence in situ hybridization (FISH) studies, demonstrating their key role in karyotype evolution in fish (for review, see Carducci et al. [186]). Overall, their localization has been observed in heterochromatin at telomeric [82,129,173,183], pericentromeric, and centromeric regions [149,177,180,181,183] and in supernumerary chromosomes [79,176]. Of extreme interest is the non-negligible number of papers underlying the localization of *Rex* retroelements at the euchromatic level [129,134,179,182,184], strongly supporting the relatively high rate of gene-linkage disruption and chromosomal rearrangements in teleost genomes [149]. In general, the distribution of *Rex* retroelements in chromosomes varies considerably between teleost orders and families [186].

All the papers reviewed herein highlight the significant role of the *Rex* retroelements in the rapid evolution of teleosts, in particular, acting on karyotype and genome structure.

5. Endogenous Retroviruses

Retroviruses are viruses constituted by a single-stranded positive-sense RNA. After infection, a retrovirus reaches the cytoplasm of the host cell, where a reverse transcriptase (RT) converts its ssRNA into cDNA, ready to be integrated into the nuclear genome of the infected host cell. Once integrated, the provirus will exploit the nuclear machinery of the host cell to transcribe and translate its components. There are four main components of the basic toolkit of retrovirus genomes: long terminal repeats (LTRs), which carry a promoter sequence that mediates the interaction with integrase for retrovirus integration into the host cell genome; *gag* (*group-specific antigen*) genes, which encode structural protein components; *pol* (*polymerase*) genes, which enclose the RT, protease, and integrase domains; *env* (*envelope*) genes, which encode coat proteins [187]. Structurally, retroviruses differ from retrotransposons by the presence of genes encoding envelope proteins. Moreover, a characteristic hallmark that allows the identification of a past retrovirus infection is provided by the solo-LTR derived from ectopic homologous recombination between two LTRs [187]. Infection by a retrovirus may occur within a germline, leading to the generation of endogenous viral elements, the so-called endogenous retroviruses (ERVs).

ERVs are inherited through vertical transmission and consequently maintained within the host genome over millions of years [188]. Identified in all vertebrate lineages [189] and all belonging to the Retroviridae family, ERVs can be approximately grouped into three main classes based on the

phylogenetic relationships between the seven exogenous retrovirus genera identified: Class I (closely related to Gammaretroviruses and Epsilonretroviruses), Class II (closely related to Betaretroviruses), and Class III (Spumavirus-like elements) [190]. Hayward and colleagues [189] have identified two further clades: human endogenous retroviruses S/L (HERVS/L)-like and snakehead fish retrovirus (SnRV)-like elements.

Naville and Volff [191] have shown that the overall ERV content in fish genomes ranges from 0.01 to 1%. In particular, epsilon-related retroviruses are the most frequent ERVs in ray-finned fishes [192]. The lowest value reported is 0.033% for *Takifugu rubripes* (with approximately 1800 insertions), and the maximum is 0.76% in *Danio rerio* (with more than 30,000 insertions) [191]. The best-studied ERV element in teleosts is Zebrafish Endogenous Retrovirus (ZFERV), isolated from zebrafish [193].

In addition to epsilon-related retroviruses and Snakehead fish retrovirus (SnRV)-like elements, endogenous foamy virus (EFV) sequences have been detected in different teleost species, including cod, platyfish, and zebrafish [194,195]. No reports of gamma or Class II elements have been described to date [189].

Whereas the evolutionary importance of ERVs, as a source of new genes [196] and, in general, as a mediator of gene expression [197,198] in catalyzing genome evolution, has been evidenced in mammals, nothing is known about the roles of ERVs in teleost evolution [191].

The complex evolutionary history of retroviruses has been recently investigated by Xu and colleagues [190]. Through an extensive genomic and phylogenetic analysis performed on species representing the main evolutionary lineages, of which 66 were ray-finned fishes, the authors unveiled the role of teleosts and turtles, as vehicles for retrovirus transmission, in overcoming the water–land barrier.

6. TEs and Sex Chromosomes

The wide chromosomal diversity in teleosts (e.g., interspecific diploid number variation; the presence or absence of sex and supernumerary chromosomes) has been suggested to be correlated with the ability to incorporate transposable elements [199]. The evolutionary success of TEs in a given population is strictly linked to their persistence, which is obtainable through TE vertical transmission in the germline, from one to the next generation [199]. Moreover, the accumulation of repetitive sequences is a common phenomenon in sex chromosomes, characterized by the absence of recombination [200].

Several papers have reported the involvement of *Rex* retrotransposons in the differentiation of sex chromosomes [149,173,175,181], with a key role played by *Rex6* [173]. These elements have been mapped on the sex chromosomes of four species belonging to the Characiformes [173,175] and one species of Perciformes [181] and on the largest pair of chromosomes recognized as sexual chromosomes in one species belonging to the Cichliformes [180]. Other convincing examples of the role of TEs in the control of sexual development and function have been recently reviewed by Dechaud and colleagues [199]. A clear example of TE control in a germline through *cis*-regulation was reported in the medaka, *Oryzias latipes*: a *LINE/Rex1* retroelement was found within the nonautonomous P element *Izanagi*, corresponding to the upstream region of the master sex-determining gene (*dmrt1bY*) in medaka. In particular, the *LINE/Rex1*-derived sequence located within the *Izanagi* element carries the binding site for Sox5, a transcriptional factor involved in the regulation of *dmrt1bY* [201,202]. A role of TEs in the determination of sex chromosome structure and evolution has also been observed in *X. maculatus*, in which the accumulation and spreading of *Texim* genes in only the Y chromosome is due to the activity of *Helitron* transposons, deeply influencing the evolution of this chromosome in platyfish [203]. Finally, in salmonids, analyses performed on the boundary regions of the master sex-determining gene (*sdY*) have shown a certain accumulation of TEs, which is probably responsible for the different *sdY* gene chromosomal locations despite their conservation [105,204].

7. Fish Transposons and the Environment

Both abiotic and biotic factors are continuously changing, resulting in new selective pressures that challenge population survival. To cope with these changes, organisms colonize new habitats and exploit

their phenotypic plasticity and/or adaptive evolutionary traits. Natural selection allows organisms with features appropriate for a specific environment to survive and, thus, to reproduce, increasing their fitness. Genetic variants will be transmitted to the next generation, increasing in frequency in the population. Genetic variation can be caused by not only point mutations and whole-genome duplications but also TE activity. Moreover, transposons can be co-opted and exapted, creating regulatory sequences, coding exons, or entirely new genes useful for the host genome [22,23,37,205]. Indeed, a great number of reports have suggested the responsiveness and susceptibility of TEs to environmental changes or stressful conditions [206–214]. Yuan and colleagues [31] analyzed 52 fishes and reported an increase in the DNA transposons in bony fish living in freshwaters and an abundance of tandem repeats in marine species that was not explained by phylogenetic relationships. In particular, among DNA transposons, *Tc1* is the most well-represented in freshwater bony fishes. This association clearly suggests a potential role of TEs in the adaptation of fish to their living environments. Freshwater environments might encourage the proliferation and spread of DNA transposons, probably because transposition can cause new genetic variants useful for host adaptation to the environment. According to these authors, the large number of repetitive elements can contribute to the generation of novel genes useful for adaptability to the environment. Moreover, the presence of such a high content of repetitive elements can cause unstable genomes due to recombination and splicing events. Due to natural selection, uncontrollable increases in genome size do not occur. Auvinet and colleagues [148] reported a preferential accumulation of four families of *DIRS1* in specific chromosomal locations of the Antarctic teleost species belonging to the *Trematomus* genus. According to these authors, the concentration of these TEs in pericentromeric and centromeric areas could have been mediated by multiple glacial-interglacial cycles that took place in the Antarctic region. The variation in temperature probably led to changes in epigenetic regulation that have allowed TE bursts. An interesting correlation between TEs and the environmental temperature has also been evidenced by our group in a recent publication [207], in which a phylogenetic analysis was performed on the partial *reverse transcriptase* of the *Rex3* retroelement in 39 teleost species. Surprisingly, in this investigation, there was a lack of correspondence with the canonical taxonomy relationships. Indeed, the *Rex3* sequences analyzed clustered into two groups, strictly related to the environmental temperature in which these species live, suggesting a selective role of temperature on specific TE sequence variants.

8. Conclusions

Actinopterygii is a taxon characterized by a high diversity of species adapted to a wide range of environments. There is generally a positive correlation between genome size and TE coverage, and the major contributors to the genome size variation are DNA transposons. The data summarized here show that the ray-finned fish genomes are unique among vertebrates in their overall TE composition. The high level of TE diversity suggests that these genetic elements represent an important evolutionary tool that has had a pivotal role in fish evolution. However, it is not clear whether repetitive elements lead to environmental adaptation or vice versa [31].

Moreover, significant differences are also evident in TE activity, which might be linked to body temperature and host defense mechanisms. Indeed, body temperature is influenced by environmental conditions, which could affect the activity of the proteins involved in transposition mechanisms; the capacity to replicate and compete with other TEs is influenced by host defense mechanisms, such as piRNAs and methylation. However, information about genome size and data on genome sequencing in ray-finned fishes is still limited. Such is the case for the deep-branching nonteleost ray-finned fishes belonging to the Polypteriformes; the investigation of the genomes of these taxa could be extremely useful for providing information on the common ancestor of TEs among actinopterygian species.

Author Contributions: Conceptualization, M.A.B. and F.C.; writing—original draft preparation, F.C., M.B., A.C., E.C. and M.A.B.; supervision, M.A.B. and A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from Ministero della Ricerca e dell'Istruzione (project number 59, 2020).

Conflicts of Interest: The authors declare no conflict of interests. The funder did not participate in any phase of the study, including the publication of data.

References

1. Froese, R.; Pauly, D. FishBase. World Wide Web Electronic Publication. Version (12/2019). Available online: www.fishbase.org (accessed on 5 June 2020).
2. Noack, K.; Zardoya, R.; Meyer, A. The complete mitochondrial DNA sequence of the bichir (*Polypterus ornatipinnis*), a basal ray-finned fish: Ancient establishment of the consensus vertebrate gene order. *Genetics* **1996**, *144*, 1165–1180.
3. Venkatesh, B.; Erdmann, M.V.; Brenner, S. Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11382–11387.
4. Inoue, J.G.; Miya, M.; Tsukamoto, K.; Nishida, M. Basal actinopterygian relationships: A mitogenomic perspective on the phylogeny of the "ancient fish". *Mol. Phylogenet. Evol.* **2003**, *26*, 110–120.
5. Kikugawa, K.; Katoh, K.; Kuraku, S.; Sakurai, H.; Ishida, O.; Iwabe, N.; Miyata, T. Basal jawed vertebrate phylogeny inferred from multiple nuclear DNA-coded genes. *BMC Evol. Biol.* **2004**, *2*, 3.
6. Gregory, T.R. Animal Genome Size Database. Version (12/2019). Available online: http://www.genomesize.com (accessed on 5 June 2020).
7. Betancur-R, R.; Wiley, E.O.; Arratia, G.; Acero, A.; Bailly, N.; Miya, M.; Lecointre, G.; Ortí, G. Phylogenetic classification of bony fishes. *BMC Evol. Biol.* **2017**, *17*, 162.
8. Bian, C.; Huang, Y.; Li, J.; You, X.; Yi, Y.; Ge, W.; Shi, Q. Divergence, evolution and adaptation in ray-finned fish genomes. *Sci. China Life Sci.* **2019**, *62*, 1003–1018.
9. Dehal, P.; Boore, J.L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **2005**, *3*, e314.
10. Postlethwait, J.H.; Woods, I.G.; Ngo-Hazelett, P.; Yan, Y.L.; Kelly, P.D.; Chu, F.; Huang, H.; Hill-Force, A.; Talbot, W.S. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res.* **2000**, *10*, 1890–1902.
11. Vandepoele, K.; de Vos, W.; Taylor, J.S.; Meyer, A.; Van de Peer, Y. 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* **2004**, *101*, 1638–1643.
12. Christoffels, A.; Koh, E.G.L.; Chia, J.M.; Brenner, S.; Aparicio, S.; Venkatesh, B. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.* **2004**, *21*, 1146–1151.
13. Meyer, A.; Van de Peer, Y. From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). *Bioessays* **2005**, *27*, 937–945.
14. Marburger, S.; Alexandrou, M.A.; Taggart, J.B.; Creer, S.; Carvalho, G.; Oliveira, C.; Taylor, M.I. Whole genome duplication and transposable element proliferation drive genome expansion in Corydoradinae catfishes. *Proc. R. Soc. B Biol. Sci.* **2018**, *285*, 20172732.
15. Alexandrou, M.A.; Swartz, B.A.; Matzke, N.J.; Oakley, T.H. Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Mol. Phylogenet. Evol.* **2013**, *69*, 514–523.
16. Macqueen, D.J.; Johnston, I.A. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. R. Soc. B* **2014**, *281*, 2013–2881.
17. Wang, J.T.; Li, J.T.; Zhang, X.F.; Sun, X.W. Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (*Cyprinus carpio*). *BMC Genom.* **2012**, *13*, 96.
18. Costantini, M.; Cammarano, R.; Bernardi, G. The evolution of isochore patterns in vertebrate genomes. *BMC Genom.* **2009**, *10*, 146.
19. Symonová, R.; Suh, A. Nucleotide composition of transposable elements likely contributes to AT/GC compositional homogeneity of teleost fish genomes. *Mob. DNA* **2019**, *10*, 49.
20. Symonová, R.; Majtánová, Z.; Arias-Rodriguez, L.; Morkovský, L.; Korìnková, T.; Cavin, L.; Johnson Pokorná, M.; Dolezálková, M.; Flajshans, M.; Normandeau, E.; et al. Genome compositional organization in

gars shows more similarities to mammals than to other ray-finned fish. *J. Exp. Zool. Mol. Dev. Evol.* **2016**, *328*, 607–619.
21. Canapa, A.; Barucca, M.; Biscotti, M.A.; Forconi, M.; Olmo, E. Transposons, genome size, and evolutionary insights in animals. *Cytogenet. Genome Res.* **2015**, *147*, 217–239.
22. Carducci, F.; Biscotti, M.A.; Barucca, M.; Canapa, A. Transposable elements in vertebrates: Species evolution and environmental adaptation. *Eur. Zool. J.* **2019**, *86*, 497–503.
23. Warren, I.A.; Naville, M.; Chalopin, D.; Levin, P.; Berger, C.S.; Galiana, D.; Volff, J.N. Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. *Chrom. Res.* **2015**, *23*, 505–531.
24. Pujolar, J.M.; Astolfi, L.; Boscari, E.; Vidotto, E.; Barbisan, F.; Bruson, A.; Congiu, L. *Tana1*, a new putatively active *Tc1*-like transposable element in the genome of sturgeons. *Mol. Phylogenet. Evol.* **2013**, *66*, 223–232.
25. Cheng, P.; Huang, Y.; Du, H.; Li, C.; Lv, Y.; Ruan, R.; Ye, H.; Bian, C.; You, X.; Xu, J.; et al. Draft genome and complete *Hox*-cluster characterization of the sterlet (*Acipenser ruthenus*). *Front. Genet.* **2019**, *10*, 776.
26. Braasch, I.; Gehrke, A.R.; Smith, J.J.; Kawasaki, K.; Manousaki, T.; Pasquier, J.; Amores, A.; Desvignes, T.; Batze, P.; Catchen, J.; et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* **2016**, *48*, 4.
27. Hao, S.; Han, K.; Meng, L.; Huang, X.; Shi, C.; Zhang, M.; Wang, Y.; Liu, Q.; Zhang, Y.; Seim, I.; et al. Three genomes of Osteoglossidae shed light on ancient teleost evolution. *BioRxiv* **2020**. [CrossRef]
28. Jiang, J.; Yuan, H.; Zheng, X.; Wang, Q.; Kuang, T.; Li, J.; Liu, J.; Song, S.; Wang, W.; Cheng, F.; et al. Gene markers for exon capture and phylogenomics in ray-finned fishes. *Ecol. Evol.* **2019**, *9*, 3973–3983.
29. Rodriguez, F.; Arkhipova, I.R. Transposable elements and polyploid evolution in animals. *Curr. Opin. Genet. Dev.* **2018**, *49*, 115–123.
30. Symonová, R.; Howell, W.M. Vertebrate genome evolution in the light of fish cytogenomics and rDNAomics. *Genes* **2018**, *9*, 96.
31. Yuan, Z.; Liu, S.; Zhou, T.; Tian, C.; Bao, L.; Dunham, R.; Liu, Z. Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. *BMC Genom.* **2018**, *19*, 141.
32. Chalopin, D.; Volff, J.N. Analysis of the spotted gar genome suggests absence of causative link between ancestral genome duplication and transposable element diversification in teleost fish. *J. Exp. Zool. B Mol. Dev. Evol.* **2017**, *328*, 629–637.
33. Pasquier, J.; Braasch, I.; Batzel, P.; Cabau, C.; Montfort, J.; Nguyen, T.; Jouanno, E.; Berthelot, C.; Klopp, C.; Journot, L.; et al. Evolution of gene expression afterwhole-genome duplication: New insights from the spotted gar genome. *J. Exp. Zool. Mol. Dev. Evol.* **2017**, *328*, 709–721.
34. Gornung, E. Twenty years of physical mapping of major ribosomal RNA genes across the teleosts: A review of research. *Cytogenet. Genome Res.* **2013**, *141*, 90–102.
35. Spaink, H.P.; Jansen, H.J.; Dirks, R.P. Advances in genomics of bony fish. *Brief. Funct. Genomics* **2013**, *13*, 144–156.
36. Kajikawa, M.; Okada, N. LINEs mobilize SINEs in the eel through a shared 3′ sequence. *Cell* **2002**, *111*, 433–444.
37. Chalopin, D.; Naville, M.; Plard, F.; Galiana, D.; Volff, J.N. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* **2015**, *7*, 567–580.
38. Coluccia, E.; Pichiri, G.; Nieddu, M.; Coni, P.; Manconi, S.; Deiana, A.M.; Salvadori, S.; Mezzanotte, R. Identification of two new repetitive elements and chromosomal mapping of repetitive DNA sequences in the fish *Gymnothorax unicolor* (Anguilliformes: Muraenidae). *Eur. J. Histochem.* **2011**, *55*, e12.
39. Durif, C.M.F.; Dufour, S.; Elie, P. Impact of silvering stage, age, body size and condition on reproductive potential of the European eel. *Mar. Ecol. Prog. Ser.* **2006**, *327*, 171–181.
40. Frankowski, J.; Bastrop, R. Identification of *Anguilla anguilla* (L.) and *Anguilla rostrata* (LeSueur) and their hybrids based on a diagnostic single nucleotide polymorphism in nuclear 18SrDNA. *Mol. Ecol. Res.* **2010**, *10*, 173–176.
41. Geeraerts, C.; Belpaire, C. The effects of contaminants in European eel: A review. *Ecotoxicology* **2010**, *19*, 239–266.
42. Gong, X.L.; Li, S.F.; Cai, W.C.; Wang, C.H. Isolation and characterization of polymorphic microsatellite markers from Australia short-finned eel (*Anguilla australis* Richardson). *Mol. Ecol. Res.* **2008**, *8*, 777–779.

43. Ishikawa, S.; Tsukamoto, K.; Nishida, M. Characterization of microsatellite loci from the Japanese eel *Anguilla japonica*. *Mol. Ecol. Notes* **2001**, *1*, 140–142.
44. Itoi, S.; Nakaya, M.; Kaneko, G.; Kondo, H.; Sezaki, K.; Watabe, S. Rapid identification of eels *Anguilla japonica* and *Anguilla anguilla* by polymerase chain reaction with single nucleotide polymorphism-based specific probes. *Fish Sci.* **2005**, *71*, 1356–1364.
45. Jansen, H.J.; Liem, M.; Jong-Raadsen, S.A.; Dufour, S.; Weltzien, F.A.; Swinkels, W.; Koelewijn, A.; Palstra, A.P.; Pelster, B.; Spaink, H.P.; et al. Rapid *de novo* assembly of the European eel genome from nanopore sequencing reads. *Sci. Rep.* **2017**, *7*, 7213.
46. Kitaoka, M.; Okamura, N.; Ichinose, H.; Goto, M. Detection of SNPs in fish DNA: Application of the Fluorogenic Ribonuclease Protection (FRIP) assay for the authentication of food contents. *J. Agric. Food Chem.* **2008**, *56*, 6246–6251.
47. Li, Z.; Chen, F.; Huang, C.; Zheng, W.; Yu, C.; Cheng, H.; Zhou, R. Genome-wide mapping and characterization of microsatellites in the swamp eel genome. *Sci. Rep.* **2017**, *7*, 3157.
48. Maes, G.E.; Pujolar, J.M.; Raeymaekers, J.A.M.; Dannewitz, J.; Volckaert, F.A.M. Microsatellite conservation and bayesian individual assignment in four *Anguilla* species. *Mar. Ecol. Prog. Ser.* **2006**, *319*, 251–261.
49. Pierron, F.; Daffe, G.; Lambert, P.; Couture, P.; Baudrimont, M. Retrotransposon methylation and activity in wild fish (*A. anguilla*): A matter of size. *Environ. Poll.* **2019**, *245*, 494–503.
50. Pujolar, J.M.; Maes, G.E.; Volckaert, F.A.M. Genetic patchiness among recruits in the European eel *Anguilla anguilla*. *Mar. Ecol. Prog. Ser.* **2006**, *307*, 209–217.
51. Suntronpong, A.; Singchat, W.; Kruasuwan, W.; Prakhongcheep, O.; Sillapaprayoon, S.; Muangmai, N.; Somyong, S.; Indananda, C.; Kraichak, E.; Peyachoknagul, S.; et al. Characterization of centromeric satellite DNAs (MALREP) in the Asian swamp eel (*Monopterus albus*) suggests the possible origin of repeats from transposable elements. *Genomics* **2020**, *112*, 3097–3107.
52. Pujolar, J.M.; Maes, G.E.; van Houdt, J.K.J.; Zane, L. Isolation and characterization of expressed sequence tag-linked microsatellite loci for the European eel (*Anguilla anguilla*). *Mol. Ecol. Res.* **2009**, *9*, 233–235.
53. Pujolar, J.M.; Jacobsen, M.W.; Als, T.D.; Frydenberg, J.; Magnussen, E.; Jonsson, B.; Jiang, X.; Cheng, L.; Bekkevold, D.; Maes, G.E.; et al. Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. *Heredity* **2014**, *112*, 627–637.
54. Rozenfeld, C.; Blanca, J.; Gallego, V.; Garcìa-Carpintero, V.; Herranz-Jusdado, J.G.; Pérez, L.; Asturiano, J.F.; Cañizares, J.; Peñaranda, D.S. *De novo* European eel transcriptome provides insights into the evolutionary history of duplicated genes in teleost lineages. *PLoS ONE* **2019**, *14*, e0218085.
55. Suntronpong, A.; Thapana, W.; Twilprawat, P.; Prakhongcheep, O.; Somyong, S.; Muangmai, N.; Peyachoknagul, S.; Srikulnath, K. Karyological characterization and identification of four repetitive element groups (the 18S–28S rRNA gene, telomeric sequences, microsatellite repeat motifs, *Rex* retroelements) of the Asian swamp eel (*Monopterus albus*). *Comp. Cytogenet.* **2017**, *11*, 435–462.
56. Tseng, M.C.; Chen, C.A.; Kao, H.W.; Tzeng, W.N.; Lee, S.C. Polymorphisms of GA/GT microsatellite loci from *Anguilla japonica*. *Mar. Biotechnol.* **2001**, *3*, 275–280.
57. Tu, Z.; Li, S.; Mao, C. The changing tails of a novel short interspersed element in *Aedes aegypti*: Genomic evidence for slippage retrotransposition and the relationship between 3′ tandem repeats and the poly(dA) tail. *Genetics* **2004**, *168*, 2037–2047.
58. Utsunomia, R.; Pansonato-Alves, J.C.; Cardim Scacchetti, P.; Oliveira, C.; Foresti, F. Scattered organization of the histone multigene family and transposable elements in *Synbranchus*. *Genet. Mol. Biol.* **2014**, *37*, 30–36.
59. Volff, J.N.; Körting, C.; Sweeney, K.; Schartl, M. The non-LTR retrotransposon *Rex3* from the Fish *Xiphophorus* is widespread among teleosts. *Mol. Biol. Evol.* **1999**, *16*, 1427–1438.
60. Volff, J.N.; Körting, C.; Schartl, M. Multiple lineages of the non-LTR retrotransposon *Rex1* with varying success in invading fish genomes. *Mol. Biol. Evol.* **2000**, *17*, 1673–1684.
61. Volff, J.N.; Körting, C.; Froschauer, A.; Sweeney, K.; Schartl, M. Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. *J. Mol. Evol.* **2001**, *52*, 351–360.
62. Yi, M.; Chen, F.; Luo, M.; Cheng, Y.; Zhao, H.; Cheng, H.; Zhou, R. Rapid Evolution of piRNA Pathway in the Teleost Fish: Implication for an Adaptation to Transposon Diversity. *Genome Biol. Evol.* **2014**, *6*, 1393–1407.
63. Ráb, P.; Yano, C.F.; Lavoué, S.; Jegede, O.I.; Bertollo, L.A.C.; Ezaz, T.; Majtánová, Z.; de Oliveira, E.A.; Cioffi, M.B. Karyotype and mapping of repetitive DNAs in the African butterfly fish *Pantodon buchholzi*, the sole species of the family Pantodontidae. *Cytogenet. Genome Res.* **2016**, *149*, 312–320.

64. de Souza, F.H.S.; Perez, M.F.; Bertollo, L.A.C.; de Oliveira, E.A.; Lavoué, S.; Gestich, C.C.; Ráb, P.; Ezaz, T.; Liehr, T.; Viana, P.F.; et al. Interspecific Genetic Differences and Historical Demography in South American Arowanas (Osteoglossiformes, Osteoglossidae, *Osteoglossum*). *Genes* **2019**, *10*, 693.
65. Mohindra, V.; Dangi, T.; Tripathi, R.K.; Kumar, R.; Singh, R.K.; Jena, J.K.; Mohapatra, T. Draft genome assembly of *Tenualosa ilisha*, Hilsa shad, provides resource for osmoregulation studies. *Sci. Rep.* **2019**, *9*, 16511.
66. Machado, A.M.; Tørresen, O.K.; Kabeya, N.; Couto, A.; Petersen, B.; Felício, M.; Campos, P.F.; Fonseca, E.; Bandarra, N.; Lopes-Marques, M.; et al. "Out of the can": A draft genome assembly, liver transcriptome, and nutrigenomics of the european sardine, *Sardina pilchardus*. *Genes* **2018**, *9*, 485.
67. Swathi, A.; Shekhar, M.S.; Katneni, V.K.; Vijayan, K.K. Genome size estimation of brackishwater fishes and penaeid shrimps by flow cytometry. *Mol. Biol. Rep.* **2018**, *45*, 951–960.
68. Gao, B.; Shen, D.; Xue, S.; Chen, C.; Cui, H.; Song, C. The contribution of transposable elements to size variations between four teleost genomes. *Mob. DNA* **2016**, *7*, 4.
69. Guo, X.M.; Zhang, Q.Q.; Sun, Y.W.; Jiang, X.Y.; Zou, S.M. *Tc1*-like Transposase *Thm3* of silver carp (*Hypophthalmichthys molitrix*) can mediate gene transposition in the genome of blunt snout bream (*Megalobrama amblycephala*). *G3* **2015**, *5*, 2601–2610.
70. Micale, L.; Loviglio, M.N.; Manzoni, M.; Fusco, C.; Augello, B.; Migliavacca, E.; Cotugno, G.; Monti, E.; Borsani, G.; Reymond, A.; et al. A fish-specific transposable element shapes the repertoire of *p53* target genes in zebrafish. *PLoS ONE* **2012**, *7*, e46642.
71. Pereira, C.S.A.; Pazian, M.F.; Ráb, P.; Collares-Pereira, M.J. Dynamics of *Rex3* in the genomes of endangered Iberian Leuciscinae (Teleostei, Cyprinidae) and their natural hybrids. *Mol. Cytogenet.* **2015**, *8*, 81.
72. Ahn, S.J.; Kim, M.S.; Jang, J.H.; Lim, S.U.; Lee, H.H. MMTS, a new subfamily of *Tc1*-like transposons. *Mol. Cells.* **2008**, *26*, 387–395.
73. Ivics, Z.; Izsvak, Z.; Minter, A.; Hackett, P.B. Identification of functional domains and evolution of *Tc1*-like transposable elements. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5008–5013.
74. Liu, D.; You, C.; Liu, S.; Liu, L.; Duan, W.; Chen, S.; Yan, J.; Liu, Y. Characterization of a novel *Tc1*-like transposon from bream (Cyprinidae, Megalobrama) and its genetic variation in the polyploidy progeny of bream-red crucian carp crosses. *J. Mol. Evol.* **2009**, *69*, 395–403.
75. Ishiyama, S.; Yamazaki, K.; Kurihara, F.; Yamashita, D.; Sao, K.; Hattori, A.; Koga, A. DNA-based transposable elements with nucleotide sequence similar to *Tol2* from medaka fish are prevalent in cyprinid fishes. *Gene Rep.* **2017**, *9*, 37–45.
76. Xu, P.; Xu, J.; Liu, G.; Chen, L.; Zhou, Z.; Peng, W.; Jiang, Y.; Zhao, Z.; Jia, Z.; Sun, Y.; et al. The allotetraploid origin and asymmetrical genome evolution of the common carp *Cyprinus carpio*. *Nat. Commun.* **2019**, *10*, 4625.
77. Dulz, T.A.; Lorscheider, C.A.; Nascimento, V.D.; Noleto, R.B.; Moreira-Filho, O.; Nogaroto, V.; Vicari, M.R. Comparative cytogenetics among *Leporinus friderici* and *Leporellus vittatus* populations (Characiformes, Anostomidae): Focus on repetitive DNA elements. *Comp. Cytogenet.* **2019**, *13*, 105–120.
78. Barbosa, P.; de Oliveira, L.A.; Pucci, M.B.; Santos, M.E.; Moreira-Filho, O.; Vicari, M.R.; Nogaroto, V.; de Almeida, M.C.; Ferreira Artoni, R. Identification and chromosome mapping of repetitive elements in the *Astyanax scabripinnis* (Teleostei: Characidae) species complex. *Genetica* **2015**, *143*, 55–62.
79. Daniel, S.N.; Penitente, M.; Silva, D.M.Z.A.; Hashimoto, D.T.; Ferreira, D.C.; Foresti, F.; Porto-Foresti, F. Organization and chromosomal distribution of histone genes and transposable *Rex* elements in the genome of *Astyanax bockmanni* (Teleostei, Characiformes). *Cytogenet. Genome Res.* **2015**, *146*, 311–318.
80. Gao, B.; Chen, W.; Shen, D.; Wang, S.; Chen, C.; Zhang, L.; Wang, W.; Wang, X.; Song, C. Characterization of autonomous families of Tc1/*Mariner* transposon in neoteleost genomes. *Mar. Genom.* **2017**, *34*, 67–77.
81. Martins, C.; Ferreira, I.A.; Oliveira, C.; Foresti, F.; Galetti, P.M., Jr. A tandemly repetitive centromeric DNA sequence of the fish *Hoplias malabaricus* (Characiformes: Erythrinidae) is derived from 5S rDNA. *Genetica* **2006**, *127*, 133–141.
82. Silva, D.M.Z.A.; Pansonato-Alves, J.C.; Utsunomia, R.; Daniel, S.N.; Hashimoto, D.T.; Oliveira, C.; Porto-Foresti, F.; Foresti, F. Chromosomal organization of repetitive DNA sequences in *Astyanax bockmanni* (Teleostei, Characiformes): Dispersive location, association and co-localization in the genome. *Genetica* **2013**, *141*, 329–336.

83. Pessenda, G.; Piscor, D.; da Silva, E.L.; Centofante, L.; Parise-Maltempi, P.P. Chromosomal organization of retrotransposon *Rex1* in *Astyanax* species (Characiformes, Characidae). *Caryologia* **2018**, *71*, 80–84.
84. Ferreira, V.A.M.; Marajo, L.; Matoso, D.A.; Ribeiro, L.B.; Feldberg, E. Chromosomal mapping of *Rex* retrotransposons in Tambaqui (*Colossoma macropomum* Cuvier, 1818) exposed to three climate change scenarios. *Cytogenet. Genome Res.* **2019**, *159*, 39–47.
85. Pucci, M.B.; Nogaroto, V.; Moreira-Filho, O.; Vicari, M.R. Dispersion of transposable elements and multigene families: Microstructural variation in Characidium (Characiformes: Crenuchidae) genomes. *Genet. Mol. Biol.* **2018**, *41*, 585–592.
86. Schemberger, M.O.; Nascimento, V.D.; Coan, R.; Ramos, E.; Nogaroto, V.; Ziemniczak, K.; Valente, G.T.; Moreira-Filho, O.; Martins, C.; Vicari, M.R. DNA transposon invasion and microsatellite accumulation guide W chromosome differentiation in a Neotropical fish genome. *Chromosoma* **2019**, *128*, 547–560.
87. Schemberger, M.O.; Oliveira, J.I.; Nogaroto, V.; Almeida, M.C.; Artoni, R.F.; Cestari, M.M.; Moreira-Filho, O.; Vicari, M.R. Construction and characterization of a repetitive DNA library in Parodontidae (Atinopterygii:Characiformes): A genomic and evolutionary approach to the degeneration of the W sex chromosome. *Zebrafish* **2014**, *11*, 518–527.
88. Schemberger, M.O.; Nogaroto, V.; Almeida, M.C.; Ferreira Artoni, R.; Valente, G.T.; Martins, C.; Moreira-Filho, O.; Cestari, M.M.; Vicari, M.R. Sequence analyses and chromosomal distribution of the *Tc1/Mariner* element in Paraodontidae fish (Teleostei: Characiforme). *Gene* **2016**, *593*, 308–314.
89. Terencio, M.L.; Schneider, C.H.; Gross, M.C.; Vicari, M.R.; Farias, I.P.; Passos, K.B.; Feldberg, E. Evolutionary dynamics of repetitive DNA in Semaprochilodus (Characiformes, Prochilodontidae): A fish model for sex chromosome differentiation. *Sex Dev.* **2013**, *7*, 325–333.
90. Yano, C.; Bertollo, L.A.C.; Ezaz, T.; Trifonov, V.; Sember, A.; Liehr, T.; Cioffi, M.B. Highly conserved Z and molecularly diverged W chromosomes in the fish genus *Triportheus* (Characiformes, Triportheidae). *Heredity* **2017**, *118*, 276–283.
91. Yano, C.F.; Merlo, M.A.; Portela-Bens, S.; Cioffi, M.B.; Bertollo, L.A.C.; Santos-Júnior, C.D.; Rebordinos, L. Evolutionary dynamics of multigene families in *Triportheus* (Characiformes, Triportheidae): A transposon mediated mechanism? *Front. Mar. Sci.* **2020**, *7*, 6.
92. de Sene, V.F.; Pansonato-Alves, J.C.; Ferreira, D.C.; Utsumonia, R.; Oliveira, C.; Foresti, F. Mapping of the retrotransposable elements *Rex1* and *Rex3* in chromosomes of eigenmannia (Teleostei, Gymnotigformes, Sternopygidae). *Cytogenet. Genome Res.* **2015**, *146*, 319–324.
93. da Silva, M.; Matoso, D.A.; Vicari, M.R.; de Almeida, M.C.; Margarido, V.P.; Artoni, R.F. Physical Mapping of 5S rDNA in two species of knifefishes: *Gymnotus pantanal* and *Gymnotus paraguensis* (Gymnotiformes). *Cytogenet. Genome Res.* **2011**, *134*, 303–307.
94. da Silva, M.; Barbosa, P.; Artoni, R.F.; Feldberg, E. Evolutionary dynamics of 5S rDNA and recurrent association of transposable elements in electric fish of the family Gymnotidae (Gymnotiformes): The case of *Gymnotus mamiraua*. *Cytogenet. Genome Res.* **2016**, *149*, 297–303.
95. Ferreira, D.C.; Oliveira, C.; Foresti, F. Chromosome mapping of retrotransposable elements *Rex1* and *Rex3* in three fish species in the subfamily Hypoptopomatinae (Teleostei, Siluriformes, Loricariidae). *Cytogenet. Genome Res.* **2011**, *132*, 64–70.
96. Gonzalez-Gouveia, J.; Wolf, I.R.; Vilas-Boas, L.A.; Heslop-Harrison, J.S.; Schwarzacher, T.; Dias, A.L. Repetitive DNA in the catfish genome: rDNA, microsatellites, and *Tc1-Mariner* transposon sequences in *Imparfinis* species (Siluriformes, Heptapteridae). *J. Hered.* **2017**, *108*, 650–657.
97. Liu, Z.; Li, P.; Kucuktas, H.; Dunham, R. Characterization of nonautonomous *Tc1*-like transposable elements of channel catfish (*Ictalurus punctatus*). *Fish Physiol. Biochem.* **1999**, *21*, 65–72.
98. Primo, C.C.; Glugoski, L.; Vicari, M.R.; Nogaroto, V. Chromosome mapping and molecular characterization of the *Tc1/Mariner* element in Rineloricaria (Siluriformes: Loricariidae). *Braz. Arch. Biol. Technol.* **2018**, *61*, e18170623.
99. Traldi, J.B.; Lui, R.L.; de Martinez, J.F.; Vicari, M.R.; Nogaroto, V.; Filho, O.M.; Blanco, D. R Chromosomal distribution of the retroelements *Rex1*, *Rex3* and *Rex6* in species of the genus *Harttia* and *Hypostomus* (Siluriformes: Loricariidae). *Neotrop. Ichthyol.* **2019**, *17*, e190010.
100. da Silva, K.R.; Mariotto, S.; Centofante, L.; Parise-Maltempi, P.P. Chromosome mapping of a *Tc1*-like transposon in species of the catfish *Ancistrus*. *Comp. Cytogenet.* **2017**, *11*, 65–79.

101. Hughes, L.C.; Ortí, G.; Huang, Y.; Sun, Y.; Baldwin, C.C.; Thompson, A.W.; Arcila, D.; Betancur-R, R.; Li, C.; Becker, L.; et al. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6249–6254.
102. Ybazeta, G. Molecular systematics and biogeography of the Galaxiidae. Ph.D. Thesis, Univeristy of Toronto, Toronto, ON, Canada, 2012.
103. Pan, Q.; Feron, R.; Yano, A.; Guyomard, R.; Jouanno, E.; Vigouroux, E.; Wen, M.; Busnel, J.M.; Bobe, J.; Concordet, J.P.; et al. Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLoS Genet.* **2019**, *15*, e1008013.
104. de Boer, J.G.; Yazawa, R.; Davidson, W.S.; Koop, B.F. Bursts and horizontal evolution of DNA transposon in the speciation of pseudotetraploid salmonids. *BMC Genom.* **2007**, *8*, 422.
105. Faber-Hammond, J.J.; Phillips, R.B.; Brown, K.H. Comparative analysis of the shared sex-determination region (SDR) among salmonid fishes. *Genome Biol. Evol.* **2015**, *7*, 1972–1987.
106. Lien, S.; Koop, B.K.; Sandve, S.R.; Miller, J.R.; Kent, M.P.; Nome, T.; Hvidsten, T.R.; Leong, J.S.; Minkley, D.R.; Zimin, A.; et al. The Atlantic salmon genome provides insights into rediploidization. *Nature* **2016**, *533*, 200–205.
107. Sävilammi, T.; Primmer, C.R.; Varadharajan, S.; Guyomard, R.; Guiguen, Y.; Sandve, S.R.; Vøllestad, L.A.; Papakostas, S.; Lien, S. The chromosome-level genome assembly of european grayling reveals aspects of a unique genome evolution process within salmonids. *G3* **2019**, *9*, 1283–1294.
108. Liu, K.; Xu, D.; Li, J.; Bian, C.; Duan, J.; Zhou, Y.; Zhang, M.; You, X.; You, Y.; Chen, J.; et al. Whole genome sequencing of Chinese clearhead icefish, *Protosalanx hyalocranius*. *GigaScience* **2017**, *6*, giw012.
109. Malmstrøm, M.; Matschiner, M.; Tørresen, O.K.; Jakobsen, K.S.; Jentoft, S. Whole genome sequencing data and *de novo* draft assemblies for 66 teleost species. *Sci. Data* **2017**, *4*, 160132.
110. Ota, K.; Tateno, Y.; Gobojori, T. Highly differentiated and conserved sex chromosome in fish species (*Aulopus japonicus*: Teleostei, Aulopidae). *Gene* **2003**, *317*, 187–193.
111. Brynildsen, W. Transposable Elements in Teleost Fish: In Silico Exploration of TE Activity, Diversity and Abundance across 74 Teleost Fish Genomes. Master's Thesis, University Oslo, Oslo, Norway, 2016. Available online: http://urn.nb.no/URN:NBN:no-55565 (accessed on 4 May 2020).
112. Tørresen, O.K.; Star, B.; Jentoft, S.; Reinar, W.B.; Grove, H.; Miller, J.R.; Walenz, B.P.; Knight, J.; Ekholm, J.M.; Peluso, P.; et al. An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genom.* **2017**, *18*, 95.
113. Lima, L.C.B.; de Martinez, P.A.; Molina, W.F. Cytogenetic characterization of three Balistoidea fish species from the Atlantic with inferences on chromosomal evolution in the families Monacanthidae and Balistidae. *Comp. Cytogenet.* **2011**, *5*, 61–69.
114. Nirchio, M.; Rossi, A.R.; Foresti, F.; Oliveira, C. Chromosome evolution in fishes: A new challenging proposal from Neotropical species. *Neotrop. Ichthyol.* **2014**, *12*, 761–770.
115. Pérez Bielsa, N. Is the Small Genome Size of Pleuronectiformes Related to Small Intron Size? Master's Thesis, Universitat de Girona, Girona, Spain, 2013.
116. Tarallo, A. Genome Composition Plasticity in Marine Organisms. Ph.D. Thesis, University of Naples "Federico II", Naples, Italy, 2016.
117. Dettleff, P.; Hormazabal, E.; Aedo, J.; Fuentes, M.; Meneses, C.; Molina, A.; Valdes, J.A. Identification and evaluation of long noncoding RNAs in response to handling stress in red cusk-eel (*Genypterus chilensis*) via RNA-seq. *Mar. Biotechnol.* **2020**, *22*, 94–108.
118. Small, C.M.; Bassham, S.; Catchen, J.; Amores, A.; Fuiten, A.M.; Brown, R.S.; Jones, A.G.; Cresko, W.A. The genome of the Gulf pipefish enables understanding of evolutionary innovations. *Genome Biol.* **2016**, *17*, 258.
119. Cioffi, M.; Bertollo, L.A.; Villa, M.A.; de Oliveira, E.A.; Tanomtong, A.; Yano, C.F.; Supiwong, W.; Chaveerach, A. Genomic organization of repetitive DNA elements and its implications for the chromosomal evolution of channid fishes (Actinopterygii, Perciformes). *PLoS ONE* **2015**, *10*, e0130199.
120. García, E.; Cross, I.; Portela-Bens, S.; Rodríguez, M.E.; García-Angulo, A.; Molina, B.; Cuadrado, A.; Liehr, T.; Rebordinos, L. Integrative genetic map of repetitive DNA in the sole *Solea senegalensis* genome shows a *Rex* transposon located in a proto-sex chromosome. *Sci. Rep.* **2019**, *9*, 17146.

121. Roberts, B.H.; Morrongiello, J.R.; King, A.J.; Morgan, D.L.; Saunders, T.M.; Woodhead, J.; Crook, D.A. Migration to freshwater increases growth rates in a facultatively catadromous tropical fish. *Oecologia* **2019**, *191*, 253–260.
122. Rodríguez, M.E.; Molina, B.; Merlo, M.A.; Arias-Pérez, A.; Portela-Bens, S.; García-Angulo, A.; Cross, I.; Liehr, T.; Rebordinos, L. Evolution of the proto sex-chromosome in *Solea senegalensis*. *Int. J. Mol. Sci.* **2019**, *20*, 5111.
123. Bohne, A.; Zhou, Q.; Darras, A.; Schmidt, C.; Schartl, M.; Galiana-Arnoux, D.; Volff, J.N. Zisupton—A novel superfamily of DNA transposable elements recently active in fish. *Mol. Biol. Evol.* **2011**, *29*, 631–645.
124. Koga, A.; Shimada, A.; Shima, A.; Sakaizumi, M.; Tachida, H.; Hori, H. Evidence for recent invasion of the medaka fish genome by the *Tol2* transposable element. *Genetics* **2000**, *155*, 273–281.
125. Koga, A.; Hori, H. The *Tol2* transposable element of the medaka fish: An active DNA-based element naturally occurring in a vertebrate genome. *Genes Genet. Syst.* **2001**, *76*, 1–8.
126. Koga, A.; Sakaizumi, M.; Hori, H. Transposable elements in medaka fish. *Zool. Sci.* **2002**, *19*, 1–6.
127. Dennenmoser, S.; Sedlazeck, F.J.; Iwaszkiewicz, E.; Li, X.-Y.; Altmuller, J.; Nolte, A.W. Copy number increases of transposable elements and protein-coding genes in an invasive fish of hybrid origin. *Mol. Ecol.* **2017**, *26*, 4712–4724.
128. Xiao, Y.; Xiao, Z.; Ma, D.; Liu, J.; Li, J. Genome sequence of the barred knifejaw *Oplegnathus fasciatus* (Temminck & Schlegel, 1844): The first chromosome-level draft genome in the family Oplegnathidae. *GigaScience* **2019**, *8*, giz013.
129. Costa, G.W.W.F.; Cioffi, M.B.; Bertollo, L.A.C.; Molina, W.F. Transposable elements in fish chromosomes: A study in the marine cobia species. *Cytogenet. Genome Res.* **2013**, *141*, 126–132.
130. Almeida, L.A.H.; Nunes, L.A.; Bitencourt, J.A.; Molina, W.F.; Affonso, P.R.A.M. Chromosomal evolution and cytotaxonomy in wrasses (Perciformes; Labridae). *J. Hered.* **2017**, *108*, 239–253.
131. Jacobina, U.P.; Cioffi, M.B.; Souza, L.G.R.; Calado, L.L.; Tavares, M.; Manzella, J., Jr.; Bertollo, L.A.C.; Molina, W.F. Chromosome mapping of repetitive sequences in *Rachycentron canadum* (Perciformes: Rachycentridae): Implications for karyotypic evolution and perspectives for biotechnological uses. *J. Biomed. Biotechnol.* **2011**, *2011*, 218231.
132. Gross, M.C.; Schneider, C.H.; Valente, G.T.; Porto, J.I.R.; Martins, C.; Feldeberg, E. Comparative cytogenetic analysis of the genus *Symphysodon* (Discus Fishes, Cichlidae): Chromosomal characteristics of retrotransposons and minor ribosomal DNA. *Cytogenet. Genome Res.* **2010**, *127*, 43–53.
133. de Freitas Mourão, A.A.; Daniel, S.N.; Hashimoto, D.T.; Ferreira, C.D.; Porto-Foresti, F. Organization and distribution of repetitive DNA classes in the *Cichla kelberi* and *Cichla piquiti* Genome. *Cytologia* **2017**, *82*, 193–197.
134. Fischer, C.; Bouneau, L.; Coutanceau, J.P.; Weissenbach, J.; Volff, J.N.; Ozouf-Costaz, C. Global heterochromatic colocalization of transposable elements with minisatellites in the compact genome of the pufferfish *Tetraodon nigroviridis*. *Gene* **2004**, *336*, 175–183.
135. Santos, F.A.; Marques, D.F.; Terencio, M.L.; Feldberg, E.; Rodrigues, L.R. Cytogenetic variation of repetitive DNA elements in *Hoplias malabaricus* (Characiformes-Erythrinidae) from white, black and clear water rivers of the Amazon basin. *Genet. Mol. Biol.* **2016**, *39*, 40–48.
136. Schneider, C.H.; Gross, M.C.; Terencio, M.L.; do Carmo, E.J.; Martins, C.; Feldberg, E. Evolutionary dynamics of retrotransposable elements *Rex1*, *Rex3* and *Rex6* in neotropical cichlid genomes. *BMC Evol. Biol.* **2013**, *13*, 152.
137. Capriglione, T.; De Paolo, S.; Cocca, E. Helinoto, a Helitron2 transposon from the icefish *Chionodraco hamatus*, contains a region with three deubiquitinase-like domains that exhibit transcriptional activity. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2014**, *11*, 49–58.
138. Capriglione, T.; Odierna, G.; Caputo, V.; Canapa, A.; Olmo, E. Characterization of a *Tc1*- like transposon in the Antarctic ice-fish, *Chionodraco hamatus*. *Gene* **2002**, *295*, 193–198.
139. Cocca, E.; De Iorio, S.; Capriglione, T. Identification of a novel helitron transposon in the genome of Antarctic fish. *Mol. Phylogenet. Evol.* **2011**, *58*, 439–446.
140. Hendry, T.A.; Freed, L.L.; Fader, D.; Fenolio, D.; Sutton, T.T.; Lopex, J.V. Ongoing transposon-mediated genome reduction in the luminous bacterial symbionts of deep-sea ceratioid anglerfishes. *Am. Soc. Microbiol. mBio* **2018**, *9*. [CrossRef]

141. Xu, D.; Molina, W.F.; Yano, C.F.; Zhang, Y.; de Oliveira, E.A.; Lou, B.; Cioffi, M.B. Comparative cytogenetics in three Sciaenid species (Teleostei, Perciformes): Evidence of interspecific chromosomal diversification. *Mol. Cytogenet.* **2017**, *10*, 37.
142. Dasilva, C.; Hadji, H.; Ozouf-Costaz, C.; Nicaud, S.; Jaillon, O.; Weissenbach, J.; Crollius, H.R. Remarkable compartmentalization of transposable elements and pseudogenes in the heterochromatin of the *Tetraodon nigroviridis* genome. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 13636–13641.
143. Guo, M.; Wang, S.; Su, Y.; Zhou, Y.; Liu, M.; Wang, J. Molecular cytogenetic analyses of *Epinephelus bruneus* and *Epinephelus moara* (Perciformes, Epinephelidae). *PeerJ* **2014**, *2*, e412.
144. Zong, W.; Gao, B.; Diaby, M.; Shen, D.; Wang, S.; Wang, Y.; Sang, Y.; Chen, C.; Wang, X.; Song, C. *Traveler*, a new DD35E family of *Tc1/Mariner* transposons, invaded vertebrates very recently. *Genome Biol. Evol.* **2020**, *12*, 66–76.
145. Costa, G.W.; Cioffi, M.B.; Bertollo, L.A.; Molina, W.F. The evolutionary dynamics of ribosomal genes, histone H3, and transposable *Rex* elements in the genome of Atlantic snappers. *J. Hered.* **2016**, *107*, 173–180.
146. Dalziel, A.C.; Tirbhowan, S.; Drapeau, H.F.; Power, C.; Jonah, L.S.; Gbotsyo, Y.A.; Dion-Côté, A.M. Using asexual vertebrates to study genome evolution and animal physiology: Banded (*Fundulus diaphanus*) x common killifish (*F. heteroclitus*) hybrid lineages as a model system. *Evol. Appl.* **2020**, *13*, 1214–1239.
147. Collares-Pereira, M.J.; Matos, I.; Morgado-Santos, M.; Coelho, M.M. Natural pathways towards polyploidy in animals: The *Squalius alburnoides* fish complex as a model system to study genome size and genome reorganization in polyploids. *Cytogenet. Genome Res.* **2013**, *140*, 97–116.
148. Auvinet, J.; Graça, P.; Belkadi, L.; Petit, L.; Bonnivard, E.; Dettai, A.; Detrich, W.H.; Ozouf-Costaz, C.; Higuet, D. Mobilization of retrotransposons as a cause of chromosomal diversification and rapid speciation: The case for the Antarctic teleost genus *Trematomus*. *BMC Genom.* **2018**, *19*, 339.
149. Cioffi, M.B.; Martins, C.; Bertollo, L.A.C. Chromosome spreading of associated transposable elements and ribosomal DNA in the fish *Erythrinus erythrinus*. Implications for genome change and karyoevolution in fish. *BMC Evol. Biol.* **2010**, *10*, 271.
150. Motta-Neto, C.C.; Cioffi, M.D.; Costa, G.W.; Amorim, K.D.; Bertollo, L.A.; Artoni, R.F.; Molina, W.F. Overview on karyotype stasis in atlantic grunts (Eupercaria, Haemulidae) and the evolutionary extensions for other marine fish groups. *Front. Mar. Sci.* **2019**, *6*, 628.
151. Auvinet, J.; Graça, P.; Ghigliotti, L.; Pisano, E.; Dettaï, A.; Ozouf-Costaz, C.; Higuet, D. Insertion hot spots of DIRS1 retrotransposon and chromosomal diversifications among the Antarctic teleosts Nototheniidae. *Int. J. Mol. Sci.* **2019**, *20*, 701.
152. Volff, J.N.; Bouneau, L.; Ozouf-Costaz, C.; Fischer, C. Diversity of retrotransposable elements in compact pufferfish genomes. *Trends Genet.* **2003**, *19*, 12.
153. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982.
154. Sun, C.; Shepard, D.B.; Chong, R.A.; López Arriaza, J.; Hall, K.; Castoe, T.A.; Feschotte, C.; Pollock, D.D.; Mueller, R.L. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol. Evol.* **2012**, *40*, 168–183.
155. Makałowski, W.; Gotea, V.; Pande, A.; Makałowska, I. Transposable elements: Classification, identification, and their use as a tool for comparative genomics. In *Evolutionary Genomics: STATISTICAL and Computational Methods, Methods in Molecular Biology*; Anisimova, M., Ed.; Springer: New York, NY, USA, 2019; Volume 1910. [CrossRef]
156. Kazazian, H.H., Jr. Mobile elements: Drivers of genome evolution. *Science* **2004**, *303*, 1626–1632.
157. Kramerov, D.A.; Vassetzky, N.S. Origin and evolution of SINEs in eukaryotic genomes. *Heredity* **2011**, *107*, 487–495.
158. Evgen'ev, M.B.; Arkhipova, I.R. *Penelope*-like elements–A new class of retroelements: Distribution, function and possible evolutionary significance. *Cytogenet. Genome Res.* **2005**, *110*, 510–521.
159. Arkhipova, I.R.; Yushenova, I.A.; Rodriguez, F. Endonuclease-containing *Penelope* retrotransposons in the bdelloid rotifer *Adineta vaga* exhibit unusual structural features and play a role in expansion of host gene families. *Mob. DNA* **2013**, *4*, 19.

160. Bourque, G.; Burns, K.H.; Gehring, M.; Goburnovam, V.; Scluanov, A.; Hammell, M.; Imbeault, M.; Izsvák, Z.; Levin, H.L.; Macfarlan, T.S.; et al. Ten things you should know about transposable elements. *Genome Biol.* **2018**, *19*, 199.
161. Goerner-Potvin, P.; Bourque, G. Computational tools to unmask transposable elements. *Nat. Rev. Genet.* **2018**, *19*, 688–704.
162. Grabundzija, I.; Messing, S.A.; Thomas, J.; Cosby, R.L.; Bilic, I.; Miskey, C.; Gogol-Döring, A.; Kapitonov, V.; Diem, T.; Dalda, A.; et al. *Helitron* transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nat. Commun.* **2016**, *7*, 10716.
163. Kapitonov, V.V.; Jurka, J. Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4540–4545.
164. Feschotte, C.; Swamy, L.; Wessler, S.R. Genome-wide analysis of *Mariner*-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). *Genetics* **2003**, *163*, 747–758.
165. Shao, F.; Han, M.; Peng, Z. Evolution and diversity of transposable elements in fish genomes. *Sci. Rep.* **2019**, *9*, 15399.
166. Du, K.; Stöck, M.; Kneitz, S.; Klopp, C.; Woltering, J.M.; Adolfi, M.C.; Feron, R.; Prokopov, D.; Makunin, A.; Kichigin, I.; et al. The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat. Ecol. Evol.* **2020**, *4*, 841–852.
167. Dion-Cote, A.M.; Renaut, S.; Normandeau, E.; Bernatchez, L. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. *Mol. Biol. Evol.* **2014**, *31*, 1188–1199.
168. Oliver, K.R.; Greene, W.K. Mobile DNA and the TE-Thrust hypothesis: Supporting evidence from the primates. *Mob. DNA* **2011**, *2*, 8.
169. Grahn, R.A.; Rinehart, T.A.; Cantrell, M.A.; Wichman, H.A. Extinction of *LINE-1* activity coincident with a major mammalian radiation in rodents. *Cytogenet. Genome Res.* **2005**, *110*, 407–415.
170. Platt, R.N.; Vandewege, M.W.; Kern, C.; Schmidt, C.J.; Hoffmann, F.G.; Ray, D.A. Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats. *Mol. Biol. Evol.* **2014**, *31*, 1536–1545.
171. Burke, W.D.; Muller, F.; Eickbush, T.H. R4, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes. *Nucleic Acid Res.* **1995**, *23*, 4628–4634.
172. Yang, J.; Malik, H.S.; Eickbush, T.H. Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 7847–7852.
173. Yano, C.F.; Poltronieri, J.; Bertollo, L.A.C.; Ferreira Artoni, R.; Liehr, T.; Cioffi, M.B. Chromosomal mapping of repetitive DNAs in *Triportheus trifurcatus* (Characidae, Characiformes): Insights into the differentiation of the Z and W chromosomes. *PLoS ONE* **2014**, *9*, e90946.
174. Pansonato-Alves, J.C.; Serrano, E.A.; Utsumomia, R.; Scacchetti, P.C.; Oliveira, C.; Foresti, F. Mapping five repetitive DNA classes in sympatric species of *Hypostomus* (Teleostei: Siluriformes: Loricariidae): Analysis of chromosomal variability. *Rev. Fish. Biol. Fisher.* **2013**, *4*, 447–489.
175. Splendore de Borba, R.; Lourenço da Silva, E.; Parise-Maltempi, P.P. Chromosome mapping of retrotransposable elements *Rex1* and *Rex3* in *Leporinus Spix*, 1829 species (Characiformes: Anostomidae) and its relationships among heterochromatic segments and W sex chromosome. *Mob. Genet. Elem.* **2013**, *3*, e27460.
176. Fantinatti, B.E.A.; Mazzuchelli, J.; Valente, G.T.; Cabral-de-Mello, D.C.; Martins, C. Genomic content and new insights on the origin of the B chromosome of the cichlid fish *Astatotilapia latifasciata*. *Genetica* **2011**, *139*, 1273–1282.
177. Valente, G.T.; Mazzuchelli, J.; Ferreira, I.A.; Poletto, A.B. Cytogenetic mapping of the retroelements *Rex1*, *Rex3* and *Rex6* among cichlid fish: New insights on the chromosomal distribution of transposable elements. *Cytogenet. Genome Res.* **2011**, *133*, 34–42.
178. Mazzuchelli, J.; Martins, C. Genomic organization of repetitive DNAs in the cichlid fish *Astronotus ocellatus*. *Genetica* **2009**, *136*, 461–469.

179. Teixeira, W.G.; Ferreira, I.A.; Cabral-de-Mello, D.C.; Mazzuchelli, J.; Valente, G.T.; Pinhal, D.; Poletto, A.B.; Venere, P.C.; Martins, C. Organization of repeated DNA elements in the genome of the cichlid fish *Cichla kelberi* and its contributions to the knowledge of fish genomes. *Cytogenet. Genome Res.* **2009**, *125*, 224–234.
180. Griffin, D.K.; Harvey, S.C.; Campos-Ramos, R.; Ayling, L.J.; Bromage, N.R.; Masabanda, J.S.; Penman, D.J. Early origins of the X and Y chromosomes: Lessons from tilapia. *Cytogenet. Genome Res.* **2002**, *99*, 157–163.
181. Ozouf-Costaz, C.; Brandt, J.; Korting, C.; Pisano, E.; Bonillo, C.; Coutanceau, J.P.; Volff, J.N. Genome dynamics and chromosomal localization of the non-LTR retrotransposons *Rex1* and *Rex3* in Antarctic fish. *Antarct. Sci.* **2004**, *16*, 51–57.
182. Supiwong, W.; Liehr, T.; Cioffi, M.B.; Chaveerach, A.; Kosyakova, N.; Pinthong, K.; Tanee, T.; Tanomtong, A. Chromosomal evolution in naked catfishes (Bagridae, Siluriformes): A comparative chromosome mapping study. *Zool. Anz.* **2014**, *253*, 316–320.
183. Favarato, R.M.; Ribeiro, L.B.; Feldberg, E.; Matoso, D.A. Chromosomal mapping of transposable elements of the *Rex* family in the bristlenose catfish, ancistrus (*Siluriformes, Loricariidae*), from the amazonian region. *J. Hered.* **2017**, *108*, 254–261.
184. Ferreira, D.C.; Porto-Foresti, F.; Oliveira, C.; Foresti, F. Transposable elements as a potential source for understanding the fish genome. *Mob. Genet. Elements* **2011**, *1*, 112–117.
185. Bouneau, L.; Fischer, C.; Ozouf-Costaz, C.; Froschauer, A.; Jaillon, O.; Coutanceau, J.P.; Korting, C.; Weissenbach, J.; Bernot, A.; Volff, J.N. An active non-LTR retrotransposon with tandem structure in the compact genome of pufferfish *Tetraodon nigroviridis*. *Genome Res.* **2003**, *13*, 1686–1695.
186. Carducci, F.; Barucca, M.; Canapa, A.; Biscotti, M.A. *Rex* retroelements and teleost genomes: An overview. *Int. J. Mol. Sci.* **2018**, *19*, 3653.
187. Malik, H.S.; Henikoff, S.; Eickbush, T.H. Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.* **2000**, *10*, 1307–1318.
188. Feschotte, C.; Gilbert, C. Endogenous viruses: Insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* **2012**, *13*, 283–296.
189. Hayward, A.; Cornwallis, C.K.; Jern, P. Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 464–469.
190. Xu, X.; Zhao, H.; Gong, Z.; Han, G.-Z. Endogenous retroviruses of non-avian/mammalian vertebrates illuminate diversity and deep history of retroviruses. *Plos Pathog.* **2018**, *14*, e1007072.
191. Naville, M.; Volff, J.N. Endogenous retroviruses in fish genomes: From relics of past infections to evolutionary innovations? *Front. Microbiol.* **2016**, *7*, 1197.
192. Basta, H.A.; Cleveland, S.B.; Clinton, R.A.; Dimitrov, A.G.; McClure, M.A. Evolution of teleost fish retroviruses: Characterization of new retroviruses with cellular genes. *J. Virol.* **2009**, *83*, 10152–10162.
193. Shen, C.H.; Steiner, L.A. Genome structure and thymic expression of an endogenous retrovirus in zebrafish. *J. Virol.* **2004**, *78*, 899–911.
194. Llorens, C.; Muñoz-Pomer, A.; Bernad, L.; Botella, H.; Moya, A. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biol. Direct.* **2009**, *4*, 41.
195. Schartl, M.; Walter, R.B.; Shen, Y.; Garcia, T.; Catchen, J.; Amores, A.; Braasch, I.; Chalopin, D.; Volff, J.N.; Lesch, K.P.; et al. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat. Genet.* **2013**, *45*, 567–572.
196. Naville, M.; Warren, I.A.; Haftek-Terreau, Z.; Chalopin, D.; Brunet, F.; Levin, P.; Galiana, D.; Volff, J.N. Not so bad after all: Retroviruses and long terminal repeat retrotransposons as a source of new genes in vertebrates. *Clin. Microbiol. Infect.* **2016**, *22*, 312–323.
197. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **2008**, *9*, 397–405.
198. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **2016**, *351*, 1083–1087.
199. Dechaud, C.; Volff, J.; Schartl, M.; Naville, M. Sex and the TEs: Transposable elements in sexual development and function in animals. *Mob. DNA* **2019**, *10*, 42.
200. Ezaz, T.; Deakin, J.E. Repetitive sequence and sex chromosome evolution in vertebrates. *Adv. Evol. Biol.* **2014**, *2014*, 104683.

201. Herpin, A.; Braasch, I.; Kraeussling, M.; Schmidt, C.; Thoma, E.C.; Nakamura, S.; Tanaka, M.; Schartl., M. Transcriptional rewiring of the sex determining *dmrt1* gene duplicate by transposable elements. *PLoS Genet.* **2010**, *6*, e1000844.
202. Schartl, M.; Schories, S.; Wakamatsu, Y.; Nagao, Y.; Hashimoto, H.; Bertin, C.; Mourot, B.; Schmidt, C.; Wilhelm, D.; Centanin, L.; et al. *Sox5* is involved in germ-cell regulation and sex determination in medaka following co-option of nested transposable elements. *BMC Biol.* **2018**, *16*, 16.
203. Tomaszkiewicz, M.; Chalopin, D.; Schartl, M.; Galiana, D.; Volff, J.N. A multicopy Y chromosomal *SGNH* hydrolase gene expressed in the testis of the platyfish has been captured and mobilized by a *Helitron* transposon. *BMC Genet.* **2014**, *15*, 44.
204. Lubieniecki, K.P.; Lin, S.; Cabana, E.I.; Li, J.; Lai, Y.Y.Y.; Davidson, W.S. Genomic instability of the sex-determining locus in Atlantic salmon (*Salmo salar*). *G3* **2015**, *5*, 2513–2522.
205. Sotero-Caio, C.G.; Platt, R.N., II; Suh, A.; Ray, D.A. Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* **2017**, *9*, 161–177.
206. Casacuberta, E.; González, J. The impact of transposable elements in environmental adaptation. *Mol. Ecol.* **2013**, *6*, 1503–1517.
207. Carducci, F.; Biscotti, M.A.; Forconi, M.; Barucca, M.; Canapa, A. An intriguing relationship between teleost *Rex3* retroelement and environmental temperature. *Biol. Lett.* **2019**, *15*, 20190279.
208. Schmidt, A.L.; Anderson, L.M. Repetitive DNA elements as mediators of genomic change in response to environmental cues. *Biol. Rev. Camb. Philos. Soc.* **2006**, *81*, 531–543.
209. Fablet, M.; Vieira, C. Evolvability, epigenetics and transposable elements. *Biomol. Concepts* **2011**, *2*, 331–341.
210. Canapa, A.; Biscotti, M.A.; Barucca, M.; Carducci, F.; Carotti, E.; Olmo, E. Shedding light upon the complex net of genome size, genome composition and environment in chordates. *Eur. Zool. J.* **2020**, *87*, 192–202.
211. Biscotti, M.A.; Barucca, M.; Canapa, A. New insights into the genome repetitive fraction of the Antarctic bivalve *Adamussium colbecki*. *PLoS ONE* **2018**, *13*, e0194502.
212. Chu, C.G.; Tan, C.T.; Yu, G.T.; Zhong, S.; Xu, S.S.; Yan, L. A novel retrotransposon inserted in the dominant *Vrn-B1* allele confers spring growth habit in tetraploid wheat (*Triticum turgidum* L.). *G3* **2011**, *1*, 637–645.
213. Fujino, K.; Hashida, S.N.; Ogawa, T.; Natsume, T.; Uchiyama, T.; Mikami, T.; Kishima, Y. Temperature controls nuclear import of *Tam3* transposase in *Antirrhinum*. *Plant J.* **2011**, *65*, 146–155.
214. Makarevitch, I.; Waters, A.J.; West, P.T.; Stitzer, M.; Hirsch, C.N.; Ross-Ibarra, J.; Springer, N.M. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* **2015**, *11*, e1004915.

3.3 Investigation of the activity of transposable elements and genes involved in their silencing in the newt *Cynops orientalis*, a species with a giant genome

scientific reports

Check for updates

OPEN

Investigation of the activity of transposable elements and genes involved in their silencing in the newt *Cynops orientalis*, a species with a giant genome

Federica Carducci[1], Elisa Carotti[1], Marco Gerdol[2], Samuele Greco[2], Adriana Canapa[1], Marco Barucca[1]✉ & Maria Assunta Biscotti[1]

Caudata is an order of amphibians with great variation in genome size, which can reach enormous dimensions in salamanders. In this work, we analysed the activity of transposable elements (TEs) in the transcriptomes obtained from female and male gonads of the Chinese fire-bellied newt, *Cynops orientalis*, a species with a genome about 12-fold larger than the human genome. We also compared these data with genomes of two basal sarcopterygians, coelacanth and lungfish. In the newt our findings highlighted a major impact of non-LTR retroelements and a greater total TE activity compared to the lungfish *Protopterus annectens*, an organism also characterized by a giant genome. This difference in TE activity might be due to the presence of young copies in newt in agreement also with the increase in the genome size, an event that occurred independently and later than lungfish. Moreover, the activity of 33 target genes encoding proteins involved in the TE host silencing mechanisms, such as *Ago/Piwi* and *NuRD* complex, was evaluated and compared between the three species analysed. These data revealed high transcriptional levels of the target genes in both newt and lungfish and confirmed the activity of NuRD complex genes in adults. Finally, phylogenetic analyses performed on *PRDM9* and *TRIM28* allowed increasing knowledge about the evolution of these two key genes of the NuRD complex silencing mechanism in vertebrates. Our results confirmed that the gigantism of the newt genomes may be attributed to the activity and accumulation of TEs.

Genome size varies considerably across eukaryotes and it is correlated neither with the number of genes, nor with the morpho-functional complexity of a species. Besides having a significant impact on the number and the size of introns[1,2], as well as on the placement of regulatory regions, genome size is profoundly influenced by the relative abundance and activity of transposable elements (TEs)[3–5]. These genetic elements constitute a large fraction of the repetitive genomic DNA and are able to move throughout genomes through mechanisms of transposition based on a DNA intermediate molecule, in the case of DNA transposons, and on an RNA intermediate molecule, in the case of retrotransposons (LTR and non-LTR retroelements).

Adaptive and non-adaptive evolutionary hypotheses have been proposed to explain the surprising genome size variation observed in eukaryotes[6–8]. However, a scientific consensus about this issue is still far from being reached, as the genome size observed in extant metazoan species is the result of the combination between complex events that may have occurred ancestrally or relatively recently. These include auto- or allo-polyploidization events[9,10], large-scale structural variations that may include deletions and insertions, expansions of tandem repeat arrays, and changes in the relative abundance of TE families[11]. Moreover, the amount of nuclear DNA does not

[1]Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy. [2]Dipartimento di Scienze della Vita, Università degli Studi di Trieste, Via L. Giorgieri, 5, 34127 Trieste, Italy. ✉email: m.barucca@univpm.it

only affect the size of both cells and nuclei in a given species, but also determines the duration of cell cycle and influences metabolic rates and the speed of development[12–14].

The amphibian order Caudata is characterized by a great variation in genome size between metamorphic and neotenic species, which show the smallest and the largest genomes, respectively. As of February 2021, the 170 urodele species listed in the Animal Genome Size database display a genome size ranging from 13.89 to 120.60 Gb[15]. The highest values belong to salamanders, which together with lungfish are the record holders for the largest genomes among extant vertebrates.

Even though several factors, such as metabolic rate and cell cycle duration, have been associated with amphibian genome size[8,16,17], the molecular mechanisms favouring the persistence of extreme repetitive DNA contents in salamander genomes remain to be addressed.

Evidence suggest that the gigantism of urodele genome is not attributable to polyploidy but rather to a reduced DNA loss rate and to the accumulation of TEs, in particular of LTR retrotransposons[18–20]. Among phylogenetically different crown salamanders, differences in genome composition have been linked to a balance between TE proliferation and host silencing mechanisms. The increase in genome size could be due to a reduced level of TE suppression, which can be likely traced back to the latest common ancestor of crown salamanders and was inherited by all the organisms belonging to this lineage. Moreover, an ongoing proliferation of *Ty3/Gypsy* LTR retroelements has been reported for three of the five species analysed[20]. In addition, the large-scale analysis of the Mexican axolotl genome has evidenced a recent burst of LTR retroelement expansion also for Ambystomatidae[1].

Although the transposition of TEs can be source of genetic diversity and regulatory innovations, such events are also known to have deleterious effects, which may lead to a decrease of host fitness. To counteract the steady threat posed by TE transposition and novel insertions, diverse lineages have evolved a number of different protective mechanisms. The Ago proteins interact with small interfering RNAs (siRNAs) and microRNAs (miRNAs) to degrade the mRNAs, to repress translation, and to form heterochromatin[21]. These small RNAs are produced through the activity of DROSHA and DICER proteins. The former acts in the nucleus, loading double-stranded RNA substrates, helped by the DiGeorge Syndrome Critical Region Gene 8 (DGCR8), and modifying them to produce miRNA precursors. DICER cleaves both these molecules and endogenous or exogenous siRNAs in the cytoplasm. In germline cells, PIWI proteins are guided by small non-coding RNAs, called piRNAs (PIWI-interacting RNAs), to transcriptionally and post-transcriptionally silence transposons through base complementarity. piRNAs are extremely heterogeneous since they derive from transposons inserted in piRNA clusters, genomic regions transcribed as long single-stranded precursor RNA molecules, later processed into mature piRNAs. The transcription of these genomic regions seems to be activated by the deposition of histone 3 lysine 9 trimethylation (H3K9me3) by the protein SETDB1[22]. Subsequently the piRNA precursor transcripts are shuttled by the HMG protein Maelstrom from the nucleus to the cytoplasm, where they are cleaved by the PLD6 nuclease to form the primary piRNAs and by the HMG protein Maelstrom to form primary and secondary piRNAs[23,24].

Additional strategies can be used to repress mobile elements. Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs), the largest family of transcriptional regulators in higher vertebrates, are notably involved in the silencing of TEs, including endogenous retroviruses, LINE and SINE retroelements[25]. KRAB-ZFPs contain the KRAB domain and bind DNA through a C-terminal array of zinc finger motifs, and the Tripartite Motif Containing 28 (TRIM28) protein through their N-terminal KRAB domain. TRIM28, known also as KRAB-associated protein 1 (KAP1), serves as a scaffold for binding the heterochromatin protein 1 (HP1) and the DNA Methyltransferases (DNMT1 and DNMT3A), which allow the deposition of transcriptional repressive marks. Moreover, TRIM28 triggers the formation of heterochromatin by also recruiting the histone methyltransferase SETDB1, which adds the H3K9me3 mark, and the nucleosome remodelling and deacetylase (NuRD) complex[26,27]. This multiprotein chromatin remodelling complex contains the histone deacetylases 1 and 2 (HDAC1 and HDAC2), the chromatin helicase DNA binding protein 4 (CHD4), the Retinoblastoma-binding protein 4 and 7 (Rbbp4 and Rbbp7), either the zinc-finger proteins GATA Zinc Finger Domain Containing 2A or 2B (GATAD2a and GATAD2b), two Metastasis-associated Proteins (MTA1, MTA2, and/or MTA3), and the Methyl-CpG Binding Domain Protein 2 or 3 (MBD2 or MBD3)[28–30]. Although this mechanism was initially presumed to only act during early development, an increasing number of papers suggests that it may be also functional in adult tissues[31–34].

The host TE silencing mechanisms evolve to counteract the rapidly mutating TEs and invasion of new TE families in the germline cells according to the "arm race" model. In mammals, the expansion of KRAB-ZFPs is positively correlated with the number of endogenous LTR elements[35]. The lack of silencing mechanisms efficiency might lead to an increase in genome size due to TE proliferation. In the lungfish giant genome, Meyer and colleagues (2021)[36] have hypothesized that transposon silencing machinery did not adapt to reduce the TE expansion. In the strawberry poison frog, it has been proposed that its large genome might be the result of TE proliferation not suppressed by Piwi proteins in female gonad[37].

Moreover, in the genomes expanded because of TEs DNA methylation seems to be essential for the long-term accommodation of these mobile elements in the host genome[38].

In light of these observations, we analysed the activity of transposable elements and 33 genes encoding proteins involved in the TE host silencing mechanisms in the transcriptomes of the Chinese fire-bellied newt *Cynops orientalis* (David, 1873), a species with a genome about 12-fold larger than the human genome. Moreover, comparisons were performed with transcriptomic data obtained from tissues of the coelacanth *Latimeria menadoensis*[39] and the lungfish *Protopterus annectens*[40], two species characterized by significant differences in the activity of TEs and in genome size[2,41]. The data obtained showed a higher TE activity in the newt, with a major impact of non-LTR retroelements. The transcriptional activity of genes involved in TE silencing mechanisms highlighted that, although these mechanisms are active in adults, differences in TE activity evidenced between the species compared in this study may be due to the presence of old and inactive copies. In addition, to increase knowledge about the TE silencing mechanisms in amphibians, phylogenetic analyses were performed on *PRDM9* and *TRIM28*, two key genes of the NuRD complex. The former gene (also named *Meisetz*) is considered the

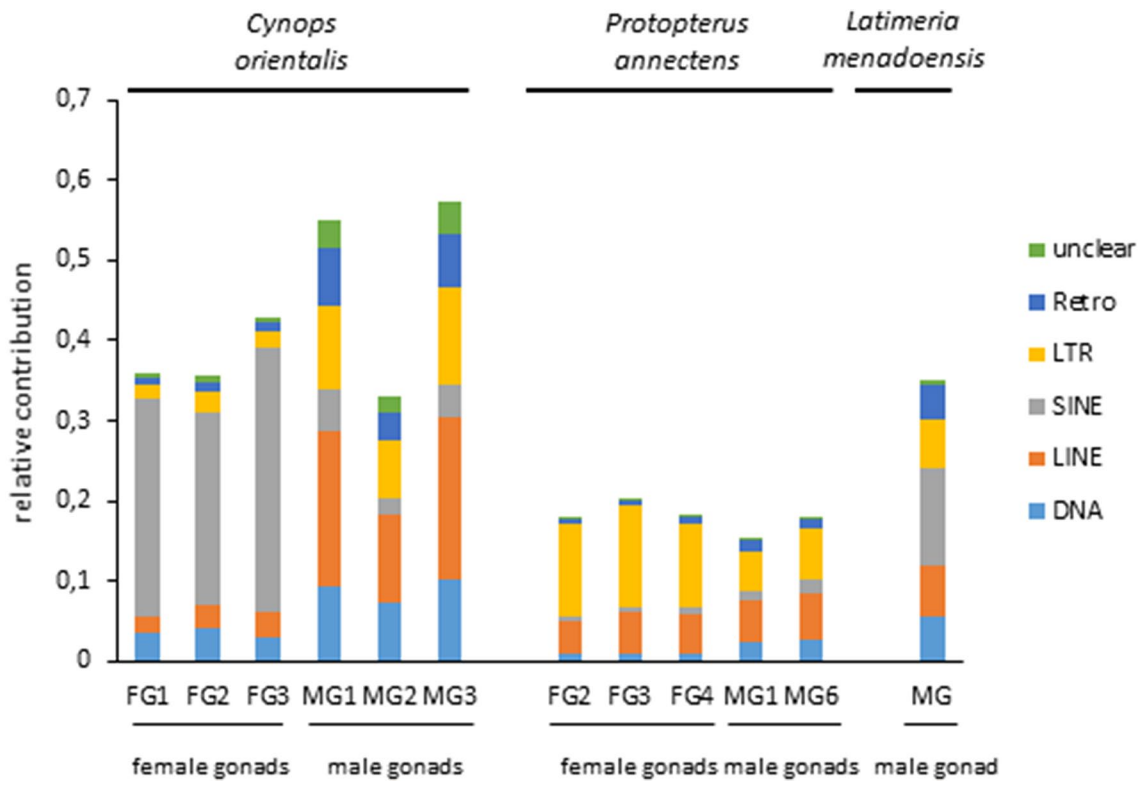

Figure 1. Relative contribution of transposon transcriptional activity in the transcriptomes obtained from gonadal tissues of *Cynops orientalis, Protopterus annectens,* and *Latimeria menadoensis*. *FG* female gonad, *MG* male gonad.

ancestor of the KRAB domain[42] and its absence in amphibians, crocodiles, and birds led to support its lost in these lineages[43]. The latter gene belongs to the *TRIM* family, highly diversified in vertebrates[44], and acts as platform in the recruitment of NuRD complex proteins. Our results revealed the presence of *PRDM9* in Gymnophiona, the sister group of Anura and Caudata, and the presence of *TRIM28* in the three amphibian clades (Gymnophiona, Urodela, Anura).

Results

Transposable element activity in gonadal tissues in *C. orientalis*. The transcriptional activity of TEs was investigated in the ovary and testis transcriptomes of three female and three male specimens of *C. orientalis* (Fig. 1). In the females, the analysis showed a higher relative contribution of SINE retroelements, in contrast with males, where the strongest impact was due to LINE retroelements. However, the transcriptional contribution of other TE types was not negligible in testes. The comparison between TE activity in the gonadal tissues of *P. annectens*, *L. menadoensis*, and *C. orientalis* revealed that the lowest activity was found in lungfish (Fig. 1). Moreover, the highest activity of LTR retroelements was observed in *P. annectens*, while SINE retroelement activity prevailed in *L. menadoensis* (for statistical support see Supplementary Table S1).

Identification and transcriptional activity of genes involved in TE silencing mechanisms. Four transcripts (*AGO1*, *AGO2*, *AGO3*, and *AGO4*) related to the *Ago* subfamily and three transcripts (*PIWIL1*, *PIWIL2*, and *PIWIL4*) related to the *Piwi* subfamily were retrieved in the female and male gonadal transcriptomes of *C. orientalis* (Supplementary Table S2). All transcripts included a complete CDS with the exception of *AGO1,* which was truncated at the 5' and 3' ends. As in vertebrates other than eutherian mammals, no sequence ascribable to the *PIWIL3* gene was identified in the newt transcriptomes. The evaluation of the expression levels of *Ago* genes in female gonads of *C. orientalis* (Fig. 2a) showed a very high activity of *AGO4*, followed by *AGO2*. The expression of the other two *Ago* genes was lower. Although *AGO4* gene was expressed in newt testis, we evidenced an expression level much lower than in female gonads (for statistical support see Supplementary Table S1). The comparison with the results previously obtained in lungfish and coelacanth revealed a similar pattern of expression for *AGO2* and *AGO4* in the female gonads of newt and lungfish, with lower expression levels in *P. annectens* (Supplementary Fig. S1).

Concerning the *Piwi* genes, investigation of their expression levels revealed a high level of expression of *PIWIL1* in both ovary and testis transcriptomes (Fig. 2b). The transcriptional activity of *PIWIL1* in *C. orientalis* female gonads was comparable to that observed in lungfish ovaries (Supplementary Fig. S2). We also investigated the presence of expressed transcripts orthologous with genes involved in miRNA and siRNA biogenesis pathways in the transcriptomes of *C. orientalis*. Three transcripts containing the complete CDS of *DICER*, *DROSHA*, and

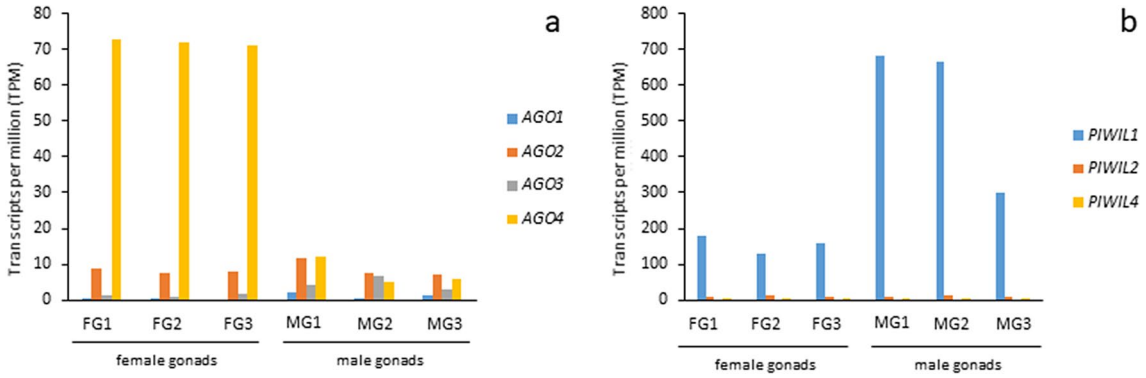

Figure 2. Expression levels of *AGO* genes. **(a)** Expression values of *Ago* genes detected in *Cynops orientalis* transcriptomes obtained from the female and male gonadal tissues. **(b)** Expression values of *Piwi* genes detected in *C. orientalis* transcriptomes obtained from the female and male gonadal tissues. *FG* female gonad, *MG* male gonad.

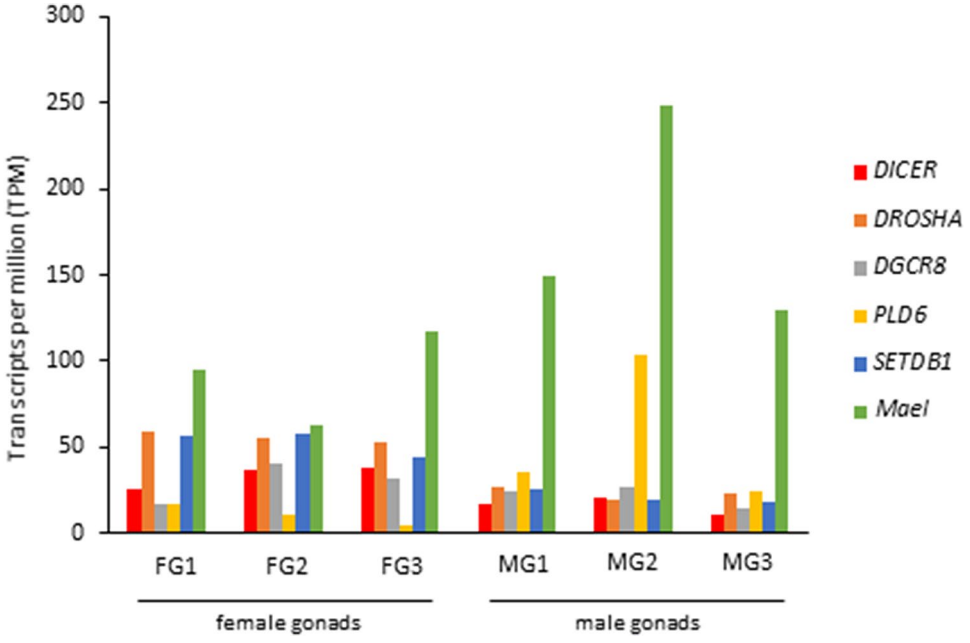

Figure 3. Expression levels of genes involved in small RNA biogenesis. Expression values of genes encoding proteins involved in small RNA production detected in *Cynops orientalis* transcriptomes obtained from the female and male gonadal tissues. *FG* female gonad, *MG* male gonad.

DGCR8 (Supplementary Table S2) were found to be expressed in all the samples analysed (Fig. 3). Similarly, three transcripts containing a complete CDS and involved in piRNA production were also expressed in all tissues, with *Mael* in particular showing high expression values in both ovary and testis (Fig. 3). Although all the six genes involved in small RNA biogenesis were actively transcribed in the fire-bellied newt, their expression levels were significantly lower than those observed in lungfish and coelacanth (Supplementary Fig. S3). Moreover, transcripts containing complete CDS were identified for *TRIM28*, the three *HP1* (*HP1a*, *HP1b*, and *HP1g*) and the two *DNMT* genes (*DNMT1* and *DNMT3A*) as well as for 12 genes encoding proteins of NuRD complex (Supplementary Table S2). The analysis of the transcriptional activity showed a high level of expression of *HP1g* and between the two methyltransferases analysed the *DNMT1* was more active in gonads (Fig. 4). Overall, the genes encoding proteins that are part of the NuRD complex were expressed at considerably high levels, with the lone exception of *GATAD2b* (Fig. 4). Moreover, these genes were also expressed in *P. annectens* and in *L. menadoensis*, but at lower levels than those observed in the newt (Supplementary Fig. S4). The transcriptional activity of these genes in the liver of the three species was found to be lower than in the gonadal tissues (Supplementary Fig. S4).

Phylogenetic analyses and evolutionary history of *PRDM9* and *TRIM28*. No sequence orthologous to *PRDM9* could be identified in the available transcriptomes. However, investigation of the genome of

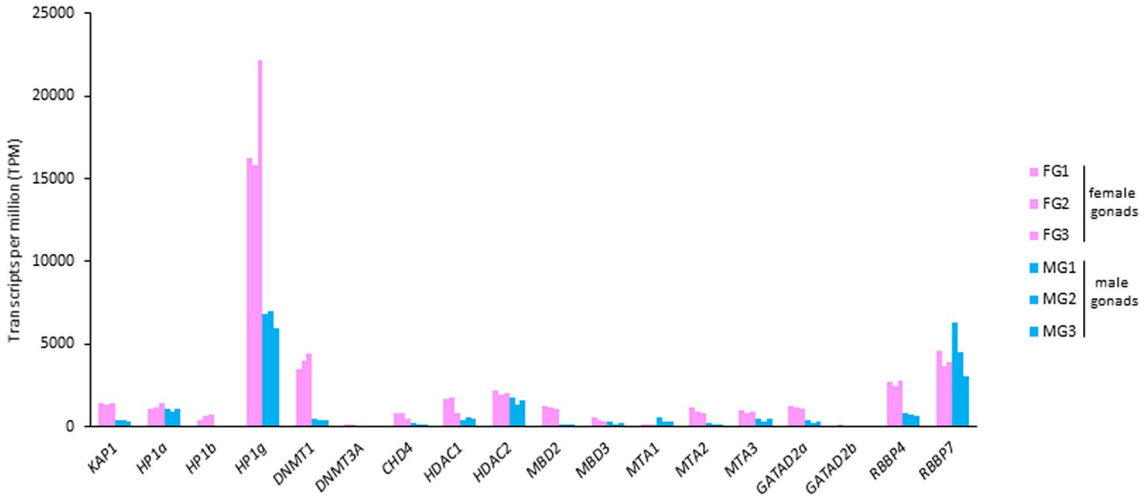

Figure 4. Expression levels of genes involved in heterochromatin formation and genes encoding proteins of the NuRD complex. Expression values of 18 genes investigated in *Cynops orientalis* transcriptomes obtained from the female and male gonadal tissues. *FG* female gonad, *MG* male gonad.

the caecilian *Microcaecilia unicolor* revealed a *bona fide PRDM9* sequence. The orthology assessment of this sequence was confirmed by phylogenetic analysis (Fig. 5A), which also demonstrated the loss of this gene in Anura and Caudata (Fig. 5B). The protein domains of PRDM9 and PRDM7 sequences considered in the phylogenetic analysis were investigated, highlighting the presence of a KRAB A box in basal sarcopterygians and in *M. unicolor* (Supplementary Fig. S5). PRDM7 sequences were present only in mammals.

Phylogenetic analysis also confirmed the presence of *TRIM28* in *C. orientalis* (Fig. 6A), as well as in *Xenopus tropicalis* and *M. unicolor*, which supports the presence of this gene in all three amphibian lineages. Moreover, a TRIM28 sequence was also identified in the elephant shark *Callorhinchus milii* (Fig. 6B).

Discussion

Salamanders and lungfish display the largest genomes among vertebrates[15]. This characteristic was specifically acquired along the Caudata lineage, since the ancestor of modern Lissamphibia (salamanders and frogs) likely had a nuclear DNA content of 5–10 pg[45,46]. In extant salamanders, genome expansion was attributed to the presence of long introns, slow rates of DNA loss, and high density of TEs, in particular of LTR retroelements[1,18–20]. In axolotl, where repetitive sequences account for 65.6% of the genome assembly size, the estimated relative age of the LTR retroelements permitted identification of a long period of activity followed by a recent burst of expansion[1]. The study of the amphibian genome size variation is even more interesting if we consider that organisms with small genomes, similar to those of birds, are present in this lineage. The recent genome sequencing of the ornate burrowing frog *Platyplectrum ornatum* has shown that reduced genome size is correlated with shorter intron lengths, low amount of TEs, and an increased expression of Piwi pathway genes[47].

The evolutionary processes underlying the accumulation and persistence of high levels of repetitive sequences in these organisms are not completely understood.

TE activity in gonadal transcriptomes of *C. orientalis*. The relative contribution of TE activity in the ovary and testis transcriptomes of three *C. orientalis* individuals of each sex suggested a major impact of non-LTR retroelements (Fig. 1). Nevertheless, we need to point out that, in absence of a complete genome sequence for this species, we cannot establish whether this observation simply reflects a higher transcriptional activity or if it indicates a parallel genome-wide expansion of this TE class. If the high amount of LTR retroelements found in common among salamander genomes is also shared by the fire-bellied newt, this finding might suggest either a lower activity of LTR elements in gonads or the presence of old/degenerate copies. Moreover, in female gonads, SINE strongly prevailed as the most active non-LTR retroelements. Although LINE retroelements were the most active elements in male gonads, all TE types showed a considerable transcriptional contribution. TEs are well-known potential regulators of gene networks that may influence the expression of neighbouring genes by hitchhiking promoters, transcription factor binding sites, insulators, splicing sites or epigenetic modifications[21]. The different pattern of TE activity observed in gonads could be due to their involvement in the regulatory toolkit of ovaries and testes in *C. orientalis*, which would be in line with the different transcriptomic expression profiles obtained in these tissues in our previous work[44]. Moreover, inter-sex variations in the activity of TEs in gonads have been also previously described in other vertebrates[49,50].

The impact of TEs and mechanisms for transposition silencing in giant genomes. The low TE activity values observed in the lungfish *P. annectens*, compared with *C. orientalis*, is particularly intriguing since the two species share a comparable genome size. A possible explanation for this discrepancy might be sought in the different expression of genes involved in TE silencing mechanisms in the two species. It is noteworthy that

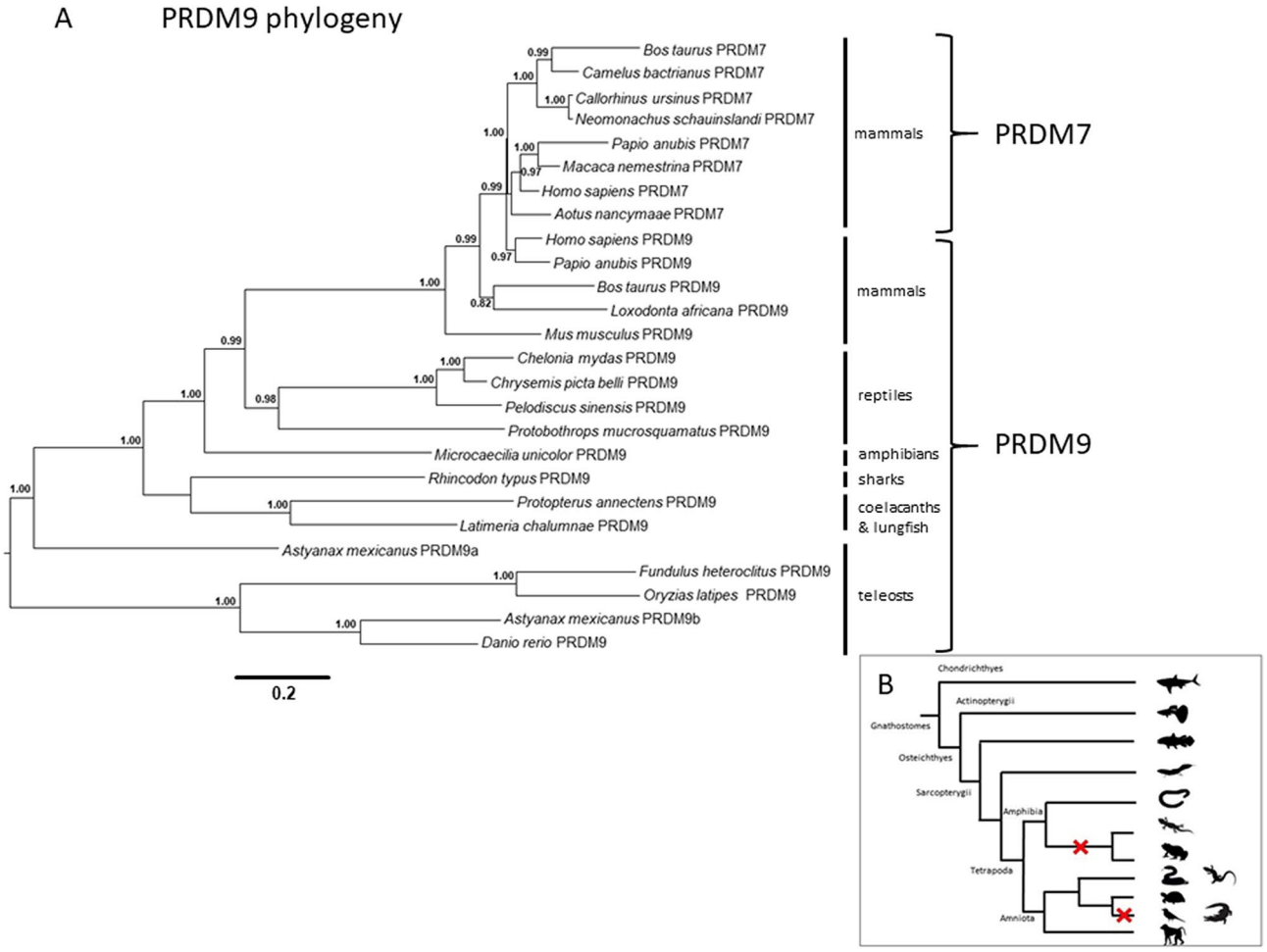

Figure 5. PRDM9 phylogeny. **(A)** Phylogenetic analysis of PRDM9 performed through Bayesian inference. The two curly brackets group PRDM7 and PRDM9 sequences. Vertical bars indicate to which animal group the sequences analysed belong to. **(B)** Schematic representation of *PRDM9* evolutionary history in gnathostomes. Red cross indicates gene loss.

the *Ago* and *Piwi* gene activity was found to be significantly higher in the newt. On the other hand, the genes involved in heterochromatin formation and those encoding proteins of the NuRD were actively transcribed in both species. In light of these observations, the lower TE activity in *P. annectens* might be consistent with the hypothesis that its genome is mainly made up of nonactive mobile elements[23,40,51,52]. Indeed, the genomic expansion in lungfish lineage occurred independently[36,45] and earlier than that of salamander lineage. In lungfish, the increasing in genome size can be dated in the Carboniferous (360–320 Mya) while in amphibians the expansion started in the Permian (280 Mya) up to the Middle Jurassic (208 Mya)[46,53,54]. The recent genome sequencing of the West African lungfish has provided further support to the hypothesis of an ancient burst of transposition followed by a long period of degeneration creating a "cemetery of TEs"[2,55]. Moreover, the high number of KRAB-ZFPs identified in the genome of the African lungfish has also suggested that this species evolved a higher ability to repress TEs[2]. Taking into account the genome size and the high transcriptional levels of the genes here investigated, we hypothesize that the TE repression system may have also been enhanced in *Cynops*. By contrast, coelacanths did not evolve a strong TE repression toolkit due to the low number of TEs in their genome, about 12-fold lower than those of the fire bellied newt and lungfish. This idea is consistent with the expression levels of TEs and genes encoding proteins involved in heterochromatin formation and silencing in this species. Overall, the activity of TEs can increase due to environmental stressors such as temperature[22]. In particular, the beginning of Carboniferous and the transition from Permian to Triassic were characterized by severe environmental changes[56] that might have provoked a reduction of the efficiency of silencing mechanisms in gonads, leading to a mobilization of TEs and thus to an increase of genome size in lungfish and newts[17].

The transcriptional activity of *TRIM28*, *HP1*, *DNMTs*, and NuRD complex genes in adults supports the evidence that these silencing mechanisms are not only active in early developmental stages[31–34]. Moreover, the levels of expression of these genes were quite high in gonads compared to the somatic hepatic tissue. This finding might be correlated with the need to repress TE activity and to preserve genome integrity in gonads. The NuRD complex is involved in endogenous retrovirus silencing[33] and consequently the high expression levels of its genes in *C. orientalis* suggested a high prevalence of these elements in the newt genome. This hypothesis is

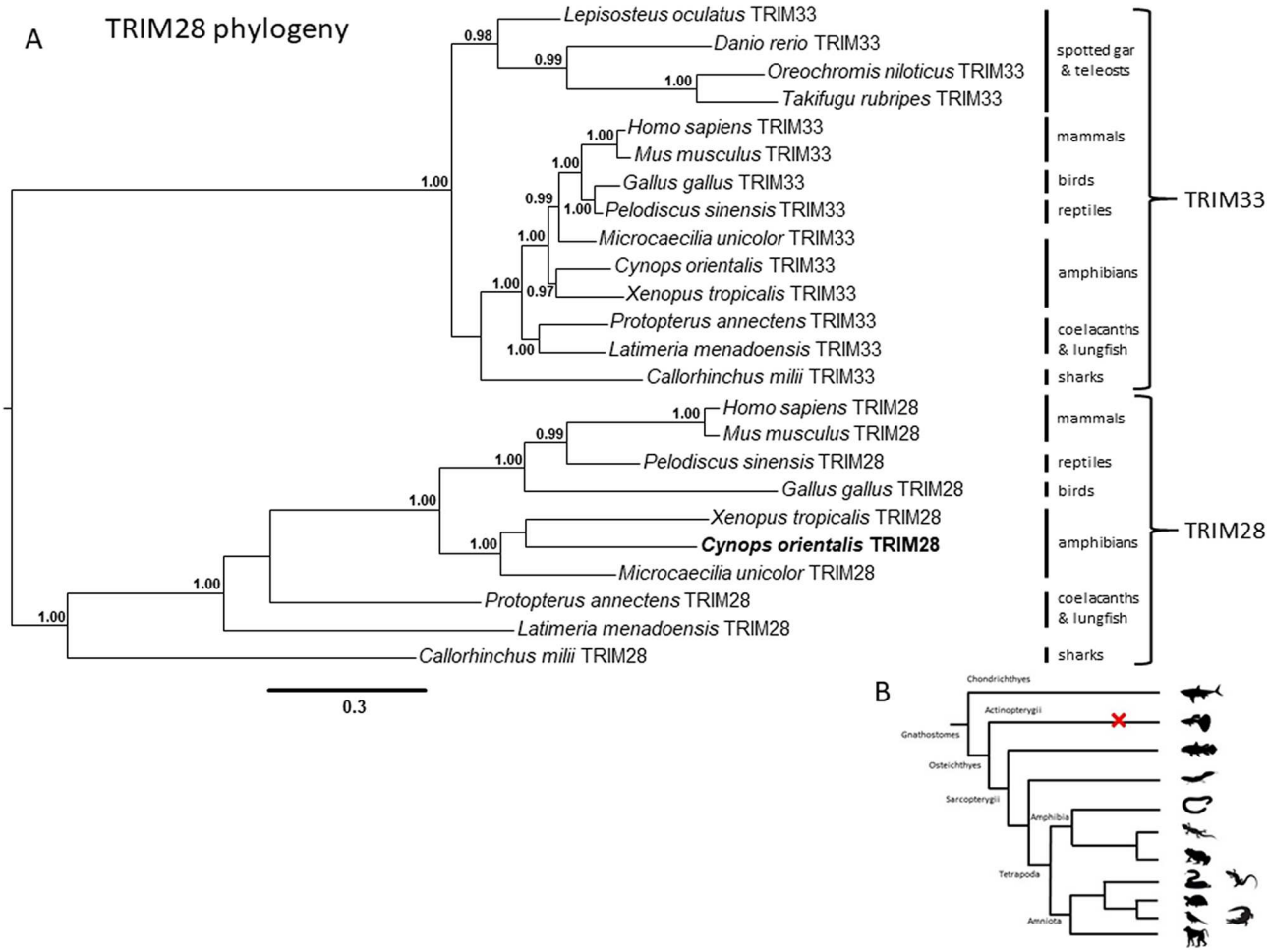

Figure 6. TRIM28 phylogeny. **(A)** Phylogenetic analysis of TRIM28 performed through Bayesian inference. The two curly brackets group TRIM33 and TRIM28 sequences. Vertical bars indicate to which animal group the sequences analysed belong to. **(B)** Schematic representation of *TRIM28* evolutionary history in gnathostomes. Red cross indicates gene loss. *Cynops orientalis* TRIM28 is in bold.

in line with the results that emerged from the complete genome sequencing of axolotl, which revealed that LTR retroelements and endogenous retroviruses were the most abundant classes of repetitive sequences[1]. Moreover, the high transcriptional activity of the *DNMT1* methyltransferase suggested this gene as the main candidate regulator of TE methylation in *C. orientalis*. This mechanism has been proposed as essential for the long-term accommodation of TEs in the host genome[38].

TE silencing mechanisms evolved in amphibians as in other tetrapods. Phylogenetic analyses were focused on *PRDM9* and *TRIM28*, two genes of interest to increase our knowledge about the evolution of the KRAB-ZFPs and NuRD complex, involved in TE silencing mechanisms in vertebrates. It is known that the KRAB domain contained in the KRAB-ZFPs derived from *PRDM9* gene[42]. This gene shows an interesting evolutionary history since it was lost in crocodiles, birds and, among amphibians, in salamanders and frogs[43]. Our analysis confirmed the absence of this gene in *C. orientalis*. However, we identified *PRDM9* in the caecilian *M. unicolor,* demonstrating that its loss in the amphibian lineage might have occurred in the common ancestor of frogs and salamanders, after the split from the Gymnophiona (caecilian) lineage. On the other hand, TRIM28 was found in all three amphibian lineages. Moreover, the identification of a *TRIM28* sequence in the elephant shark *C. milii* suggests that this gene was already present in the common ancestor of gnathostomes and consequently hints that the absence of this gene in Actinopterygii was due to a secondary loss. The presence of *PRDM9* and *TRIM28* genes suggested that the TE silencing mechanisms evolved in amphibians as in other tetrapods.

Conclusions

The findings here obtained highlighted a major impact of non-LTR retroelements and a greater total TE activity in the newt *C. orientalis* compared to the lungfish *Protopterus annectens*, another organism characterized by a giant genome. Therefore, the activity and accumulation of TEs likely played a key role in the genomic gigantism

in urodeles. In particular, the difference in TE activity between the newt *Cynops* and the lungfish *Protopterus* might be due to the presence of younger and active copies in newt.

The transcriptional activity of target genes encoding proteins involved in the TE host silencing mechanisms confirmed that the NuRD complex was active also in adults and that their expression was high in gonads to preserve genome integrity. Further comparative studies that may include species characterized by smaller genomes will provide new insights on the evolutionary "arms-race" between TEs proliferation and silencing mechanisms acting in host genome.

Finally, our phylogenetic analyses showed that *PRDM9* was present in the common ancestor of amphibians and *TRIM28* was present in the common ancestor of gnathostomes. Therefore, these findings allowed to increase knowledge about the evolution of these two key genes of NuRD complex silencing mechanism in vertebrates.

Material and methods

Identification of transposable elements and estimate of transcriptional activity. The identification of expressed transposable elements was carried out using the methodology described in a previous work[40]. Briefly, the de novo assembled transcriptome of *C. orientalis*[48] was first masked with RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) against the Dfam[57] database, and then scanned with RepeatScout[58] to identify novel active repeated elements. Following a false positive decontamination step aimed at removing protein-coding sequences not ascribable to TE-associated conserved domains, the elements were classified with TEClass[59] into the following categories: DNA transposons, LTR retroelements, LINEs, SINEs, retroelements (whenever a retroelement could not be discriminated with certainty as a LTR or non –LTR retroelements) or unclear. The likely collapse of nearly identical TEs belonging to the same family, as well as the frequent fragmentation of transcripts containing repeats, typical of transcriptome de novo assembly approaches[60], prevented the classification of the identified TEs on a finer scale.

The cumulative activity of each class of TEs was calculated in *C. orientalis*, *P. annectens*, and *L. menadoensis* available gonadal tissue samples, as their relative contribution to the total transcriptional effort, i.e. by calculating the fraction of reads mapped to the elements included in the repeat library, relative to the total number of reads mapped to the reference transcriptome assemblies in each sample. This task was carried out with the proprietary *map reads to reference* tool included in the CLC genomics Workbench v.20 (Qiagen, Hilden, Germany), using the following mapping parameters: length fraction = 0.75, similarity fraction = 0.98. TE counts for *C. orientalis* were listed in the Supplementary Table S3. The statistical significance between the relative abundance of any TE class between male and female gonads was evaluated with an unpaired t-test (Supplementary Table S1).

Characterization and expression of genes involved in TE silencing mechanisms. The de novo transcriptome assembly obtained from the ovary and testis collected from three *C. orientalis* specimens of each sex (FG1/FG2/FG3 and MG1/MG2/MG3, respectively) was used as a reference database for sequence homology searches[48] with tBLASTn[61]. In detail, the search involved transcripts orthologous to members of the *Ago* (*AGO1*, *AGO2*, *AGO3*, *AGO4*) and *Piwi* (*PIWIL1*, *PIWIL2*, *PIWL4*) subfamilies, as well as those involved in small RNA biogenesis (*DICER*, *DROSHA*, *DGCR8*, *PLD6*, *SETDB1*, *Mael*). In addition, the search was extended to *P. annectens*[40] and *L. menadoensis*[39] for *PRDM9*, *TRIM28*, *HP1a*, *HP1b*, *HP1g*, *DNMT1* and *DNMT3A*, and for the transcripts encoding proteins of the NuRD complex (*CHD4*, *HDAC1*, *HDAC2*, *MBD2*, *MBD3*, *MTA1*, *MTA2*, *MTA3*, *GATAD2a*, *GATAD2b*, *RBBP4*, *RBBP7*). All transcripts were translated using Sequence Translation (https://www.ebi.ac.uk/Tools/st/), allowing the identification of UTR and CDS regions (Supplementary Table S2). The sequences were deposited in GenBank under the accession numbers provided in the Supplementary Table S2. The presence of conserved protein domains in PRDM9 and its paralog PRDM7 was inferred with CD-Search[62].

The expression values of genes of interest are reported as transcripts per million (TPM) to allow comparisons both within and between tissues[63] (Supplementary Table S3). These values were further adjusted following the procedure described in Biscotti et al. 2016[48] to ensure full compatibility with data previously published for other species. Briefly, to take into account possible distortions of TPM calculations linked with the different structure of the transcriptomes, gene expression values were computed by using a scaling factor, based on the cumulative expression of 2111 broadly expressed verified single-copy orthologs shared by all the species taken into account. The expression levels of such genes were calculated with the CLC Genomics Workbench v.20, using the following parameters: length fraction = 0.75, similarity fraction = 0.98. Read counts obtained from the three biological replicates of male and female gonads, normalized by quantile, were used to perform a statistical analysis of differential gene expression, carried out with a Baggerlys' test[64]. The presence of the genes of interest among the subset of those differentially expressed between the gonads of the two sexes was ascertained based on a Bonferroni-corrected p-value < 0.05 (Supplementary Table S1).

Phylogenetic analyses. Phylogenetic analyses were performed to assess the orthology and the evolutionary history of genes of interest. Orthologous Similar sequences were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/) or ENSEMBL (http://www.ensembl.org/index.html) and used in the phylogenetic analyses (for accession numbers see Supplementary Table S4). Multiple alignments of the amino acid sequences were obtained with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) using default parameters. PRDM9 and TRIM28 phylogenetic analyses with midpoint rooting were performed by Bayesian inference using MrBayes (version 3.2)[65] and 1,000,000 generations were run (the burn-in was set to 2,500). The Jones amino acid model[66] was identified by the MrBayes program with a potential scale reduction factor (PSRF) of 1.00. Stationarity defined as the condition where the average standard deviation of split frequencies was of 0.0019 for PRDM9 phylogeny and of 0.0051 for TRIM28 phylogeny.

Received: 2 March 2021; Accepted: 5 July 2021
Published online: 20 July 2021

References

1. Nowoshilow, S. *et al.* The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**, 50–55 (2018).
2. Wang, K. *et al.* African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell* https://doi.org/10.1016/j.cell.2021.01.047 (2021).
3. Biscotti, M. A., Carducci, F., Olmo, E. & Canapa, A. Vertebrate genome size and the impact of transposable elements in the genome evolution. in *Evolution*, *Origin of Life, Concepts and Methods*. 233–251. (Springer, 2019).
4. Carducci, F., Biscotti, M. A., Barucca, M. & Canapa, A. Transposable elements in vertebrates: Species evolution and environmental adaptation. *Eur. Zool. J.* **86**, 497–503 (2019).
5. Chalopin, D., Naville, M., Plard, F., Galiana, D. & Volff, J. N. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* **7**, 567–580 (2015).
6. Olmo, E., Capriglione, T. & Odierna, G. Genome size evolution in vertebrates: Trends and constraints. *Comp. Biochem. Phys.* **92**, 447–453 (1989).
7. Kang, M. *et al.* Adaptive and nonadaptive genome size evolution in Karst endemic flora of China. *New. Phytol.* **202**, 1371–1381 (2014).
8. Waltari, E. & Edwards, S. V. Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. *Am. Nat.* **160**, 539–552 (2002).
9. Schreier, A. D., Gille, D., Mahardja, B. & May, B. Neutral markers confirm the octoploid origin and reveal spontaneous autopolyploidy in white sturgeon, *Acipenser transmontanus*. *J. Appl. Ichthyol.* **27**, 24–33 (2011).
10. Cunha, C. *et al.* The evolutionary history of the allopolyploid *Squalius alburnoides* (Cyprinidae) complex in the northern Iberian Peninsula. *Heredity* **106**, 100–112 (2011).
11. Canapa, A., Barucca, M., Biscotti, M. A., Forconi, M. & Olmo, E. Transposons, genome size, and evolutionary insights in animals. *Cytogenet. Genome Res.* **147**, 217–239 (2015).
12. Olmo, E. & Morescalchi, A. Genome and cell sizes in frogs: A comparison with salamanders. *Experientia* **34**, 44–46 (1978).
13. Gregory, T. R. Genome size and developmental complexity. *Genetica* **115**, 131–146 (2002).
14. Vinogradov, A. E. Nucleotypic effect in homeotherms: Body-mass independent resting metabolic rate of passerine birds is related to genome size. *Evolution* **51**, 220–225 (1997).
15. Gregory, T. R. *Animal Genome Size Database*. http://www.genomesize.com. (2020).
16. Vinogradov, A. E. & Anatskaya, O. V. Genome size and metabolic intensity in tetrapods: A tale of two lines. *Proc. Biol. Sci.* **273**, 27–32 (2006).
17. Canapa, A. *et al.* Shedding light upon the complex net of genome size, genome composition and environment in chordates. *Eur. Zool. J.* **87**, 192–202 (2021).
18. Sun, C. *et al.* LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol. Evol.* **4**, 168–183 (2012).
19. Sun, C., Arriaza, J. R. L. & Mueller, R. L. Slow DNA loss in the gigantic genomes of salamanders. *Genome Biol. Evol.* **4**, 1340–1348 (2012).
20. Sun, C. & Mueller, R. L. Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. *Genome Biol. Evol.* **6**, 1818–1829 (2014).
21. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**, 272–285 (2007).
22. Pappalardo, A. M., Ferrito, V., Biscotti, M. A., Canapa, A. & Capriglione, T. Transposable elements and stress in vertebrates: An overview. *Int. J. Mol. Sci.* **22**, 1970 (2021).
23. Biscotti, M. A. *et al.* The small non-coding RNA processing machinery of two living fossil species, lungfish and coelacanth, gives new insights into the evolution of the Argonaute protein family. *Genome Biol. Evol.* **9**, 438–453 (2017).
24. Parhad, S. S. & Theurkauf, W. E. Rapid evolution and conserved function of the piRNA pathway. *Open Biol.* **9**, 180181 (2019).
25. Turelli, P. *et al.* Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* **24**, 1260–1270 (2014).
26. Schultz, D. C., Friedman, J. R. & Rauscher, F. J. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2a subunit of NuRD. *Genes Dev.* **15**, 428–443 (2001).
27. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J. SETDB1: A novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919–932 (2002).
28. Allen, H. F., Wade, P. A. & Kutateladze, T. G. The NuRD architecture. *Cell. Mol. Life Sci.* **70**, 3513–3524 (2013).
29. Kloet, S. L. *et al.* Towards elucidating the stability, dynamics and architecture of the nucleosome remodelling and deacetylase complex by using quantitative interaction proteomics. *FEBS J.* **282**, 1774–1785 (2015).
30. Mohd-Sarip, A. *et al.* DOC1-dependent recruitment of NURD reveals antagonism with SWI/SNF during epithelial-mesenchymal transition in oral cancer cells. *Cell Rep.* **20**, 61–75 (2017).
31. Ecco, G. *et al.* Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. *Dev. Cell.* **36**, 611–623 (2016).
32. Ecco, G., Imbeault, M. & Trono, D. KRAB zinc finger proteins. *Development* **144**, 2719–2729 (2017).
33. Imbeault, M., Helleboid, P.-Y. & Trono, D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**, 550–554 (2017).
34. Grassi, D., Jönsson, M. E., Brattås, P. L. & Jakobsson, J. TRIM28 and the control of transposable elements in the brain. *Brain Res.* **1705**, 43–47 (2019).
35. Yang, P., Wang, X. & Macfarlan, T. S. The role of KRAB-ZFPs in transposable element repression and mammalian evolution. *Trends Genet.* **33**, 871–881 (2017).
36. Meyer, A. *et al.* Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* **590**, 284–289 (2021).
37. Rogers, R. L. *et al.* Genomic takeover by transposable elements in the strawberry poison frog. *Mol. Biol. Evol.* **35**, 2913–2927 (2018).
38. Zhou, W., Liang, G., Molloy, P. L. & Jones, P. A. DNA methylation enables transposable element-driven genome expansion. *Proc. Natl. Acad. Sci. USA* **117**, 19359–19366 (2020).
39. Pallavicini, A. *et al.* Analysis of the transcriptome of the Indonesian coelacanth *Latimeria menadoensis*. *BMC Genom.* **14**, 538 (2013).
40. Biscotti, M. A. *et al.* The lungfish transcriptome: A glimpse into molecular evolution events at the transition from water to land. *Sci. Rep.* **6**, 21571 (2016).

41. Amemiya, C. T. *et al.* The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**, 7445 (2013).
42. Birtle, Z. & Ponting, C. P. Meisetz and the birth of the KRAB motif. *Bioinformatics* **22**, 2841–2845 (2006).
43. Baker, Z. *et al.* Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife* **6**, e24133 (2017).
44. Meroni, G. Genomics and evolution of the TRIM gene family. *Adv. Exp. Med. Biol.* **770**, 1–9 (2012).
45. Liedtke, H. C. *et al.* Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. *Nat. Ecol. Evol.* **2**, 1792–1799 (2018).
46. Laurin, M., Canoville, A., Struble, M., Organ, C. & de Buffrénil, V. Early genome size increase in urodeles. *C.R. Palevol.* **15**, 74–82 (2016).
47. Lamichhaney, S. *et al.* A bird-like genome from a frog: Mechanism of genome size reduction in the ornate burrowing frog, *Platyplectrum ornatum*. *Proc. Natl. Acad. Sci. USA* **118**, e2011649118 (2021).
48. Biscotti, M. A. *et al.* The transcriptome of the newt *Cynops orientalis* provides new insights into evolution and function of sexual gene networks in sarcopterygians. *Sci. Rep.* **10**, 5445 (2020).
49. Dong, Y. *et al.* Expression dynamics and relations with nearby genes of rat transposable elements across 11 organs, 4 developmental stages and both sexes. *BMC Genom.* **18**, 666 (2017).
50. Zamudio, N. & Bourc'his, D. Transposable elements in the mammalian germline: A comfortable niche or a deadly trap? *Heredity* **105**, 92–104 (2010).
51. Metcalfe, C. J., File´e, J., Germon, I., Joss, J. & Casane, D. Evolution of the Australian lungfish (*Neoceratodus forsteri*) genome: A major role for *CR1* and *L2* LINE elements. *Mol. Biol. Evol.* **29**, 3529–3539 (2012).
52. Metcalfe, C. J. & Casane, D. Accommodating the load: The transposable element content of very large genomes. *Mob. Genet. Elements* **3**, e24775 (2013).
53. Organ, C. L., Struble, M., Canoville, A., de Buffrénil, V. & Laurin, M. Macroevolution of genome size in sarcopterygians during the water-land transition. *C.R. Palevol.* **15**, 65–73 (2016).
54. Liedtke, H. C. *et al.* No ecological opportunity signal on a continental scale? Diversification and life-history evolution of African true toads (anura: bufonidae). *Evolution* **70**, 1717–1733 (2016).
55. Sirijovsky, N., Woolnough, C., Rock, J. & Joss, J. M. *Nf*CR1, the first non-LTR retrotransposon characterized in the Australian lungfish genome, *Neoceratodus forsteri* shows similarities to CR1-like elements. *J. Exp. Zool.* **304**, 40–49 (2005).
56. McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792–801 (1984).
57. Storer, J., Hubley, R., Rosen, J., Wheeler, T. J. & Smit, A. F. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob. DNA* **12**, 2 (2021).
58. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
59. Abrusán, G., Grundmann, N., DeMester, L. & Makalowski, W. TEclass—A tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* **25**, 1329–1330 (2009).
60. Lima, L. *et al.* Playing hide and seek with repeats in local and global de novo transcriptome assembly of short RNA-seq reads. *Algorithms Mol. Biol.* **12**, 2 (2017).
61. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
62. Marchler-Bauer, A. & Bryant, S. H. CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res.* **32**, 327–331 (2004).
63. Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**, 281–285 (2012).
64. Baggerly, K., Deng, L., Morris, J. & Aldaz, C. Differential expression in SAGE: Accounting for normal between-library variation. *Bioinformatics* **19**, 1477–1483 (2003).
65. Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**, 2310–2314 (2001).
66. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 (1992).

Acknowledgements

This research and the APC were funded by Ministero Italiano dell'Istruzione, dell'Università e della Ricerca grant number 55_R.SCIENT.A_2020.

Author contributions

F.C., M.B., A.C., and M.A.B. conceived the study; E.C., F.C., S.G., M.G., and M.A.B. were involved in methodology; S.G. and M.G. performed the evaluation of transposable elements activity and gene expression analyses on target genes; M.G. and M.A.B. validated data; E.C. and F.C. performed phylogenetic analyses and protein domain characterization; writing—original draft preparation, all authors; A.C. and M.A.B. supervised the manuscript; M.A.B acquired funding. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-94193-6.

Correspondence and requests for materials should be addressed to M.B.

Reprints and permissions information is available at www.nature.com/reprints.

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

3.4 Transposable elements and teleost migratory behaviour

International Journal of
Molecular Sciences

MDPI

Article

Transposable Elements and Teleost Migratory Behaviour

Elisa Carotti [1,†], Federica Carducci [1,†], Adriana Canapa [1], Marco Barucca [1,*], Samuele Greco [2], Marco Gerdol [2] and Maria Assunta Biscotti [1]

[1] Department of Life and Environmental Sciences, Polytechnic University of Marche, Via Brecce Bianche, 60131 Ancona, Italy; e.carotti@pm.univpm.it (E.C.); f.carducci@univpm.it (F.C.); a.canapa@univpm.it (A.C.); m.a.biscotti@univpm.it (M.A.B.)
[2] Department of Life Sciences, University of Trieste, Via L. Giorgieri, 5-34127 Trieste, Italy; SAMUELE.GRECO@phd.units.it (S.G.); mgerdol@units.it (M.G.)
* Correspondence: m.barucca@univpm.it
† Equal contribution.

Abstract: Transposable elements (TEs) represent a considerable fraction of eukaryotic genomes, thereby contributing to genome size, chromosomal rearrangements, and to the generation of new coding genes or regulatory elements. An increasing number of works have reported a link between the genomic abundance of TEs and the adaptation to specific environmental conditions. Diadromy represents a fascinating feature of fish, protagonists of migratory routes between marine and freshwater for reproduction. In this work, we investigated the genomes of 24 fish species, including 15 teleosts with a migratory behaviour. The expected higher relative abundance of DNA transposons in ray-finned fish compared with the other fish groups was not confirmed by the analysis of the dataset considered. The relative contribution of different TE types in migratory ray-finned species did not show clear differences between oceanodromous and potamodromous fish. On the contrary, a remarkable relationship between migratory behaviour and the quantitative difference reported for short interspersed nuclear (retro)elements (SINEs) emerged from the comparison between anadromous and catadromous species, independently from their phylogenetic position. This aspect is likely due to the substantial environmental changes faced by diadromous species during their migratory routes.

Keywords: genome evolution; transposable elements; environmental adaptation; fish

Citation: Carotti, E.; Carducci, F.; Canapa, A.; Barucca, M.; Greco, S.; Gerdol, M.; Biscotti, M.A. Transposable Elements and Teleost Migratory Behaviour. *Int. J. Mol. Sci.* **2021**, *22*, 602. https://doi.org/10.3390/ijms22020602

Received: 2 December 2020
Accepted: 6 January 2021
Published: 9 January 2021

1. Introduction

Teleosts, comprising more than 32,000 extant species [1], represent an evolutionarily successful and highly diverse group of vertebrates that populate a wide range of both sea water (SW) and freshwater (FW) habitats across the world, from polar to tropical regions [2]. The genome composition of these organisms most certainly represents one of the key factors behind such evolutionary success. Two major events of whole genome duplication (WGD) occurred during the early evolution of vertebrates 500 million years ago, and thus also affected the lineage of actinopterygians. Subsequently, a third duplication (3R) took place, 300 million years ago, in the teleost lineage and specific events (4R) occurred independently only in some lineages [3–5]. The genome size of ray-finned fish is characterised by a wide range of variation, with the smallest values found in Tetraodontiformes species (~0.35 Gb) and the highest values in Acipenseriformes (~9.5 Gb). These differences are mostly ascribable to the presence of repetitive DNA [6–8]. Indeed, as postulated by the C-value paradox, the complexity of an organism is not related to the amount of DNA, because not all DNA is made up of genes, but it is mostly constituted by intergenic non-coding DNA, and in particular by repetitive DNA [9]. These large genomic regions contain highly repeated sequences that are classified into two main groups: transposable elements (TEs) and tandem repeat elements, with the latter including satellite DNA, minisatellites, and microsatellites [10–12]. TEs are DNA sequences capable of replicating, moving, and integrating into new regions

of the genome. They are divided into two main classes: (i) retrotransposons (Class I) are able to propagate their copy sequences in the host genome by reverse transcription of an RNA intermediate molecule through a copy-and-paste mechanism; (ii) DNA transposons (Class II) are generally eliminated from their original location in order to be inserted in a different portion of the genome by a cut-and-paste mechanism. The Class I retroelements include long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons, these latter encompassing long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) [13]. The structure of LTR retroelements comprises the open reading frames (ORFs) encoding for viral structural proteins (GAG), an aspartic protease (Pol), a reverse transcriptase (RT), an RNase H, and an integrase. LINE retroelements encode a reverse transcriptase and a nuclease necessary for transposition. Conversely, SINE retroelements are non-autonomous and their origins are related to tRNA, 7SL RNA, and 5S RNA. The Class II includes autonomous elements characterised by the presence of terminal inverted repeats and the ability to encode for a transposase. This Class of TEs comprises also *Helitrons* that use a rolling-circle mechanism for replication [13]. Although repetitive DNA was long considered to be devoid of any functional meaning and therefore labelled as "junk DNA", a large number of reports now support a functional and evolutionary role for a significant fraction of the non-coding portion of vertebrate genomes [14–17]. In particular, TEs can be considered important factors for the reorganisation of the genome through chromosomal rearrangements such as duplications, inversions, translocations, and creation of novel genes or regulatory elements through molecular domestication [7,18,19] but also function themselves as enhancers, promoters, silencers, and boundary elements [20,21]. Overall, the fish mobilome shows a wide variety of TE superfamilies, and, compared to other vertebrates, it is mainly dominated by DNA transposons [22–25]. In addition, some lineage-specific TEs have played a significant role in karyotype evolution and, in particular, in the genesis of sex chromosomes [26–31]. In plants, several works have highlighted a relationship between TEs and the environment to which species are adapted [32–36]; however, growing evidence suggests that TEs might be among the main drivers of adaptation even in animals [37,38]. This aspect is particularly intriguing in fish, which are characterised by a high biodiversity [39–41].

Therefore, the possible correlation between TEs and migration, one of the most intriguing behaviour of fish, is of extreme interest. To cope with aquatic environments characterised by substantial differences in salinity and temperature, migrating organisms have evolved an extraordinary physiological plasticity. In this context, diadromous fish species represent a particularly interesting case study. This group of fish has always attracted scientific attention due to their spectacular migratory routes between SW and FW, and some of them are also of commercial importance for fishing activities. Three different types of diadromy have been recognised: anadromy and catadromy are related to spawning and differ for the passage from SW to FW and vice versa, respectively. A third typology, still debated among fish biologists, is amphidromy, for which migration to SW is not correlated with spawning necessity and occurs in a specific life stage (hatched larvae) for a restricted period of time (up to 200 days), after which juveniles return to FW, where they spend their adult life [42,43]. On the contrary, oceanodromous and potamodromous species migrate within SW and FW, respectively.

While multiple studies are progressively building up a significant amount of evidence corroborating the existence of a correlation between the genomic abundance of particular TE classes and the adaptation to specific environments, to date this fascinating topic has never been comprehensively investigated in migrating fish. With the present study, we aim to fill this knowledge gap by analysing the TE content in 24 fish species, including 15 having a migratory behaviour.

2. Results

2.1. Evaluation of TE Impact in Bony Fish Genomes

The masking analysis of genome repetitive fraction showed a variable TE content in fish genomes (Figure 1, Supplementary Table S1). The highest abundances of TEs were detected in the *Petromyzon marinus*, *Callorhinchus milii*, *Neogobius melanostomus*, and *Danio rerio* genomes. The impact of each TE type on the genome of the analysed species was graphically represented as the TE relative contribution (Figure 2), that evidenced a clear difference between the two non-bony vertebrates, where LINE retroelements were clearly prevalent compared to actinopterygians. A balance between Class I retroelements (LINE, SINE, and LTR) and Class II DNA transposons was observed in the three species belonging to the *Anguilla* genus, in *Salmo salar*, *Periophthalmodon schlosseri, Dicentrarchus labrax, Thunnus orientalis,* and *Sinocyclocheilus grahami.* In the remaining ray-finned fish, retroelements were more abundant than DNA transposons, except for *Cyprinus carpio*, *D. rerio*, *Scartelaos histophorus*, *N. melanostomus*, and *Astyanax mexicanus*, where a major impact of DNA transposons emerged. We recorded a very similar amount of DNA transposons and LINE retroelements in *Oncorhynchus mykiss*, *Gadus morhua,* and *Arapaima gigas*. The impact of LTR retroelements on the genome of *A. gigas, Tenualosa ilisha*, and *Lates calcarifer* was higher than the other species. Overall, SINE retroelements had a minor relative contribution in all ray-finned fish species, except for *Lepisosteus oculatus* and *Scleropages formosus*.

We found a positive correlation between TE content and assembled genome size, supported by statistically significant Spearman and Kendall correlations (Spearman: rho = 0.536, *p*-value = 0.009; Kendall: tau = 0.352, *p*-value = 0.019; Supplementary Figure S1).

Moreover, the trends evidenced from the distribution of TE relative abundances was not consistent with species phylogeny.

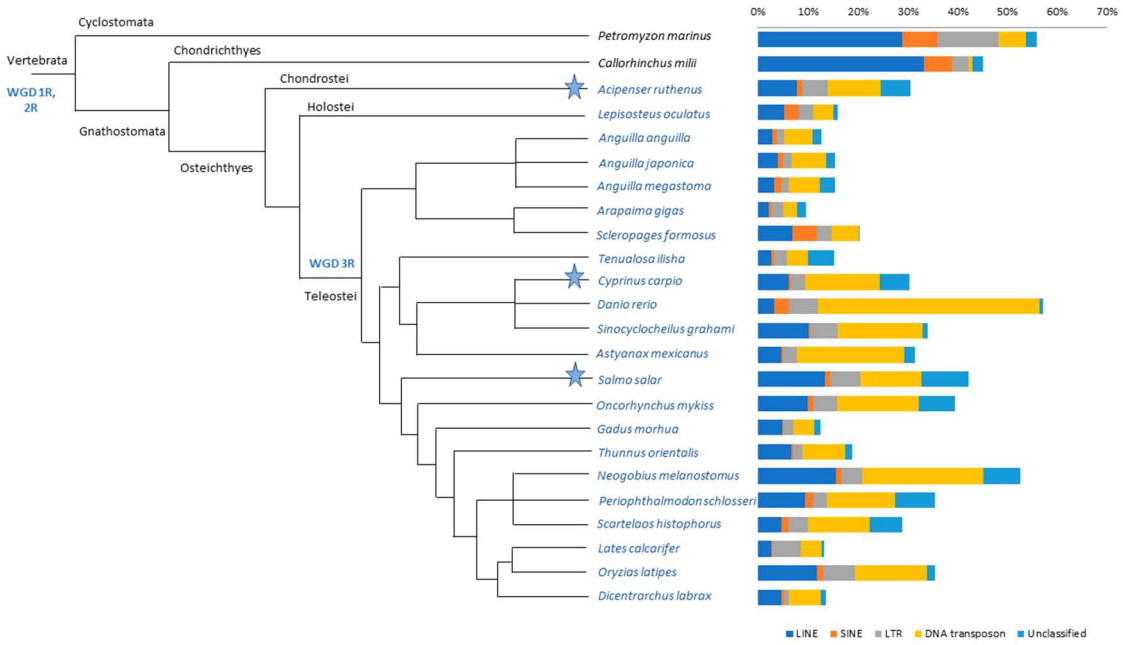

Figure 1. Cladogram showing the relationships between analysed species. Main clades are reported above the branches. 1R, 2R, and 3R whole genome duplication (WGD) events are reported, and the light-blue star indicates species that have undergone 4R-specific events. Species belonging to Actinopterygii are shown in light blue. The cladogram was modified from Betancur-R et al. [44]. The percentages of total transposable elements (TEs) masked in the genomes of the studied species are shown on the right-hand side. Each bar displays the percentage of the main TE types: DNA transposons in yellow; LTR retroelements in grey; SINE retroelements in orange; LINE retroelements in dark blue; and unclassified elements in light blue.

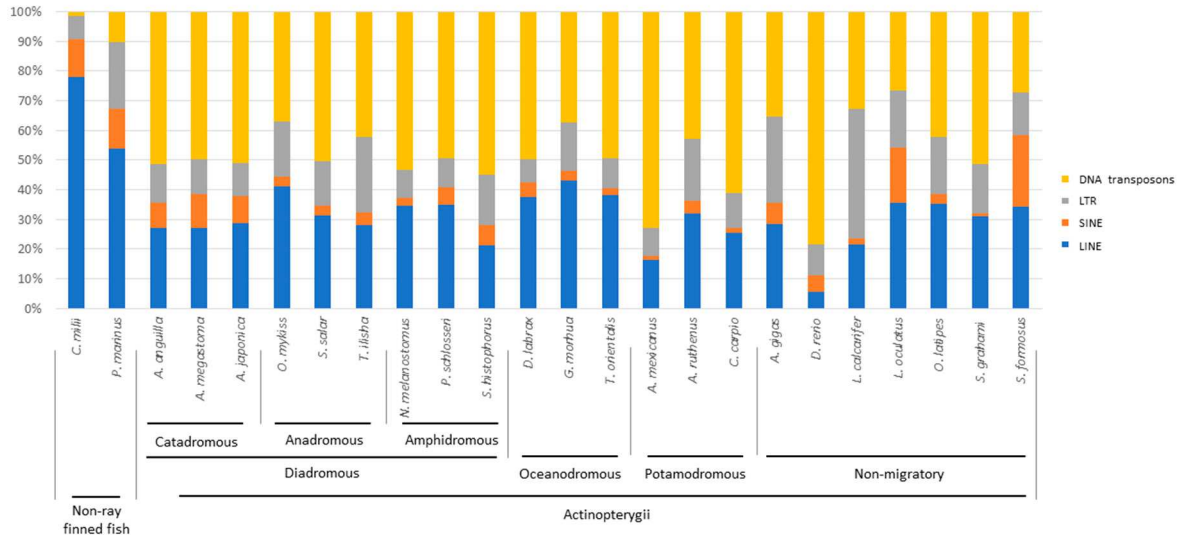

Figure 2. Relative abundance of TE types in the mobilome of the fish species analysed. The histogram shows the relative abundance of each TE type in the mobilome of two non-ray finned fish and 22 actinopterygians. The migratory behaviour of Actinopterygii is indicated.

2.2. TE Contribution in Migratory Species Genomes

Nine of the species analysed here were characterised by a diadromous behaviour: the three eel species (*Anguilla anguilla*, *Anguilla japonica*, *Anguilla megastoma*) are catadromous, *T. ilisha*, *S. salar*, and *O. mykiss* are anadromous, while *N. melanostomus, P. schlosseri,* and *S. histophorus* are amphidromous. The comparison of TE relative abundances enabled recognition of distinct patterns in the different fish groups. In particular, concerning retroelements, the three catadromous species presented a higher abundance of LINE, followed by a very similar amount of LTR and SINE retroelements, while SINE retroelements were the least represented in the other six species considered (Figure 2, Supplementary Table S2).

One-way analysis of variance (ANOVA) identified the abundance of SINE retroelements as the only statistically significant difference (p-value < 0.05) between anadromous and catadromous species, and the abundance of DNA transposons as the only significant difference between amphidromous and catadromous species.

A potential caveat of this study is linked with the congeneric status of the three catadromous species we considered, which may have led to the introduction of a phylogenetic signature in genomic TE composition in this group. Unfortunately, no fully sequenced genome of other catadromous species not belonging to *Anguilla* genus is presently available to broaden taxonomical sampling.

The variation partitioning analyses were performed to test the influence of phylogeny or migration on TE quantitative composition. Data obtained showed a marked correlation of migratory behaviour on the quantitative difference observed for SINE retroelements in the comparison between catadromous and anadromous species. Moreover, migratory behaviour was found to have a slighter correlation on the quantitative difference of DNA transposons observed between catadromous and amphidromous species (Figure 3).

To provide a fine-scale overview of the expansion of particular TE families in catadromous and anadromous species, we analysed the relative abundance of TE families, generating two Z-score-based heat maps (Supplementary Figure S2). Despite their high phylogenetic distance, the three anadromous species displayed a very similar content of SINE retroelements, much lower than the three closely related catadromous species. Curiously, some elements (*SINE/tRNA* and *SINE/tRNA-Core*) displayed a similar abundance between anadromous and catadromous species, except for *A. anguilla*, where they were particularly expanded (Supplementary Figure S2A). While larger interspecific differences were evidenced by a broader overview of the relative abundance of all TEs (Supplementary

Figure S2B), the clustering of anadromous and catadromous species in two distinct groups was still strongly supported. Moreover, the comparison between the two heat maps evidenced distinct relationships between patterns observed for the three species belonging to the *Anguilla* genus: Supplementary Figure S2B shows a cluster comprising *A. anguilla* and *A. japonica*, and in an external position *A. megastoma*, while Supplementary Figure S2A outlines a closer relationship between *A. japonica* and *A. megastoma*; for the anadromous species considered, the two heat maps did not show any difference in terms of relationships between the obtained patterns.

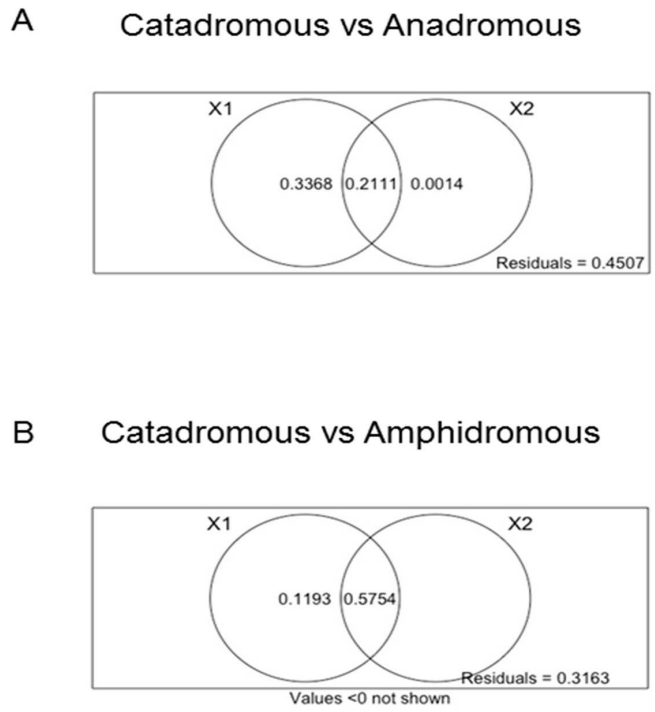

Figure 3. Venn diagrams obtained by variation partitioning analyses (VPA) using redundancy analysis (RDA). The partition of the variation of a response variable (X) between two sets of explanatory variables (X1 and X2) is shown. Each circle represents the portion of variation accounting for an explanatory variable or a combination of the explanatory matrices. The intersection between the two circles represents the amount of variation explained by both variables X1 and X2 [45]. (**A**): the response variable (X) is the quantitative difference reported for SINE retroelements from the comparison between anadromous and catadromous fish species; the explanatory variable X1 represents the migration and the explanatory variable X2 indicates the sequence divergence obtained by p-distance matrix using 16S rDNA. (**B**): the response variable (X) is the quantitative difference reported for DNA transposons from the comparison between catadromous and amphidromous fish species; the explanatory variable X1 represents the migration and the explanatory variable X2 indicates the sequence divergence obtained by p-distance matrix using 16S rDNA.

Although a specific pattern was not clear in the comparison between the relative TE abundances of diadromous, oceanodromous and potamodromous teleosts, we observed that oceanodromous species were generally characterised by a higher quantity of LINE retroelements (>37%) compared to potamodromous and diadromous species, except for *O. mykiss* (Figure 2).

2.3. Kimura Distance-Based Copy Divergence Analyses of Transposable Elements

The analysis of TE sequence divergence by Kimura distance evidenced the presence of one or two amplification bursts in all actinopterygians. The position of these bursts on the left part of the graph (K-value < 25, that indicates a relatively small degree of divergence) suggested a link with recent amplification events. Moreover, the repeat landscapes of the

analysed ray-finned fish species showed a remarkable DNA elimination rate due to the distribution of the largest part of TE copies below a K-value of 25 (Figure 4, Supplementary File S1). In most species, DNA transposons represented more than 50% out of all the TEs identified, except for *L. oculatus*, *A. gigas*, and *S. formosus*, in which all the main TE classes were represented. The contribution of SINE retroelements was notable in the spotted gar and in the Asian arowana, in contrast with the other ray-finned fish analysed. No correlation could be observed between the composition of the repeat landscapes and species phylogeny, as clearly summarised in Figure 4.

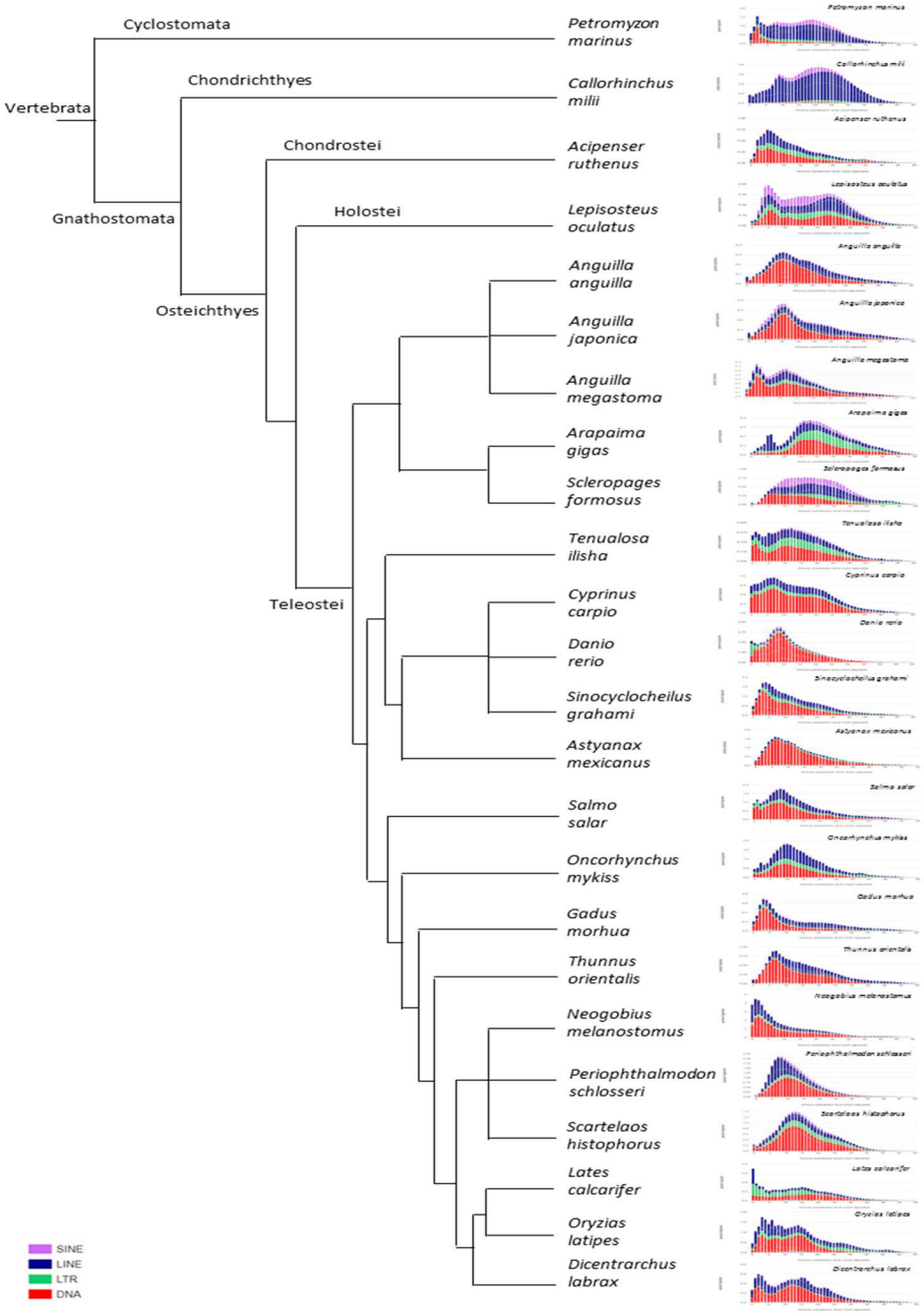

Figure 4. Cladogram with repeat landscape plots. The cladogram reported on the left summarises the phylogenetic relationships among the species analysed in this study. The graphs displayed on the right report the repeat landscape plots obtained by Kimura distance-based copy divergence analyses of transposable elements for each species. The repeat landscape plots are reported at higher resolution in Supplementary File S1.

3. Discussion

Genome architecture is one of the key factors influencing the evolutionary success of species [46]. Transposable elements are known to play a pivotal role as a dynamic component of the DNA repetitive fraction. In this framework, the evaluation of the impact of mobile elements in ray-finned fish, one of the most diversified group of vertebrates, may contribute not only to unravel the biodiversity of this taxon, but also to unveil the reason of their extreme adaptability to widely different aquatic environmental conditions. The evolution of specific TE sequences has been linked with environmental temperature [39]. Moreover, an increasing number of studies have reported the additional influence of abiotic factors on TE activity [32,33,37]. Migration is one of the most fascinating behaviours in the lifestyle of some fish species. In particular, this behaviour leads diadromous species to move between environments characterised by substantial differences in abiotic factors (e.g., salinity and temperature). In this study, we investigated the genomes of 15 migratory species: *T. ilisha, O. mykiss,* and *S. salar* are anadromous; *A. anguilla*, *A. japonica,* and *A. megastoma* are catadromous; *N. melanostomus, P. schlosseri,* and *S. histophorus* are amphidromous; *G. morhua, T. orientalis,* and *D. labrax* are oceanodromous; *Acipenser ruthenus, C. carpio,* and *A. mexicanus* are potamodromous. For comparison, two non-bony fish (the jawless *P. marinus* and the elephant shark *C. milii*) and seven additional non-migratory actinopterygian species were considered (Figure 1).

Our results showed a variable TE content both in the bony and non-bony fish genomes analysed, as well as a positive correlation between TE content and the assembled genome size. This suggested, in line with previous reports, that these elements affected the content of nuclear DNA [7,12,22,23,47–51]. In addition, differences in TE content were not associated either with the number of whole genome duplication (WGD) events the analysed species underwent, nor with the ploidy level. In particular, the genome of actinopterygians experienced two rounds of WGDs [52–54], and a third round occurred in teleosts [54–56]. Several other ray-finned fish species experienced a fourth independent WGD event [3–5,57].

In the TE relative contribution analyses, LINE retroelements had the highest impact in sea lamprey and elephant shark, as previously reported by other authors [22,25,40]. Although the known trend of higher relative abundance of DNA transposons in ray-finned fish was confirmed by our analyses, we evidenced some exceptions. In particular, *A. ruthenus, L. oculatus, A. gigas, S. formosus, T. ilisha, O. mykiss, G. morhua, L. calcarifer*, and *O. latipes* showed a lower number of DNA transposons compared with retroelements. The high relative TE content of DNA transposons observed in the only available species of Acipenseriformes, *A. ruthenus*, might reflect a condition similar to that one present in the common ancestor of Actinopterygii.

However, the patterns of TE relative abundance observed across species (Figure 2) were not consistent with phylogeny. This suggested that the accumulation of specific TEs might be related to the evolutionary adaptation of species to specific environmental niches. Indeed, variation partitioning analyses evidenced a statistically significant influence of migratory behaviour on quantitative differences of DNA transposons between catadromous and amphidromous species. The effect of migration was more pronounced on the quantitative difference reported for SINE retroelements in the comparison between anadromous and catadromous fish species. One study has reported that resident and migratory populations of the teleost *Coilia nasus* carry a different SINE copy number. These differences have been related to the anadromous ecotypes of *C. nasus* that cope with environmental challenges during their life cycle [41]. The abundance of SINE retroelements may contribute to the genomic variation in fish, affecting gene expression. Indeed, SINEs are often enriched at the boundaries of transcriptionally active or inactive domains and these elements are thought to be involved in defining high-order genomic organisation through intra- or inter-chromosomal interactions [58–60].

The analysis performed to identify fine-scale expansions of SINE retroelements between catadromous and anadromous species showed that most of the elements experienced an expansion in catadromous species. Moreover, the relationships between SINE retroele-

ments patterns observed for the three species of *Anguilla* genus were different from those obtained considering all TEs. These findings strengthen the hypothesis that the similarity in the abundance of SINE retroelements between *Anguilla* species analysed might be also due to different causes from phylogeny. However, the differences in SINE content between catadromous and anadromous species could be investigated more extensively when genomic data from other catadromous species not belonging to *Anguilla* genus are available.

On the contrary, the relative contribution of TEs did not show a clear pattern, either among oceanodromous or among potamodromous species. This finding might be linked to the fact that these species spend their entire life in marine- and freshwaters, respectively, without having to face substantial environmental changes. Overall, the comparison between the oceanodromous–potamodromous and the anadromous–amphidromous–catadromous groups evidenced a lower SINE content in the former, strengthening the hypothesis that these retroelements play a key role in the genomes of fish species characterised by diadromous migratory behaviour. However, further investigations are needed to clarify the cause–effect relationships between the co-evolution of the number of SINE retroelements and the migratory behaviour of fish.

The reconstruction of the transposition history through Kimura distance showed one or two amplification bursts in the repeat landscape profiles of catadromous and anadromous species, with the differences in the relative content of SINE that might be explained by three scenarios. The higher relative content of SINEs in catadromous species is probably related to a higher rate of their transposition. Moreover, purifying selection could have acted as the main driver in determining the discrepancy in TE transposition rate between catadromous and anadromous species. On the other hand, the effect of TE silencing and/or elimination mechanisms might be more active in anadromous species, affecting SINE retroelements content. Moreover, the role played by horizontal transfer events [61–63] might not be negligible, because these could have conveyed beneficial SINE retroelements in the host genome of catadromous species, leading to their expansion.

Finally, the importance of sequencing depth and the assembly methods need to be considered, mainly with respect to repeat-rich regions, because these technical factors might influence the estimation of transposable element content. However, the combination of species-specific repeat libraries with the de novo discovery and annotation process described in this work allowed us to improve the identification of previously unidentifiable non-canonical repeats.

4. Materials and Methods

A total of 24 fish species were considered in this study: a single species belonging to Cyclostomata (lamprey, *P. marinus*), within Gnathostomata, one member of Chondrichthyes (elephant shark, *C. millii*) and 22 Osteichthyes. These all belong to Actinopterygii and, in detail included one Chondrostei (sterlet, *A. ruthenus*), one Holostei (spotted gar, *L. oculatus*) and 20 Teleostei (European eel, *A. anguilla*; Japanese eel, *A. japonica*; Polynesian longfinned eel, *A. megastoma*; pirarucu, *A. gigas*, arowana, *S. formosus*; Hilsa Shad, *T. ilisha*; carp, *C. carpio*; zebrafish, *D. rerio*; golden line barbel, *S. grahami*; cavefish, *A. mexicanus*; salmon, *S. salar*; rainbow trout, *O. mykiss*; Atlantic cod, *G. morhua*; tuna, *T. orientalis*; round goby, *N. melanostomus*; giant mudskipper, *P. schlosseri*; walking goby, *S. histophorus*; barramundi, *L. calcarifer*; medaka, *O. latipes*; European seabass, *D. labrax*). Unmasked genomes of these species were obtained from the public database NCBI GenBank (https://www.ncbi.nlm.nih.gov/genome/), except for the Atlantic cod genome, which was downloaded from Ensembl genome browser (https://www.ensembl.org/index.html). Accession numbers are reported in Supplementary Table S3.

To identify transposable elements in the analysed species, we used species-specific TE libraries derived from the FishTEDB database (http://www.fishtedb.org/). For the species *A. ruthenus*, *A. gigas*, *T. ilisha*, *C. carpio*, *O. mykiss*, *S. salar*, *N. melanostomus, P. schlosseri,* and *S. histophorus*, de novo TE libraries were built as follows. De novo TE identification was

performed using RepeatScout v 1.0.5 [64]: the "build_lmer_table" module generated a table of lmer frequency and "RepeatScout" extracted the repeats that were then filtered with "filter-stage-1.prl" script in order to remove low complexity sequences. The filtered output file was used by RepeatMasker v 4.1.0 (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) as a library to extract repeats from each downloaded genome. The RepeatScout "filter-stage-2.prl" script was then employed to remove sequences repeated less than 10 times. The remaining sequences were then filtered by BLASTX [65] search against the Uniprot–Swissprot database [66] and by Interproscan v5.34-73.0 [67], by removing sequences with at least one hit (e-value = 1e-50), because they could be coding sequences not ascribable to TEs. In order to avoid the loss of domesticated transposons sequences filtered out in the previous step, discarded elements were analysed with HMMER [68], searching for domains ascribable to integrase, reverse transcriptase, and transposase functions. The corresponding HMM profiles (PF13333.7, PF13683.7, PF00665.27, PF00078.28, PF13843.7) were downloaded from Pfam [69]. Sequences with an e-value lower than 1e-5 were then reintegrated in the TE library. Finally, TEclass-2.13 was used to classify the remaining sequences. All the libraries were used to mask each genome with RepeatMasker, setting the—a argument in order to obtain the alignment file for each species.

An extended library comprising all TEs identified in catadromous and anadromous fish species was used to mask their genomes, obtaining a comparable dataset among species, representative of the relative genomic abundance of specific TE families. This was calculated by dividing the total amount of masked nucleotides for each family by the total amount of masked nucleotides using the entire library on each genome.

To estimate TE age and transposition history in actinopterygian genomes, Kimura distances (rate of transition and transversions) were calculated between genome total sequence length and TE consensus from the library using the scripts "calcDivergenceFromAlign.pl" and "createRepeatLandscape.pl" provided by the RepeatMasker package.

The main steps of the pipeline are graphically summarised in Supplementary Figure S3.

To test a relationship between genome size (total sequence length) and total TE content (total TE, reported here as a percentage of the total genome size), we applied the Spearman and Kendall rank correlation tests using "cor.test()" function in R.

Variation partitioning analysis was performed using vegan package 2.5.5 [70] to evaluate the influence of transposable elements and/or phylogeny on migratory behaviour. Migration was assigned as the explanatory variable X1 and phylogeny (considered as the sequence divergence obtained by the p-distance matrix using 16S rDNA) as the explanatory variable X2. For each pair of species, both explanatory variables were compared with data related to the difference calculated for each TE type (response variable X).

5. Conclusions

The evolutionary success of species is strictly related to the composition and functionality of their genome. Transposable elements undoubtedly have a key role in shaping genome architecture by generating the genetic innovations responsible for species' adaptability. Ray-finned fish are characterised by an extremely high diversity and are adapted to a wide range of environments. The data presented here did not show any significant correlation between TE composition and the phylogenetic relationships among species, although they indicated an interesting link between the genomic content of mobile elements and environmental conditions. It is not easy to establish whether the content of a specific TE-type is determined by the environment in which a given species lives or vice versa. However, these findings follow the well-established trends that envision TEs as primary drivers of the exceptional biodiversity of species and lay the foundation for further experimental analyses towards the comprehension of mechanisms and factors involved in this fascinating relationship.

Supplementary Materials: Supplementary materials can be found at https://www.mdpi.com/1422-0067/22/2/602/s1.

Author Contributions: Conceptualization, E.C., F.C., M.B. and M.A.B.; methodology, E.C., F.C., S.G., M.G., and M.A.B.; software, S.G. and M.G.; validation, M.G. and M.A.B.; formal analysis, E.C., F.C. and S.G.; data curation, E.C. and F.C.; writing—original draft preparation, all authors; writing—review and editing, all authors; supervision, A.C. and M.A.B.; funding acquisition, M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC were funded by Ministero Italiano dell'Istruzione, dell'Università e della Ricerca grant number 040017_R.SCIENT.A_2018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Froese, R.; Pauly, D. (Eds.) FishBase, Version 12/2019. Available online: www.fishbase.org (accessed on 10 May 2020).
2. Glasauer, S.M.K.; Neuhauss, S.C. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genom.* **2014**, *289*, 1045–1060. [CrossRef] [PubMed]
3. Alexandrou, M.A.; Swartz, B.A.; Matzke, N.J.; Oakley, T.H. Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Mol. Phylogenet. Evol.* **2013**, *69*, 514–523. [CrossRef] [PubMed]
4. MacQueen, D.J.; Johnston, I.A. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. Biol. Sci.* **2014**, *281*, 20132881. [CrossRef] [PubMed]
5. Wang, J.-T.; Li, J.-T.; Zhang, X.; Sun, X.-W. Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (Cyprinus carpio). *BMC Genom.* **2012**, *13*, 96. [CrossRef]
6. Biscotti, M.A.; Carducci, F.; Olmo, E.; Canapa, A. Vertebrate genome size and the impact of transposable elements in genome evolution. In *Evolution, Origin of Life, Concepts and Methods*; Springer International Publishing: Cham, Switzerland, 2019; pp. 233–251.
7. Canapa, A.; Barucca, M.; Biscotti, M.A.; Forconi, M.; Olmo, E. Transposons, genome size, and evolutionary insights in animals. *Cytogenet. Genome Res.* **2015**, *147*, 217–239. [CrossRef]
8. Carducci, F.; Barucca, M.; Canapa, A.; Carotti, E.; Biscotti, M.A. Mobile elements in ray-finned fish genomes. *Life* **2020**, *10*, 221. [CrossRef]
9. Thomas, C.A. The genetic organization of chromosomes. *Annu. Rev. Genet.* **1971**, *5*, 237–256. [CrossRef]
10. Biscotti, M.A.; Olmo, E.; Heslop-Harrison, J.S. Repetitive DNA in eukaryotic genomes. *Chromosom. Res.* **2015**, *23*, 415–420. [CrossRef]
11. Franck, J.P.; Harris, A.S.; Bentzen, P.; Denovan-Wright, E.M.; Wright, J.M. Organization and evolution of satellite, minisatellite and microsatellite DNAs in teleost fishes. *Oxf. Surv. Eukaryot. Genes* **1991**, *7*, 51–82.
12. Garrido-Ramos, M.A. Satellite DNA: An evolving topic. *Genes* **2017**, *8*, 230. [CrossRef]
13. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.J.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [CrossRef] [PubMed]
14. Pierron, F.; Daffe, G.; Lambert, P.; Couture, P.; Baudrimont, M. Retrotransposon methylation and activity in wild fish (A. anguilla): A matter of size. *Environ. Pollut.* **2019**, *245*, 494–503. [CrossRef] [PubMed]
15. Biscotti, M.A.; Canapa, A.; Forconi, M.; Olmo, E.; Barucca, M. Transcription of tandemly repetitive DNA: Functional roles. *Chromosom. Res.* **2015**, *23*, 463–477. [CrossRef] [PubMed]
16. Canapa, A.; Biscotti, M.A.; Barucca, M.; Carducci, F.; Carotti, E.; Olmo, E. Shedding light upon the complex net of genome size, genome composition and environment in chordates. *Eur. Zool. J.* **2020**, *87*, 192–202. [CrossRef]
17. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [CrossRef] [PubMed]
18. Carducci, F.; Biscotti, M.A.; Barucca, M.; Canapa, A. Transposable elements in vertebrates: Species evolution and environmental adaptation. *Eur. Zool. J.* **2019**, *86*, 497–503. [CrossRef]
19. Warren, I.A.; Naville, M.; Chalopin, D.; Levin, P.; Berger, C.S.; Galiana, D.; Volff, J.-N. Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. *Chromosom. Res.* **2015**, *23*, 505–531. [CrossRef]
20. Sundaram, V.; Wysocka, J. Transposable elements as a potent source of diverse cis -regulatory sequences in mammalian genomes. *Philos. Trans. R. Soc. B Biol. Sci.* **2020**, *375*, 20190347. [CrossRef]
21. Moschetti, R.; Palazzo, A.; Lorusso, P.; Viggiano, L.; Marsano, R.M. "What you need, baby, I got it": Transposable elements as suppliers of cis-operating sequences in drosophila. *Biology* **2020**, *9*, 25. [CrossRef]
22. Chalopin, D.; Naville, M.; Plard, F.; Galiana, D.; Volff, J.-N. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* **2015**, *7*, 567–580. [CrossRef]
23. Chalopin, D.; Volff, J.-N. Analysis of the spotted gar genome suggests absence of causative link between ancestral genome duplication and transposable element diversification in teleost fish. *J. Exp. Zool. Part. B Mol. Dev. Evol.* **2017**, *328*, 629–637. [CrossRef] [PubMed]

24. Gao, B.; Shen, D.; Xue, S.; Chen, C.; Cui, H.; Song, C. The contribution of transposable elements to size variations between four teleost genomes. *Mob. DNA* **2016**, *7*, 1–16. [CrossRef] [PubMed]
25. Shao, F.; Han, M.; Peng, Z. Evolution and diversity of transposable elements in fish genomes. *Sci. Rep.* **2019**, *9*, 1–8. [CrossRef] [PubMed]
26. Volff, J.N.; Korting, C.; Sweeney, K.; Schartl, M. The non-LTR retrotransposon Rex3 from the fish Xiphophorus is widespread among teleosts. *Mol. Biol. Evol.* **1999**, *16*, 1427–1438. [CrossRef] [PubMed]
27. Volff, J.N.; Körting, C.; Schartl, M. Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. *Mol. Biol. Evol.* **2000**, *17*, 1673–1684. [CrossRef]
28. Volff, J.N.; Körting, C.; Froschauer, A.; Sweeney, K.; Schartl, M. Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. *J. Mol. Evol.* **2001**, *52*, 351–360. [CrossRef]
29. Carducci, F.; Barucca, M.; Canapa, A.; Biscotti, M.A. Rex retroelements and teleost genomes: An overview. *Int. J. Mol. Sci.* **2018**, *19*, 3653. [CrossRef]
30. Dechaud, C.; Volff, J.-N.; Schartl, M.; Naville, M. Sex and the TEs: Transposable elements in sexual development and function in animals. *Mob. DNA* **2019**, *10*, 1–15. [CrossRef]
31. Furman, B.L.S.; Metzger, D.C.H.; Darolti, I.; Wright, A.E.; Sandkam, B.A.; Almeida, P.; Shu, J.J.; E Mank, J. Sex chromosome evolution: So many exceptions to the rules. *Genome Biol. Evol.* **2020**, *12*, 750–763. [CrossRef]
32. Fujino, K.; Hashida, S.-N.; Ogawa, T.; Natsume, T.; Uchiyama, T.; Mikami, T.; Kishima, Y. Temperature controls nuclear import of Tam3 transposase in Antirrhinum. *Plant. J.* **2010**, *65*, 146–155. [CrossRef]
33. Makarevitch, I.; Waters, A.J.; West, P.T.; Stitzer, M.; Hirsch, C.N.; Ross-Ibarra, J.; Springer, N.M. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* **2015**, *11*, e1004915.
34. Feschotte, C.E.D.; Jiang, N.; Wessler, S.R. Plant transposable elements: Where genetics meets genomics. *Nat. Rev. Genet.* **2002**, *3*, 329–341. [CrossRef] [PubMed]
35. Sahebi, M.; Hanafi, M.M.; van Wijnen, A.J.; Rice, D.; Rafii, M.Y.; Azizi, P.; Osman, M.; Taheri, S.; Bakar, M.F.A.; Isa, M.N.M.; et al. Contribution of transposable elements in the plant's genome. *Gene* **2018**, *665*, 155–166. [CrossRef] [PubMed]
36. Quadrana, L. The contribution of transposable elements to transcriptional novelty in plants: The FLC affair. *Transcription* **2020**, *11*, 192–198. [CrossRef] [PubMed]
37. Casacuberta, E.; González, J. The impact of transposable elements in environmental adaptation. *Mol. Ecol.* **2013**, *22*, 1503–1517. [CrossRef]
38. Pimpinelli, S.; Piacentini, L. Environmental change and the evolution of genomes: Transposable elements as translators of phenotypic plasticity into genotypic variability. *Funct. Ecol.* **2019**, *34*, 428–441. [CrossRef]
39. Carducci, F.; Biscotti, M.A.; Forconi, M.; Barucca, M.; Canapa, A. An intriguing relationship between teleost Rex3 retroelement and environmental temperature. *Biol. Lett.* **2019**, *15*, 20190279. [CrossRef]
40. Yuan, Z.; Liu, S.; Zhou, T.; Tian, C.; Bao, L.; Dunham, R.A.; Liu, Z. Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. *BMC Genom.* **2018**, *19*, 1–10. [CrossRef]
41. Liu, D.; Yang, J.-Q.; Tang, W.-Q.; Zhang, X.; Royster, C.M.; Zhang, M. SINE Retrotransposon variation drives Ecotypic disparity in natural populations of Coilia nasus. *Mob. DNA* **2020**, *11*, 4–14. [CrossRef]
42. McDowall, R. The evolution of diadromy in fishes (revisited) and its place in phylogenetic analysis. *Rev. Fish. Biol. Fish.* **1997**, *7*, 443–462. [CrossRef]
43. McDowall, R.M. On amphidromy, a distinct form of diadromy in aquatic organisms. *Fish. Fish.* **2007**, *8*, 1–13. [CrossRef]
44. Betancur, R.R.; Wiley, E.O.; Arratia, G.; Acero, A.; Bailly, N.; Miya, M.; Lecointre, G.; Ortí, G. Phylogenetic classification of bony fishes. *BMC Evol. Biol.* **2017**, *17*, 1–40. [CrossRef] [PubMed]
45. Legendre, P. Studying beta diversity: Ecological variation partitioning by multiple regression and canonical analysis. *J. Plant. Ecol.* **2007**, *1*, 3–8. [CrossRef]
46. Metcalfe, C.J.; Filée, J.; Germon, I.; Joss, J.; Casane, D. Evolution of the Australian lungfish (Neoceratodus forsteri) genome: A major role for CR1 and L2 LINE elements. *Mol. Biol. Evol.* **2012**, *29*, 3529–3539. [CrossRef] [PubMed]
47. Lynch, M.; Walsh, B. *The Origins of Genome Architecture*; Sinauer Associates: Sunderland, MA, USA, 2007.
48. Venkatesh, B.; Lee, A.P.; Ravi, V.; Maurya, A.K.; Lian, M.M.; Swann, J.B.; Ohta, Y.; Flajnik, M.F.; Sutoh, Y.; Kasahara, M. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **2014**, *505*, 174–179. [CrossRef] [PubMed]
49. Amemiya, C.; Alfoeldi, J.; Lee, A.P.; Fan, S.; Philippe, H.; Maccallum, I.; Braasch, I.; Manousaki, T.; Schneider, I.; Rohner, N.; et al. The African coelacanth genome provides insights into tetrapod evolution. *Nat. Cell Biol.* **2013**, *496*, 311–316. [CrossRef]
50. Elliot, T.A.; Gregory, T.R. Do larger genomes contain more diverse transposable elements? *BMC Evol. Biol.* **2015**, *15*, 69. [CrossRef]
51. Smith, J.J.; Kuraku, S.; Holt, C.; Sauka-Spengler, T.; Jiang, N.; Campbell, M.S.; Li, W. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insightinto vertebrate evolution. *Nat. Genet.* **2013**, *45*, 415–421. [CrossRef]
52. Dehal, P.; Boore, J.L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **2005**, *3*, e314. [CrossRef]
53. Postlethwait, J.H.; Woods, I.G.; Ngo-Hazelett, P.; Yan, Y.-L.; Kelly, P.D.; Chu, F.; Huang, H.; Hill-Force, A.; Talbot, W.S. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res.* **2000**, *10*, 1890–1902. [CrossRef]
54. Vandepoele, K.; De Vos, W.; Taylor, J.S.; Meyer, A.; Van De Peer, Y. 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* **2004**, *101*, 1638–1643. [CrossRef] [PubMed]

55. Christoffels, A.; Koh, E.G.L.; Chia, J.-M.; Brenner, S.; Aparicio, S.; Venkatesh, B. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.* **2004**, *21*, 1146–1151. [CrossRef] [PubMed]
56. Meyer, A.; Van de Peer, Y. From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). *BioEssays* **2005**, *27*, 937–945. [CrossRef] [PubMed]
57. Du, K.; Stöck, M.; Kneitz, S.; Klopp, C.; Woltering, J.M.; Adolfi, M.C.; Feron, R.; Prokopov, D.; Makunin, A.; Kichigin, I.; et al. The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat. Ecol. Evol.* **2020**, *4*, 841–852. [CrossRef]
58. Schmidt, D.; Schwalie, P.C.; Wilson, M.D.; Ballester, B.; Gonçalves, Â.; Kutter, C.; Brown, G.D.; Marshall, A.; Flicek, P.; Odom, D.T. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **2012**, *148*, 335–348. [CrossRef] [PubMed]
59. Crepaldi, L.; Policarpi, C.; Coatti, A.; Sherlock, W.T.; Jongbloets, B.C.; Down, T.A.; Riccio, A. Binding of TFIIIC to SINE elements controls the relocation of activity-dependent neuronal genes to transcription factories. *PLoS Genet.* **2013**, *9*, e1003699. [CrossRef]
60. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **2017**, *18*, 71–86. [CrossRef]
61. Boissinot, S.; Bourgeois, Y.; Manthey, J.D.; Ruggiero, R.P. The mobilome of reptiles: Evolution, structure, and function. *Cytogenet. Genome Res.* **2019**, *157*, 21–33. [CrossRef]
62. Wallau, G.L.; Ortiz, M.F.; Loreto, E.L.S. Horizontal transposon transfer in Eukarya: Detection, bias, and perspectives. *Genome Biol. Evol.* **2012**, *4*, 801–811. [CrossRef]
63. Bourgeois, Y.; Boissinot, S. On the population dynamics of junk: A review on the population genomics of transposable elements. *Genes* **2019**, *10*, 419. [CrossRef]
64. Price, A.L.; Jones, N.C.; Pevzner, P.A. De novo identification of repeat families in large genomes. *Bioinformatics* **2005**, *21*, i351–i358. [CrossRef] [PubMed]
65. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
66. The UniProt Consortium UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef] [PubMed]
67. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [CrossRef] [PubMed]
68. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef]
69. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2018**, *47*, D427–D432. [CrossRef]
70. Oksanen, J.; Guillaume Blanchet, F.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; et al. Vegan: Community ecology package. R Package Version 2.5–5. 2019. Available online: https://CRAN.R-project.org/package=vegan (accessed on 10 May 2020).

3.5 Transcriptional contribution of transposable elements in relation to salinity conditions in teleosts and silencing mechanisms involved

International Journal of
Molecular Sciences

MDPI

Article

Transcriptional Contribution of Transposable Elements in Relation to Salinity Conditions in Teleosts and Silencing Mechanisms Involved

Elisa Carotti [1,†], Federica Carducci [1,†], Samuele Greco [2], Marco Gerdol [2], Daniele Di Marino [1,3], Nunzio Perta [1,3], Anna La Teana [1,3], Adriana Canapa [1], Marco Barucca [1,*] and Maria Assunta Biscotti [1]

1 Dipartimento di Scienze della Vita e Dell'Ambiente, Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy; e.carotti@pm.univpm.it (E.C.); f.carducci@univpm.it (F.C.); d.dimarino@univpm.it (D.D.M.); nunzioperta95@gmail.com (N.P.); a.lateana@univpm.it (A.L.T.); a.canapa@univpm.it (A.C.); m.a.biscotti@univpm.it (M.A.B.)
2 Dipartimento di Scienze della Vita, Università degli Studi di Trieste, Via L. Giorgieri, 5, 34127 Trieste, Italy; samuele.greco@phd.units.it (S.G.); mgerdol@units.it (M.G.)
3 New York-Marche Structural Biology Center (NY-MaSBiC), Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy
* Correspondence: m.barucca@univpm.it
† These authors contributed equally to this work.

Abstract: Fish are an interesting taxon comprising species adapted to a wide range of environments. In this work, we analyzed the transcriptional contribution of transposable elements (TEs) in the gill transcriptomes of three fish species exposed to different salinity conditions. We considered the giant marbled eel *Anguilla marmorata* and the chum salmon *Oncorhynchus keta*, both diadromous, and the marine medaka *Oryzias melastigma*, an euryhaline organism sensu stricto. Our analyses revealed an interesting activity of TEs in the case of juvenile eels, commonly adapted to salty water, when exposed to brackish and freshwater conditions. Moreover, the expression assessment of genes involved in TE silencing mechanisms (six in heterochromatin formation, fourteen known to be part of the nucleosome remodeling deacetylase (NuRD) complex, and four of the *Argonaute* subfamily) unveiled that they are active. Finally, our results evidenced for the first time a krüppel-associated box (KRAB)-like domain specific to actinopterygians that, together with TRIM33, might allow the functioning of NuRD complex also in fish species. The possible interaction between these two proteins was supported by structural prediction analyses.

Keywords: transposable elements; fish; silencing mechanisms; salinity; Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs)

Citation: Carotti, E.; Carducci, F.; Greco, S.; Gerdol, M.; Di Marino, D.; Perta, N.; La Teana, A.; Canapa, A.; Barucca, M.; Biscotti, M.A. Transcriptional Contribution of Transposable Elements in Relation to Salinity Conditions in Teleosts and Silencing Mechanisms Involved. *Int. J. Mol. Sci.* **2022**, *23*, 5215. https://doi.org/10.3390/ijms23095215

Academic Editor: Gail B. Mahady

Received: 13 April 2022
Accepted: 6 May 2022
Published: 6 May 2022

1. Introduction

A considerable portion of eukaryote genomes is composed of transposable elements (TEs). They are able to move within genomes through transposition using either a DNA intermediate molecule, for example DNA transposons, or an RNA intermediate molecule, known as retrotransposons, that includes Long Terminal Repeats (LTRs) and non LTR elements. These latter comprise Long Interspersed Nuclear Elements (LINEs) and Short Interspersed Nuclear Elements (SINEs).

It is known that TEs can affect genome size, can be co-opted to create novel genes, or be involved in chromosome rearrangements. Not negligible is also the increasing number of evidence reporting a transposition activity that seems to be influenced by abiotic factors such as temperature, salinity, and pH [1–8], suggesting a role of these genetic elements in the regulation of mechanisms responsible for environmental adaptation [2,8–11]. Moreover, the presence of a specific TE class [10,12] as well as of a particular TE sequence variant [9] has been proposed to be related with species adaptation to a specific environment.

Salinity is an abiotic factor that influences the adaptation of marine and freshwater organisms, and its changes can also represent a possible stress that threatens their survival. Teleosts are an excellent model to investigate the relationship between this environmental parameter and TEs. Indeed, this highly diverse group populates a wide range of habitats across the world, from tropical to polar regions of both sea water (SW) and freshwater (FW) and some species are characterized by the ability to migrate during their lifespan. This fascinating behavior can occur in SW (oceanodromous), in FW (potamodromous) or between these two environments (diadromous). In the latter case, the migratory behavior is carried out for reproductive purposes and organisms have to face, in a defined stage of their life cycle, variations in salinity that require changes in osmotic regulation. Among fish, this aspect is better tolerated by euryhaline species sensu stricto that have acquired the ability to adapt to various salinity conditions.

The giant marbled eel *Anguilla marmorata* is a catadromous species widely distributed in tropical, subtropical, and temperate areas that moves from freshwater to salt water for spawning [13]. The chum salmon *Oncorhynchus keta* is an anadromous species that spends much of its life in salt water and goes up rivers to spawn [14,15]. The marine medaka *Oryzias melastigma* is an euryhaline species able to survive in environments with different salinities.

In this study, we compared the activity of TEs between gill tissues of the considered species analyzing RNA-Seq data available in the public databases obtained from organisms treated at different salinity conditions, FW, brackish water (BW), and SW. Gills are primary organs that are highly sensitive to salinity level, can detect changes in external osmotic pressure, and promote ionic compensatory mechanisms to maintain the osmolarity in body fluids. Moreover, they play crucial roles in physical processes such as gas exchange, nitrogenous waste excretion, and acid-base balance [14,16].

Interestingly, our findings evidenced a variation in TE transcriptional contribution in the case of the giant marbled eel with respect to *O. melastigma* and *O. keta*, and thus the transcriptional activity of genes involved in TE silencing mechanisms was investigated. It is known that to counteract the negative effects due to transposition, host genomes evolved various TE silencing mechanisms involving small RNAs, chromatin and DNA modification pathways, and sequence-specific repressors such as those based on the KRAB-ZFPs recruiting the NuRD complex [11,17]. Here, the evaluation of the expression of genes involved in these mechanisms (six in heterochromatin formation, fourteen known to be part of the NuRD complex, and four of the *Argonaute* subfamily) unveiled that they are active. Intriguingly, our results evidenced for the first time a KRAB-like domain specific of actinopterygians that together with TRIM33 might allow the functioning of NuRD complex.

2. Results

2.1. Transcriptional Contribution of Transposable Elements in Gill Transcriptomes of A. marmorata, O. melastigma, and O. keta

The transcriptional activity of TEs, assessed as percentage of mapped reads, was evaluated through the analysis of gill RNA-Seq data available for *A. marmorata*, *O. melastigma*, and *O. keta* treated in FW, BW, and SW conditions (Figure 1).

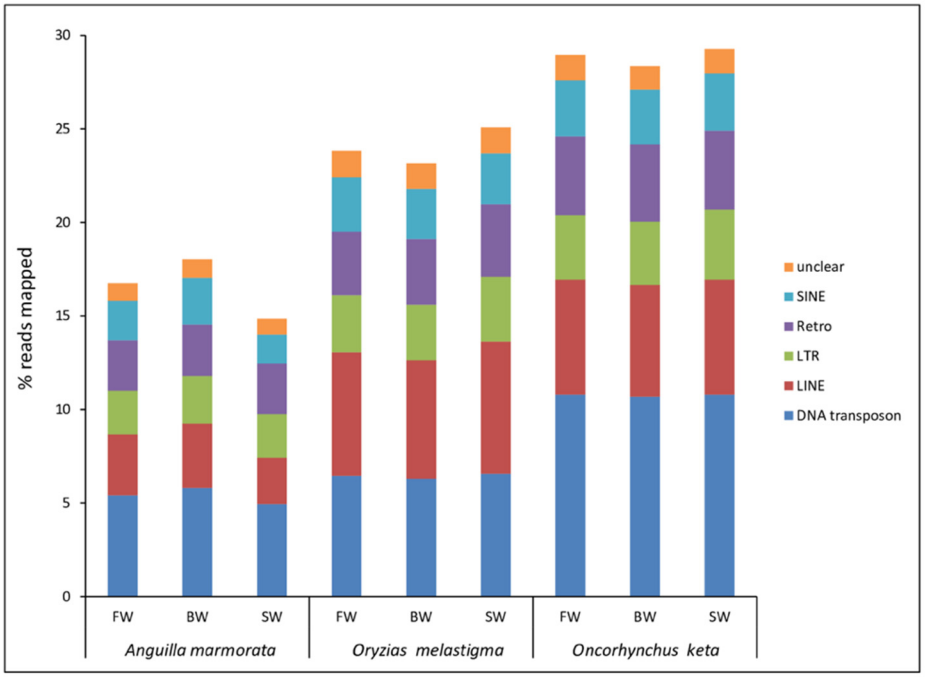

Figure 1. Transcriptional contribution of transposable elements in *Anguilla marmorata, Oryzias melastigma,* and *Oncorhynchus keta* gill transcriptomes. FW: freshwater; BW: brackish water; SW: salt water.

The total TE contribution was higher in *O. keta*, followed by that of *O. melastigma* and *A. marmorata*. These differences were mainly ascribable to the activity of LINE retroelements and DNA transposons. In *A. marmorata* and *O. keta*, the highest impact was due to DNA transposons while in *O. melastigma* the two major TE types were DNA transposons and LINE retroelements. The comparison of TE transcriptional contribution showed similar levels between the three tested conditions in *O. keta* (maximum percentage variation of total TEs was 3.22 in the SW vs. BW comparison) and *O. melastigma* (maximum percentage variation of total TEs was 8.31 in the SW vs. BW comparison), while in *A. marmorata* a lower level was detected in the SW condition compared to FW (the increase of percentage variation of total TEs was 12.79) and to BW (the increase of percentage variation of total TEs was 21.43). Regarding percentage variation of single TE class, in *A. marmorata* the main difference was ascribable to SINE retroelements that increased 37.09% in the comparison SW vs. FW and 63.31% in the SW vs. BW comparison. For the other two species considered, minimum percentage variations were reported for the single TE class (Table S1).

2.2. Identification and Transcriptional Activity of Genes Involved in TE Silencing Mechanisms

TE silencing mechanisms were investigated in order to explain the different TE contribution observed for the SW condition in the giant marbled eel. The transcriptomes of *A. marmorata*, *O. melastigma*, and *O. keta* were screened to retrieve transcripts with coding sequence (CDS) orthologous to *AGO1, AGO2, AGO3,* and *AGO4* genes of the *Ago* subfamily, to genes encoding proteins involved in heterochromatin formation (*HP1α*, *HP1βa*, *HP1βb*, *HP1γ*, *DNMT1*, *DNMT3Aa*), and in NuRD complex (*CHD3*, *CHD4α*, *CHD4β*, *HDAC1*, *HDAC2*, *MBD2*, *MBD3b*, *MTA1*, *MTA2*, *MTA3*, *GATAD2A*, *GATAD2B*, *RBBP4*, and *RBBP7*) (Table S2).

Four transcripts encoding the Ago proteins were retrieved in the analyzed species. *AGO4* of *A. marmorata*, *AGO3* of *O. keta*, and *AGO1*, *AGO2*, and *AGO3* of *O. melastigma* showed a complete CDS. The evaluation of the transcriptional activity of the four *AGO* genes evidenced the same trend between the three tested conditions in chum salmon and in marine medaka, while in giant marbled eel the expression of *AGO* genes was variable between FW, BW, and SW conditions (Figure 2). Overall, *AGO4* showed the highest

transcriptional value in all conditions of analyzed species. However, in the case of eel, a decrease in the expression level of this gene was observable from FW to SW condition (Figure 2).

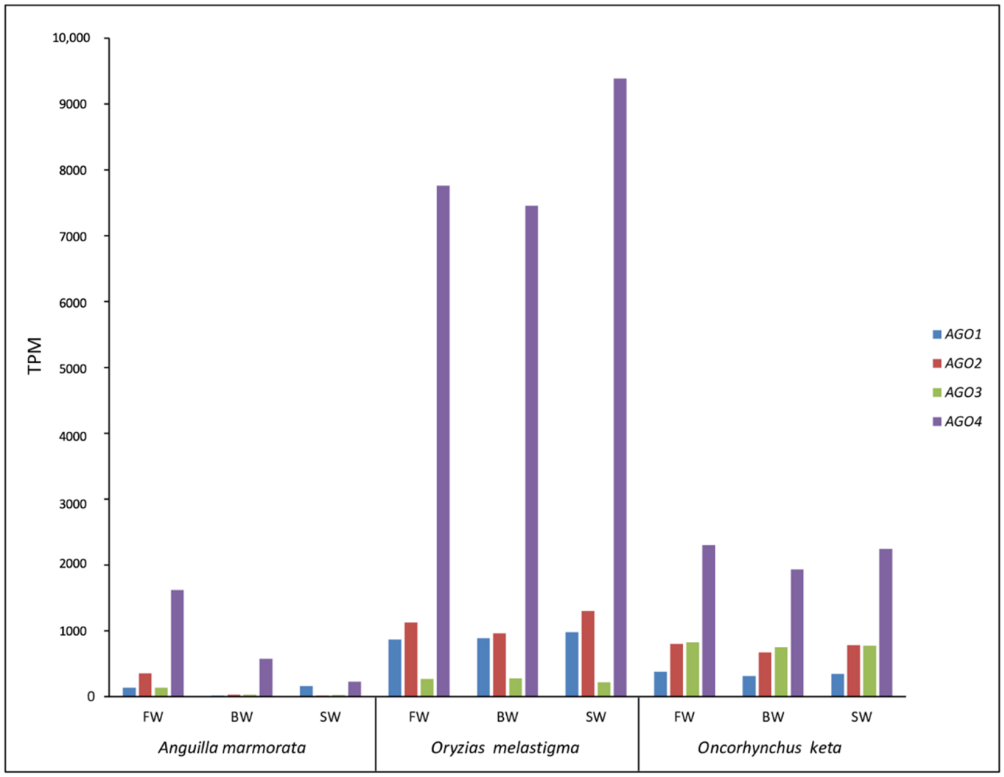

Figure 2. Transcriptional activity of *Argonaute* genes in *Anguilla marmorata*, *Oryzias melastigma*, and *Oncorhynchus keta* gill transcriptomes. FW: freshwater; BW: brackish water; SW: salt water.

Concerning genes related to heterochromatin formation, *HP1α* transcript was identified in *O. keta* and *O. melastigma* with a complete CDS; for *HP1β*, two transcripts corresponding to *HP1βa* and *HP1βb* were retrieved in *A. marmorata* and *O. melastigma*, and one transcript corresponding to *HP1βb* was identified in *O. keta*. All *HP1β* sequences presented a complete CDS. For *HP1γ*, a transcript was retrieved in all species (Table S2). This gene was the most expressed in *A. marmorata* and *O. keta*. In the case of *O. melastigma*, *HP1βa* showed the highest transcriptional levels. The *HP1α* was the second gene having a remarkable transcriptional activity in *O. melastigma* and in *O. keta* (Figure 3). A sequence homologous to *DNMT1* was retrieved in all considered species while *DNMT3Aa* was identified in chum salmon and marine medaka transcriptomes. The expression of *DNMT1* was appreciable in all analyzed RNA-Seq data. Moreover, in *O. melastigma*, among the DNA methyltransferases, the *DNMT3Aa* showed higher transcriptional levels than *DNMT1* (Figure 3).

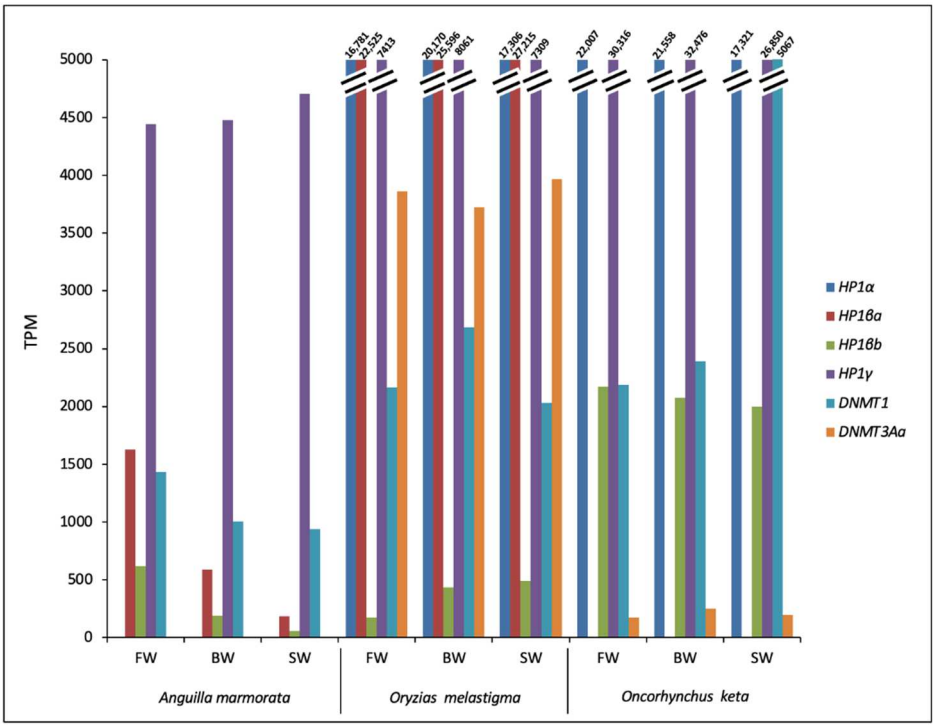

Figure 3. Transcriptional activity of genes involved in heterochromatin formation in *Anguilla marmorata, Oryzias melastigma,* and *Oncorhynchus keta* gill transcriptomes. FW: freshwater; BW: brackish water; SW: salt water.

Interestingly, for the NuRD complex, the whole set of genes was retrieved in all species investigated. The identified transcripts showed a complete CDS for *HDAC1*, *MBD3b*, *MTA1*, *RBBP4*, and *RBBP7* in *A. marmorata*, for *HDAC2*, *MBD2*, *MBD3b*, *MTA1*, *MTA2*, *RBBP4*, and *RBBP7* in *O. keta*, and for all genes of NuRD complex with the exception of *CHD3* and *CHD4* in *O. melastigma* (Table S2). The expression analysis of the NuRD complex genes showed that they are active in all the three species considered (Figure 4). Moreover, many of them did not show considerable expression variations between the three tested conditions for *O. keta* as well as those for *O. melastigma*. In the case of *A. marmorata*, a different trend was observed between the three conditions with an overall increase in gene expression levels in BW and FW.

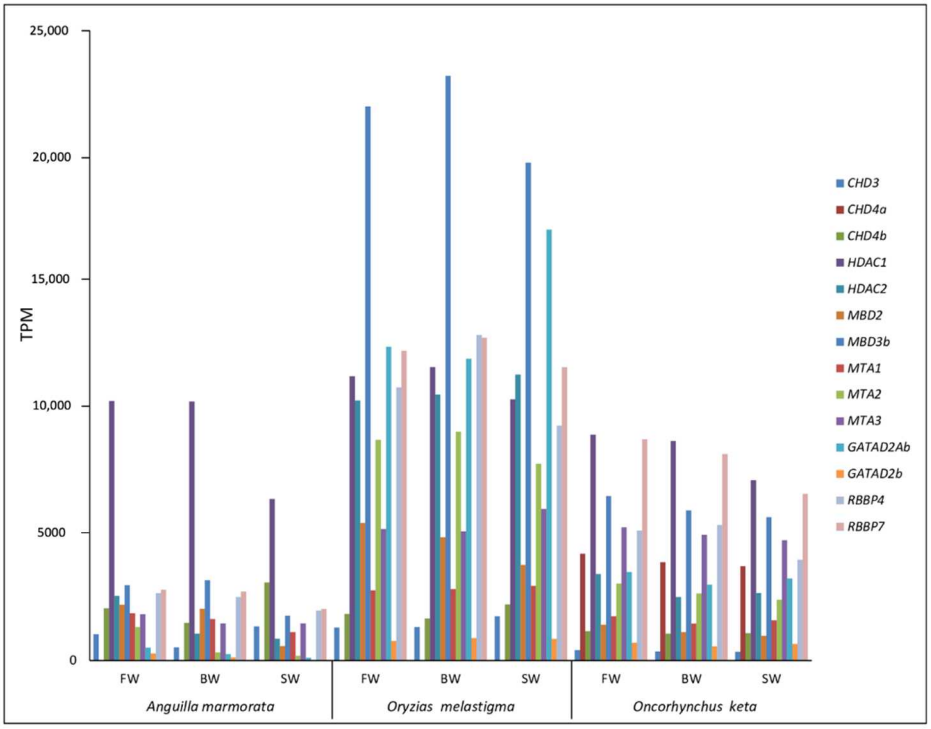

Figure 4. Transcriptional activity of NuRD complex genes in *Anguilla marmorata*, *Oryzias melastigma*, and *Oncorhynchus keta* gill transcriptomes. FW: freshwater; BW: brackish water; SW: salt water.

In sarcopterygians, the NuRD complex is recruited at TE sequences through the involvement of KRAB-ZFPs and TRIM28, a protein belonging to the Tripartite Motif family. The former binds the TE sequence through their C-terminal ZF domains and the TRIM28 through their N-terminal KRAB domain, which in turn serves as a scaffold to recruit proteins of the NuRD complex. In actinopterygians, both KRAB-ZFPs and TRIM28 are not present. Therefore, we searched for proteins playing analogous functions in fish species here analyzed. TRIM33 is a protein that has the same domain architecture as TRIM28, being composed of an N-terminal RING Finger/B-Box/coiled coil (RBCC), the Plant Homeo Domain (PHD), and the Bromodomain (BROMO) [18] and is widely spread among ray-finned fish (Figure 5A). The *TRIM33* transcripts showed a complete CDS only for *O. melastigma* (Table S2). Interestingly, our findings highlighted for *TRIM33* the same trend reported for NuRD complex related genes (Figure 6).

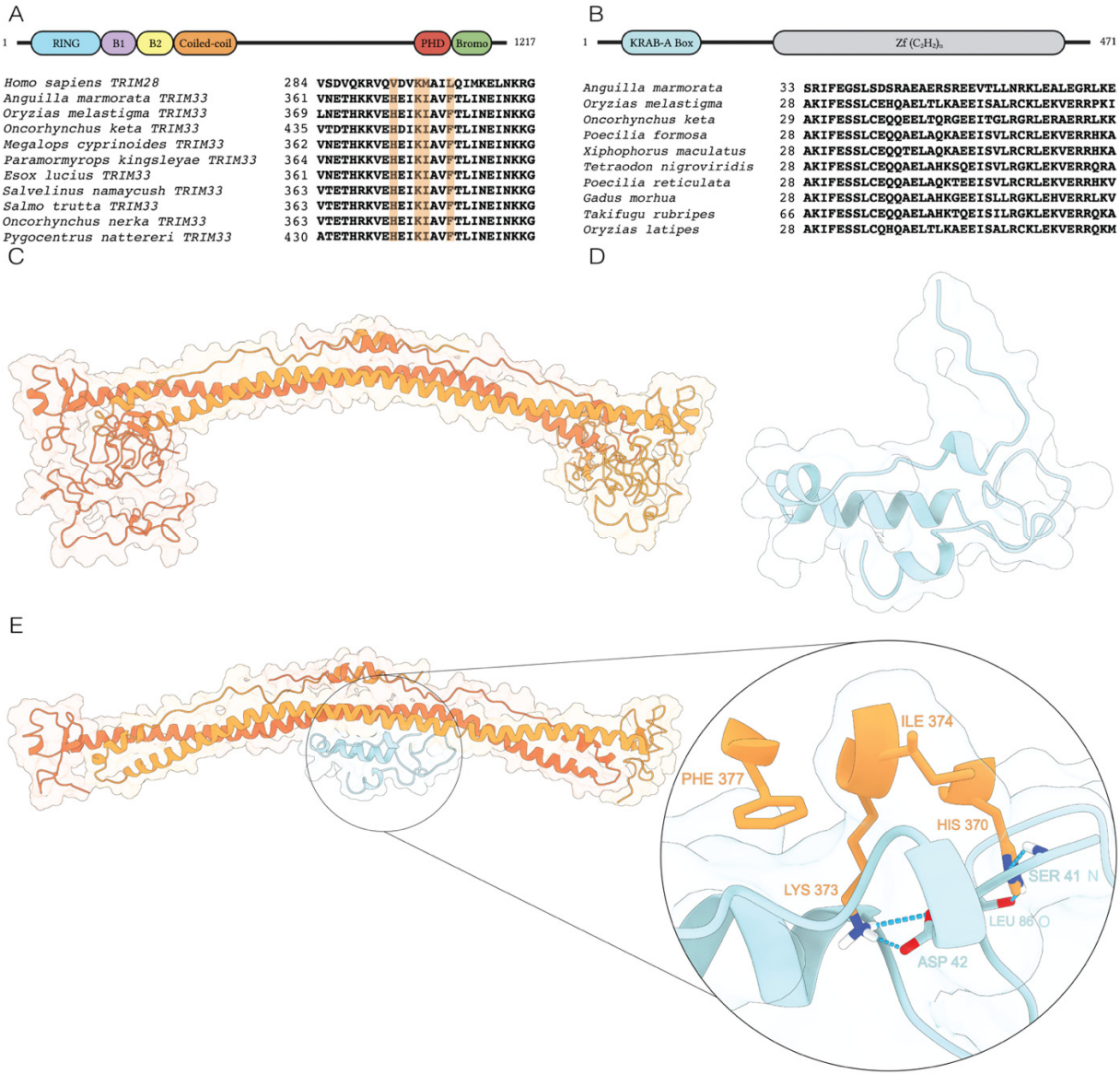

Figure 5. Multiple sequence alignment and 3D structures of the KRAB domain contained in the KRAB-like protein and that of TRIM33 protein. (**A**) In the upper side, a schematic representation of the domain architecture is reported for the *A. marmorata* TRIM33 protein. In the lower side, multiple sequence alignment related to the coiled-coil region of *Homo sapiens* TRIM28 protein and of ten actinopterygian TRIM33 sequences is showed. The main interacting residues at the interface are highlighted in orange. (**B**) In the upper side, a schematic representation of the domain architecture is reported for the *Anguilla marmorata* KRAB-like protein. In the lower side, the N-terminal region multiple sequence alignment of ten actinopterygian KRAB-like sequences is reported. (**C**) Ribbon and surface representation of *A. marmorata* TRIM33 structural model. Chain A and chain B are colored in dark and light orange, respectively. (**D**) Ribbon and surface representation of the putative *A. marmorata* KRAB-like domain. (**E**) Ribbon and surface representation of docked TRIM33/KRAB-like complex with a zoom on the residues H370/K373/I374/F377.

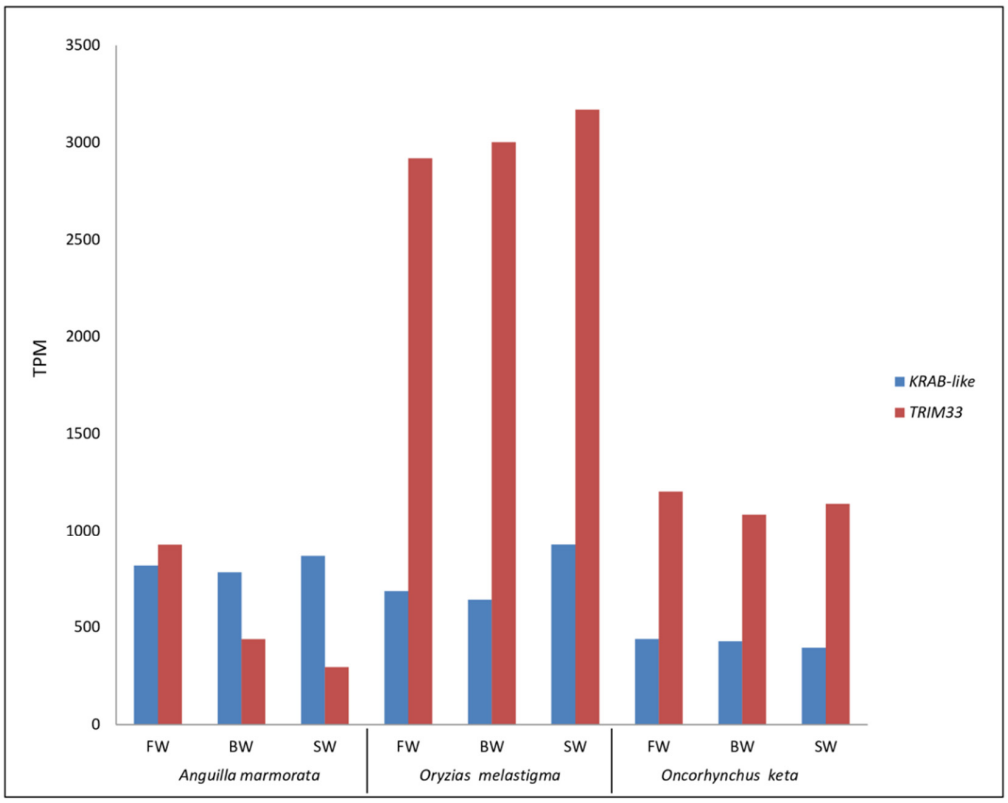

Figure 6. Transcriptional activity of *KRAB*-like and *TRIM33* genes in *Anguilla marmorata*, *Oryzias melastigma*, and *Oncorhynchus keta* gill transcriptomes. FW: freshwater; BW: brackish water; SW: salt water.

Moreover, in the transcriptomes of the considered species, we also searched sequences with a region similar to KRAB domain and ZF motifs. This analysis allowed us to identify a sequence named *KRAB-like* showing these features and widely spread in ray-finned fish (Figure 5B). Its transcriptional expression levels were uniform between the three tested conditions in the two diadromous species, different from *O. melastigma* that showed a higher activity in SW condition (Figure 6).

2.3. Structural Prediction of the TRIM33/KRAB-like Complex

To understand the nature of interactions between TRIM33 and KRAB-like proteins we applied a series of structural bioinformatics methodologies. Since there is a lack of structural information available for actinopterygians, TRIM33 was modelled on the human TRIM28 (Figure 5C) while the KRAB-like sequence was modelled on the human KRAB-ZNF93 (Figure 5D). In the wide range of the human ZNFs, we selected ZNF93, since this was identified by Stoll et al. (2019) [19] to strictly bind TRIM28.

We predicted the 3D assembly of the actinopterygian TRIM33/KRAB-like complex (Figure 5E) and used the HADDOCK score to evaluate the poses obtained, it being a valuable parameter of the strength of protein–protein interaction [20,21]. In this case, the HADDOCK score for the best predicted complex was −17.5 +/− 11 with a Z-score of −2.3 for the best prediction. To have a more robust structural reference for the predicted TRIM33/KRAB-like complex, we also built a 3D structure of the *H. sapiens* TRIM28/KRAB-ZNF93 complex.

The superimposition between the human TRIM28/KRAB-ZNF93 and the actinopterygian TRIM33/KRAB-like complexes showed a comparable assembly between both complexes (Figure S1A). Interestingly, the interacting residues of both TRIM28 and TRIM33 were arranged in a similar mode interacting with the cleft of the KRAB domain of KRAB-ZNF93 and KRAB-like, respectively (Figure S1B).

The putative amino acids directly involved in the stabilization of the complex were obtained by combining the results obtained from analyzing the multiple sequence alignment (Figure 5A) between the sequence of human TRIM28 and a set of other actinopterygian TRIM33 sequences and the structural information provided by Stoll et al., (2019) [19]. Indeed, starting from the human TRIM28, we identified the residues H370/K373/I374/F377 for TRIM33. Notably, the H370 and the K373 were oriented into the cleft formed by the KRAB domain helices, where these residues can establish crucial non-covalent interactions (Figure 5E). Moreover, the K373 showed a peculiar role in the TRIM33/KRAB-like complex stabilization since this positively charged residue established an electrostatic interaction (i.e., salt bridge) with D42 (Figure 5E, inset). In order to better characterize the role of K373, we performed a multiple sequence alignment using fifty-one actinopterygian and ten sarcopterygian sequences (Figure S2) to evaluate the degree of conservation of such residue. We found that the K373 was a highly conserved residue both in actinopterygian TRIM33 and sarcopterygian TRIM28.

In addition, we identified other conserved residues from the multiple sequence alignment (highlighted in green in Figure S2), and their positions on the structural model of the complex were shown in Figure S3A,B (i.e., TRIM33 residues: K359/E363/K366/I374/N384/K385/G387/K388; TRIM28 residues: R282/D286/K289/M297/N307/K308/G310/R311).

Altogether these data corroborate our initial hypothesis regarding the ability of TRIM33 and KRAB-like actinopterygian proteins to assemble in a complex comparable to that of humans and thus play a similar biological function.

3. Discussion

In this paper, the TE activity was investigated in three fish species showing different salinity tolerance. *A. marmorata*, a catadromous species, and *O. keta*, an anadromous species, migrate for reproduction and thus they have to face changes in salinity in a defined stage of their life cycle. Contrarily, *O. melastigma* is an euryhaline species sensu stricto. Species belonging to *Oryzias* and *Anguilla* genera present a comparable genome size (0.9 pg/N and 1.3 pg/N, respectively) with a TE amount of 40% and 15%, respectively; while species of the *Oncorhynchus* genus show a genome size that is approximately twofold (2.6 pg/N), probably due to the genome duplication event that occurs in salmonids, and TEs occupy a large fraction of their genome with about 40% [10]. Therefore, differences emerging from the comparison of percentage of total TE mapped reads, obtained from the gill transcriptomes of the analyzed species, can be ascribable to their genome size and TE amount. Analyzing the three tested conditions in each species, the lack of remarkable variations in TE contribution in *O. melastigma* might be related to its strong ability to survive in aquatic environments with a wide range of salinity. *O. keta* specimens selected for this study were at the fry stage and thus adapted to freshwater. Therefore, it was interesting to investigate if the exposure at higher salinity concentrations caused changes in TE transcription. However, our findings showed no appreciable differences in TE contribution between the three tested conditions in this species. The lack of this evidence might be attributable to short periods of exposure [14], not sufficient to provoke a TE transcriptional response. An appreciable difference in percentage of TE mapped reads emerged comparing the SW with FW and BW conditions of *A. marmorata*. Since specimens used for this experiment were at the juvenile stage and thus adapted to living in seawater [16], the exposure to low salinity levels might have determined the increase in TE transcription observed. Moreover, the major impact in these changes was due to the expression of SINE retroelements (37.09% for SW vs. FW, 63.31% for SW vs. BW). This finding could be in line also with our previous results that highlighted a role of these kinds of elements in the catadromous behavior of eels [10]. Indeed, it has been reported that SINE retroelements affect gene expression since they are located at the boundaries of transcriptionally active or inactive domains favoring intra- or inter-chromosomal interactions [22–24].

The transcriptional activation of TEs is strictly related to silencing mechanisms [11,25–27]. Therefore, the activity of genes encoding proteins involved in TE repression was inves-

tigated. It is noteworthy that no appreciable differences were detected in the expression levels of genes belonging to the *Ago* subfamily between the analyzed conditions in chum salmon and marine medaka. In giant marbled eel, the expression of these genes was variable and, in particular, that of *AGO4* showed a decreasing level from FW to SW. This gene and *AGO2* showed an expression trend similar to that of TEs. Interestingly, *AGO4* has been shown to play a role in transposon silencing in gonads [28] and our results suggested that this function might also be conserved in somatic tissues. Moreover, a slicing activity for *AGO2* has been reported and the endosiRNAs derived seem to be involved in the transposon repression [29].

Among heterochromatinization related genes, *HP1βb* and *HP1γ* were transcriptionally active in all species and tested conditions. The high expression levels of *HP1γ* in *O. keta* might explain the low difference detected in total TE contribution levels, compared to the other two species. Indeed, a higher TE activity was expected in chum salmon due to the high number of mobile elements present in its genome. Between the two genes selected for *DNMTs*, the expression analysis suggested that DNMT1 might be the candidate protein in TE methylation.

The expression analysis of genes encoding proteins of NuRD complex showed that this system is active in the fish species investigated. To date, to our knowledge, the activity of these genes in fish has been reported in the blastema in fin regeneration [30]. The involvement of NuRD complex in TE silencing in tetrapods has been investigated at the embryonic stage [31–34]. However, our data obtained from gills of juveniles and adults were in agreement with recent works showing the expression of NuRD complex genes also in adult tissues [35,36]. Overall, the expression of genes investigated did not show remarkable variations between the three tested conditions both in *O. melastigma* and *O. keta* while a major variability was detected in *A. marmorata*. This trend reflected that obtained for TE transcription, suggesting a possible relationship between NuRD complex and TEs also in fish. In tetrapods, the NuRD complex is recruited at TE sequence through the involvement of KRAB-ZFPs and TRIM28. In fish lineage, these components have not been identified. For this reason, we have investigated the possibility that TRIM33 and our KRAB-like could play the same function of TRIM28 and KRAB-ZFPs of tetrapods. Indeed, TRIM33 is a protein belonging to the Tripartite Motif family, it shows the same domain architecture of TRIM28 and is spread among ray-finned fish [18]. Recently, Helleboid and colleagues (2019) [37], analyzing the interactome of KRAB-ZFPs in human, have evidenced that ZFPs might interact with other members of the TRIM family. Intriguingly, our analysis revealed an increased transcription of *TRIM33* with salinity decrease in giant marbled eel, following the same trend observed for genes involved in NuRD complex. This finding supported our hypothesis that in fish TRIM33 might play a role similar to that of tetrapod TRIM28. Moreover, the KRAB-like sequence here reported showed a KRAB-like domain at the N terminus and several Zinc Finger motifs at the C terminus. This protein domain architecture is similar to that of tetrapod KRAB-ZFPs. The docking analysis supported the ability of TRIM33 and KRAB-like proteins to assemble in a complex in actinopterygians, comparable to that of human TRIM28/KRAB-ZFP93.

4. Materials and Methods

RNA-Seq raw data were obtained from the public database NCBI GenBank (https://www.ncbi.nlm.nih.gov/ accessed on 21 September 2021) deposited under the accession number GSE95803 for *A. marmorata* [16], in Sequence Read Archive (SRA) under the accession numbers SRX3932910-SRX3932912 for *O. keta* [14], and under the BioProject ID PRJNA745044 for *O. melastigma* [38]. Data were obtained from juvenile specimens in the case of *A. marmorata* and *O. keta* and from adult specimens in the case of *O. melastigma*. Exposures were conducted at FW, BW, and SW (Table S3). For each condition, available data were referred to pooled biological replicates [14,16,38]. Raw paired-end reads were imported in the CLC Genomics Workbench v.12 (Qiagen, Hilden, Germany) and trimmed with the proper internal tool by removing sequencing adapters, low quality bases, and low

quality read ends using default parameters. Trimmed reads were then de novo assembled with the proper tool within the CLC using default parameters. Completeness of the de novo assembled transcriptomes was evaluated through BUSCO v5 using the Actinopterygii OrthoDB v10 database as reference [39].

4.1. Estimation of TE Activity

To estimate the TE transcriptional activity, we first identified TEs in the de novo assembled transcriptomes with RepeatMasker v.4.1.0 using a custom library created following the methodology described in our previous work [10]. After RepeatMasker analysis, the trimmed reads related to FW, BW, and SW conditions for *A. marmorata*, *O. keta,* and *O. melastigma* were mapped against reference transcriptomes to calculate expression values, using the proprietary *RNA-Seq Analysis tool* included in the CLC Genomics Workbench v.12 to set the following mapping parameters: length fraction = 0.75 and similarity fraction = 0.98. In order to remove redundancy, the RepeatMasker output file was filtered removing entries not classified as TEs, keeping those with the highest score and length values. The overall expression of each TE class was calculated by summing the expression values of each TE type (SINE, Retro, LTR, LINE, DNA transposons, and unclear). The expression values were then transformed in percentage of mapped reads to achieve comparability between species.

4.2. Identification and Expression Analysis of Genes of Interest

Focusing on genes of interest, they were retrieved through TBLASTN [40] from the de novo assembled transcriptomes obtained from gills of *A. marmorata*, *O. melastigma*, and *O. keta*. In particular, the search was made for genes involved in heterochromatinization (*HP1α*, *HP1βa*, *HP1βb*, *HP1γ*, *DNMT1*, *DNMT3aa*), genes related to the NuRD complex (*KRAB-like*, *TRIM33*, *CHD3*, *CHD4a*, *CHD4b*, *HDAC1*, *HDAC2, MBD2*, *MBD3b*, *MTA1*, *MTA2*, *MTA3*, *GATAD2ab*, *GATAD2b*, *RBBP4*, *RBBP7*), and four genes of the *Argonaute* subfamily (*AGO1*, *AGO2*, *AGO3*, and *AGO4*). Transcripts were translated using the EMBOSS Transeq translation tool (https://www.ebi.ac.uk/Tools/st/emboss_transeq/ accessed on 18 November 2021) and UTR and CDS regions were identified (Table S2). The aforementioned sequences were deposited in GenBank under the accession numbers listed in the Table S2. To ensure the comparison between species, gene expression values were computed using a scaling factor, based on the cumulative expression of a dataset composed of 3640 orthologs derived from the Actinopterygii OrthoDB v10 database.

In detail, the dataset was created as follows: for each transcriptome of the three species considered in this study, expression levels of genes attributed from BUSCO analyses as "*complete and single copy*" and "*fragmented*" genes were kept as the number of mapped reads; expression values of genes classified as "*duplicated*" (most probably derived from transcriptional isoforms) were computed as the sum of each copy of single BUSCO and the expression levels of "missing" genes were set to 0. This dataset was then used as a calibration set, computing a scaling factor that was applied to the original expression values of the genes of interest as described in Biscotti et al., (2016) [41].

4.3. Molecular Modelling

The computational analysis was performed to investigate the putative interaction between TRIM33 and KRAB-like sequence identified in actinopterygians. The three-dimensional (3D) model for TRIM33 was built using SwissModel web server (https://swissmodel.expasy.org/ accessed on 12 January 2022) [42–44], while the putative KRAB domain contained in the KRAB-like sequence was built through Modeller web server (https://salilab.org/modeller/, https://swissmodel.expasy.org/ accessed on 12 January 2022) [45]. Because the template used for the modelling of the KRAB domain was a predicted structure of the *Homo sapiens* ZNF93 (UniProtKB accession number AF-P35789), we validated this model comparing it to the NMR murine structure (PDB ID 1V65).

The docking was performed using the HADDOCK web server [46,47]. We selected as active residues (AIRs) the whole TRIM33/TRIM28 coiled-coil region and all the KRAB-

like/KRAB-ZNF93 residues pointing out that our docking protocol was completely unbiased. The HADDOCK score was used to classify the poses of the complexes sampled for each docking run. In principle, the lowest HADDOCK score corresponds to the highest predicted protein–protein affinity [47,48].

For the multiple sequence alignment, we used COBALT, a constraint-based alignment tool for multiple protein sequences included in the NCBI toolkit [49].

Molecular graphics and analyses were obtained with UCSF ChimeraX [50].

5. Conclusions

TEs represent one of the most intriguing genome components and the analysis of their activity in somatic tissues contributes to understanding the role of these genetic elements in regulating the physiology of organisms [51]. Our data evidenced a variation in TE contribution in the case of juvenile eels, commonly adapted to salty water, when exposed to brackish and freshwater conditions. Indeed, it seems that gill cells activate TE controlling systems in response to the TE increase occurred from salt water to freshwater. Interestingly, our analyses suggested for the first time that besides *Argonaute 4*, *HP*, *DNMT*, the NuRD complex might also be involved in fish TE silencing. Therefore, we propose for the first time the existence of a KRAB-like domain specific to actinopterygians that together with TRIM33 allows the functioning of the NuRD complex also in fish lineage.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23095215/s1.

Author Contributions: Methodology, transcriptomic data analysis, and writing original draft, E.C. and F.C.; supervised transcriptomic data analysis, S.G. and M.G.; computational structural analyses, D.D.M., N.P. and A.L.T.; project supervision, A.C. and M.B.; conceptualization, funding acquisition, project administration, M.A.B. All authors discussed the results, wrote the manuscript, and commented on the final version of the manuscript prior to submission. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC were funded by Ministero dell'Università e della Ricerca grant number 55_R. SCIENT.A_2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Sequences were deposited in GenBank under the accession numbers provided in the Table S2.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fujino, K.; Hashida, S.N.; Ogawa, T.; Natsume, T.; Uchiyama, T.; Mikami, T.; Kishima, Y. Temperature controls nuclear import of Tam3 transposase in *Antirrhinum*. *Plant J.* **2011**, *65*, 146–155. [CrossRef] [PubMed]
2. Casacuberta, E.; González, J. The impact of transposable elements in environmental adaptation. *Mol. Ecol.* **2013**, *6*, 1503–1517. [CrossRef] [PubMed]
3. Makarevitch, I.; Waters, A.J.; West, P.T.; Stitzer, M.; Hirsch, C.N.; Ross-Ibarra, J.; Springer, N.M. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* **2015**, *11*, e1004915.
4. Schmidt, A.L.; Anderson, L.M. Repetitive DNA elements as mediators of genomic change in response to environmental cues. *Biol. Rev. Camb. Philos. Soc.* **2006**, *81*, 531–543. [CrossRef] [PubMed]
5. Fablet, M.; Vieira, C. Evolvability, epigenetics and transposable elements. *Biomol. Concepts* **2011**, *2*, 331–341. [CrossRef]
6. Canapa, A.; Biscotti, M.A.; Barucca, M.; Carducci, F.; Carotti, E.; Olmo, E. Shedding light upon the complex net of genome size, genome composition and environment in chordates. *Eur. Zool. J.* **2020**, *87*, 192–202. [CrossRef]
7. Chu, C.G.; Tan, C.T.; Yu, G.T.; Zhong, S.; Xu, S.S.; Yan, L. A novel retrotransposon inserted in the dominant *Vrn-B1* allele confers spring growth habit in tetraploid wheat (*Triticum turgidum* L.). *G3* **2011**, *1*, 637–645. [CrossRef]
8. Carducci, F.; Barucca, M.; Canapa, A.; Carotti, E.; Biscotti, M.A. Mobile elements in ray-finned fish genomes. *Life* **2020**, *10*, 221. [CrossRef]
9. Carducci, F.; Biscotti, M.A.; Forconi, M.; Barucca, M.; Canapa, A. An intriguing relationship between teleost *Rex3* retroelement and environmental temperature. *Biol. Lett.* **2019**, *15*, 20190279. [CrossRef]

10. Carotti, E.; Carducci, F.; Canapa, A.; Barucca, M.; Greco, S.; Gerdol, M.; Biscotti, M.A. Transposable elements and teleost migratory behaviour. *Int. J. Mol. Sci.* **2021**, *22*, 602. [CrossRef]
11. Pappalardo, A.M.; Ferrito, V.; Biscotti, M.A.; Canapa, A.; Capriglione, T. Transposable elements and stress in vertebrates: An overview. *Int. J. Mol. Sci.* **2021**, *22*, 1970. [CrossRef] [PubMed]
12. Liu, D.; Yang, J.; Tang, W.; Zhang, X.; Royster, C.M.; Zhang, M. SINE Retrotransposon variation drives Ecotypic disparity in natural populations of *Coilia nasus*. *Mob. DNA* **2020**, *11*, 4. [CrossRef] [PubMed]
13. Steendam, C.; Verhelst, P.; Van Wassenbergh, S.; De Meyer, J. Burrowing behaviour of the European eel (*Anguilla anguilla*): Effects of life stage. *J. Fish Biol.* **2020**, *97*, 1332–1342. [CrossRef] [PubMed]
14. Lee, S.Y.; Lee, H.J.; Kim, Y.K. Comparative transcriptome profiling of selected osmotic regulatory proteins in the gill during seawater acclimation of chum salmon (*Oncorhynchus keta*) fry. *Sci. Rep.* **2020**, *10*, 1987. [CrossRef] [PubMed]
15. Beechie, T.J.; Nicol, C.; Fogel, C.; Jorgensen, J.; Timpane-Padgham, B. *Chum Salmon Life-Cycle Model Description and Results for the Chehalis River Basin*; NOAA Contract Report NMFS-NWFSC-CR-2021-02; U.S. Department of Commerce: Washington, DC, USA, 2021.
16. Cao, Q.; Wang, H.; Fan, C.; Sun, Y.; Li, J.; Cheng, J.; Chu, P.; Yin, S. Environmental salinity influences the branchial expression of TCR pathway related genes based on transcriptome of a catadromous fish. *Comp. Biochem. Physiol. D* **2021**, *38*, 100815. [CrossRef]
17. Bourque, G.; Burns, K.H.; Gehring, M.; Gorbunova, V.; Seluanov, A.; Hammell, M.; Imbeault, M.; Izsvák, Z.; Levin, H.L.; Macfarlan, T.S.; et al. Ten things you should know about transposable elements. *Genome Biol.* **2018**, *19*, 199. [CrossRef]
18. Boudinot, P.; van der Aa, L.M.; Jouneau, L.; Du Pasquier, L.; Pontarotti, P.; Briolat, V.; Benmansour, A.; Levraud, J.P. Origin and evolution of TRIM proteins: New insights from the complete TRIM repertoire of zebrafish and pufferfish. *PLoS ONE* **2011**, *6*, 7. [CrossRef]
19. Stoll, G.A.; Oda, S.I.; Chong, Z.S.; Yu, M.; McLaughlin, S.H.; Modis, Y. Structure of KAP1 tripartite motif identifies molecular interfaces required for retroelement silencing. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 15042–15051. [CrossRef]
20. Di Marino, D.; Chillemi, G.; De Rubeis, S.; Tramontano, A.; Achsel, T.; Bagni, C. MD and Docking Studies Reveal That the Functional Switch of CYFIP1 is mediated by a Butterfly-like Motion. *J. Chem. Theory Comput.* **2015**, *11*, 3401–3410. [CrossRef]
21. Trucchi, E.; Gratton, P.; Mafessoni, F.; Motta, S.; Cicconardi, F.; Mancia, F.; Bertorelle, G.; D'Annessa, I.; Di Marino, D. Population dynamics and structural effects at short and long range support the hypothesis of the selective advantage of the G614 SARS-CoV-2 Spike Variant. *Mol. Biol. Evol.* **2021**, *38*, 1966–1979. [CrossRef]
22. Schmidt, D.; Schwalie, P.C.; Wilson, M.D.; Ballester, B.; Gonçalves, A.; Kutter, C.; Brown, G.D.; Marshall, A.; Flicek, P.; Odom, D.T. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **2012**, *148*, 335–348. [CrossRef] [PubMed]
23. Crepaldi, L.; Policarpi, C.; Coatti, A.; Sherlock, W.T.; Jongbloets, B.C.; Down, T.A.; Riccio, A. Binding of TFIIIC to sine elements controls the relocation of activity-dependent neuronal genes to transcription factories. *PLoS Genet.* **2013**, *9*, 8. [CrossRef] [PubMed]
24. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **2017**, *18*, 71–86. [CrossRef] [PubMed]
25. Miousse, I.R.; Chalbot, M.C.; Lumen, A.; Ferguson, A.; Kavouras, I.G.; Koturbash, I. Response of transposable elements to environmental stressors. *Mutat. Res. Rev. Mutat. Res.* **2015**, *765*, 19–39. [CrossRef]
26. Fedoroff, N.V. Transposable elements, epigenetics, and genome evolution. *Science* **2012**, *338*, 758–767. [CrossRef]
27. Slotkin, R.K.; Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **2007**, *8*, 272–285. [CrossRef]
28. Leebonoi, W.; Sukthaworn, S.; Panyim, S.; Udomkit, A. A novel gonad-specific Argonaute 4 serves as a defense against transposons in the black tiger shrimp *Penaeus monodon*. *Fish Shellfish Immunol.* **2015**, *42*, 280–288. [CrossRef]
29. Berrens, R.V.; Andrews, S.; Spensberger, D.; Santos, F.; Dean, W.; Gould, P.; Sharif, J.; Olova, N.; Chandra, T.; Koseki, H.; et al. An endosiRNA-based repression mechanism counteracts transposon activation during global DNA demethylation in embryonic stem cells. *Cell Stem Cell* **2017**, *21*, 694–703. [CrossRef]
30. Pfefferli, C.; Jaźwińska, A. The art of fin regeneration in zebrafish. *Regeneration* **2015**, *2*, 72–83. [CrossRef]
31. Wang, J.; Huang, J.; Shi, G. Retrotransposons in pluripotent stem cells. *Cell Regen.* **2020**, *9*, 4. [CrossRef]
32. Grassi, D.A.; Jönsson, M.E.; Brattås, P.L.; Jakobsson, J. TRIM28 and the control of transposable elements in the brain. *Brain Res.* **2019**, *1705*, 43–47. [CrossRef] [PubMed]
33. Seah, M.; Wang, Y.; Goy, P.A.; Loh, H.M.; Peh, W.J.; Low, D.; Han, B.Y.; Wong, E.; Leong, E.L.; Wolf, G.; et al. The KRAB-zinc-finger protein ZFP708 mediates epigenetic repression at RMER19B retrotransposons. *Development* **2019**, *146*, dev170266. [CrossRef] [PubMed]
34. Imbeault, M.; Helleboid, P.Y.; Trono, D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **2017**, *543*, 550–554. [CrossRef]
35. Ecco, G.; Cassano, M.; Kauzlaric, A.; Duc, J.; Coluccio, A.; Offner, S.; Imbeault, M.; Rowe, H.M.; Turelli, P.; Trono, D. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. *Dev. Cell* **2016**, *36*, 611–623. [CrossRef]
36. Carducci, F.; Carotti, E.; Gerdol, M.; Greco, S.; Canapa, A.; Barucca, M.; Biscotti, M.A. Investigation of the activity of transposable elements and genes involved in their silencing in the newt *Cynops orientalis*, a species with a giant genome. *Sci. Rep.* **2021**, *11*, 14743. [CrossRef] [PubMed]

37. Helleboid, P.Y.; Heusel, M.; Duc, J.; Piot, C.; Thorball, C.W.; Coluccio, A.; Pontis, J.; Imbeault, M.; Turelli, P.; Aebersold, R.; et al. The interactome of KRAB zinc finger proteins reveals the evolutionary history of their functional diversification. *EMBO J.* **2019**, *38*, 18. [CrossRef] [PubMed]
38. Liang, P.; Saqib, H.; Lin, Z.; Zheng, R.; Qiu, Y.; Xie, Y.; Ma, D.; Shen, Y. RNA-seq analyses of marine medaka (*Oryzias melastigma*) reveals salinity responsive transcriptomes in the gills and livers. *Aquat. Toxicol.* **2021**, *240*, 105970. [CrossRef]
39. Manni, M.; Berkeley, M.R.; Seppey, M.; Zdobnov, E.M. BUSCO: Assessing Genomic Data Quality and Beyond. *Curr. Protoc.* **2021**, *1*, e323. [CrossRef]
40. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
41. Biscotti, M.A.; Gerdol, M.; Canapa, A.; Forconi, M.; Olmo, E.; Pallavicini, A.; Barucca, M.; Schartl, M. The lungfish transcriptome: A glimpse into molecular evolution events at the transition from water to land. *Sci. Rep.* **2016**, *6*, 21571. [CrossRef]
42. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, 296–303. [CrossRef] [PubMed]
43. Guex, N.; Peitsch, M.C.; Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **2009**, *30*, 162–173. [CrossRef] [PubMed]
44. Bienert, S.; Waterhouse, A.; de Beer, T.A.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res.* **2017**, *45*, 313–319. [CrossRef] [PubMed]
45. Šali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815. [CrossRef] [PubMed]
46. Dominguez, C.; Boelens, R.; Bonvin, A.M. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737. [CrossRef] [PubMed]
47. van Zundert, G.C.P.; Rodrigues, J.; Trellet, M.; Schmitz, C.; Kastritis, P.L.; Karaca, E.; Melquiond, A.; van Dijk, M.; de Vries, S.J.; Bonvin, A. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720–725. [CrossRef] [PubMed]
48. Honorato, R.V.; Koukos, P.I.; Jiménez-García, B.; Tsaregorodtsev, A.; Verlato, M.; Giachetti, A.; Rosato, A.; Bonvin, A. Structural biology in the clouds: The WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.* **2021**, *8*, 729513. [CrossRef]
49. Papadopoulos, J.S.; Agarwala, R. COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **2007**, *23*, 1073–1079. [CrossRef]
50. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Meng, E.C.; Couch, G.S.; Croll, T.I.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*, 70–82. [CrossRef]
51. Friedli, M.; Trono, D. The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell. Dev. Biol.* **2015**, *31*, 429–451. [CrossRef]

3.6 Transposable element tissue-specific response to temperature stress in the stenothermal fish *Puntius tetrazona*

animals

MDPI

Article

Transposable Element Tissue-Specific Response to Temperature Stress in the Stenothermal Fish *Puntius tetrazona*

Elisa Carotti †, Federica Carducci *,†, Adriana Canapa, Marco Barucca and Maria Assunta Biscotti

Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy

* Correspondence: f.carducci@univpm.it

† These authors contributed equally to this work.

Simple Summary: Teleosts are one of the most diversified group of vertebrates, as they colonized different aquatic environments worldwide. Stenothermal fish are characterized by a limited ability to tolerate temperature changes, therefore variations in this abiotic factor represent a threat for their survival. In the present study, we selected the tiger barb *Puntius tetrazona*, for which RNA-Seq data were available for brain, gill, and liver, to investigate the possible role of the transcriptional activity of transposable elements (TEs) in the rapid adaptation of this species. Our findings highlighted a tissue-specific response of TEs with a remarkable increase at 13 °C recorded in liver, strengthening the view of TEs as source of genetic variability acting positively in species resilience.

Abstract: Ray-finned fish represent a very interesting group of vertebrates comprising a variety of organisms living in different aquatic environments worldwide. In the case of stenothermal fish, thermal fluctuations are poorly tolerated, thus ambient temperature represents a critical factor. In this paper, we considered the tiger barb *Puntius tetrazona*, a freshwater fish belonging to the family Cyprinidae, living at 21–28 °C. We analyzed the available RNA-Seq data obtained from specimens exposed at 27 °C and 13 °C to investigate the transcriptional activity of transposable elements (TEs) and genes encoding for proteins involved in their silencing in the brain, gill, and liver. TEs are one of the tools generating genetic variability that underlies biological evolution, useful for organisms to adapt to environmental changes. Our findings highlighted a different response of TEs in the three analyzed tissues. While in the brain and gill, no variation in TE transcriptional activity was observed, a remarkable increase at 13 °C was recorded in the liver. Moreover, the transcriptional analysis of genes encoding proteins involved in TE silencing such as heterochromatin formation, the NuRD complex, and the RISC complex (e.g., AGO and GW182 proteins) highlighted their activity in the hepatic tissue. Overall, our findings suggested that this tissue is a target organ for this kind of stress, since TE activation might regulate the expression of stress-induced genes, leading to a better response of the organism to temperature changes. Therefore, this view corroborates once again the idea of a potential role of TEs in organism rapid adaptation, hence representing a promising molecular tool for species resilience.

Keywords: transposable elements; fish; silencing mechanisms; stress temperature; tissue-specific response; phenotypic plasticity

Citation: Carotti, E.; Carducci, F.; Canapa, A.; Barucca, M.; Biscotti, M.A. Transposable Element Tissue-Specific Response to Temperature Stress in the Stenothermal Fish *Puntius tetrazona*. *Animals* **2023**, *13*, 1. https://doi.org/10.3390/ani13010001

Academic Editor: Gonzalo Martínez-Rodríguez

Received: 4 November 2022
Revised: 14 December 2022
Accepted: 15 December 2022
Published: 20 December 2022

1. Introduction

The evolutionary success of a species is strictly related to their ability to cope with changes in environmental conditions. Phenotypic plasticity is the main mechanism used by organisms to respond to rapid environmental perturbations. Indeed, it can produce well-adapted phenotypes that, in turn, might improve the fitness of organisms [1,2]. Phenotypic plasticity is also the result of environmental sensitivity of the genome that, through modifications in the heterochromatin structure, varies the expression not only of genes

but also of transposable elements (TEs) [3]. TEs are interspersed repeated sequences that constitute a considerable fraction in several eukaryotic genomes. They adopt different transposition mechanisms to move throughout the host genome and are classified as TEs of class I, also called retrotransposons, including Long Terminal Repeats (LTRs) and non-LTRs (LINE and SINE), which use an RNA intermediate molecule that is reverse transcribed into complementary DNA by a reverse transcriptase (RT), followed by the reintegration into the host genome through a copy and paste mechanism, and class II of TEs, also called DNA transposons, use a DNA intermediate molecule to transpose and are characterized by the cut and paste mechanism of transposition, with an exception made for Helitron and Maverick/Polinton, which mobilize by a self-synthesizing mechanism involving direct synthesis of a DNA copy [4]. TEs are molecular elements that are activated by biotic and abiotic factors, contributing to the flexible response of the genome to external stimuli. They can cause chromosome rearrangements, increase mutation rates, or influence the expression of nearby genes, triggering the rapid adaptation of organisms to new conditions [5–8]. Barbara McClintock (1984) [9] was the first to propose a helpful role of TE-induced mutations in response to stress. However, transposition events might have deleterious effects; hence, several protective mechanisms are turned on by the host genome for its safeguard. DNA methylation is the most widely adopted mechanism for silencing TEs and consists of the addition of methyl groups to DNA. Another epigenetic mechanism regulating TE transcription regards histone modifications mediated by the Nucleosome Remodeling and Deacetylase (NuRD) complex [10]. In mammals, this complex is recruited at the TE sequence level by Krüppel-associated box domain zinc finger proteins (KRAB-ZFPs) through the involvement of the KRAB-associated protein-1 (KAP1) co-factor. Although these two latter proteins are absent in actinopterygians, our group recently proposed a KRAB-like and tripartite motif containing 33 (TRIM33) proteins as substitutes for the NuRD complex functioning in this evolutionary lineage [11]. Another conserved mechanism of TE silencing is based on the action of Argonaute proteins (AGOs) that use small non-coding RNAs (ncRNAs), such as microRNAs (miRNAs) and short interference RNAs (siRNAs) [12,13]. AGOs, in association with members of trinucleotide repeat containing adaptor 6 (TNRC6), part of the GW182 protein family [14–16], form a large multiprotein complex known as RNA-Induced Silencing Complex (RISC) that performs a post-translational regulation through RNA degradation and/or translation inhibition [17,18].

While a relationship between TEs and environment has been well established for plants [19–26] and *Drosophila* [27], this issue has only recently been investigated in vertebrates [8,11,27–30]. Actinopterygii or ray-finned fish are a very diversified group adapted to aquatic environments worldwide, from salt to fresh waters, from cold to warm seas, and from high-elevation mountain lakes to extreme sea depths. Studies performed to date in this taxon have provided evidence that the evolution of a specific TE type is the result of an intricate relationship with the environment. Rhee and colleagues (2017) [31] reported a high content of RC/helitrons that might be responsible for the high genetic diversity observed in *Kryptolebias marmoratus* populations. Moreover, they have found a maintenance of the TE genomic content in the self-fertilizing hermaphroditic *K. marmoratus*. This feature seems to promote genome recombination that, in addition to a mixed mating strategy, might have contributed to the evolutionary adaptation to ecological pressure of a species. Analyzing 52 fish species, Yuan and co-workers (2018) [32] evidenced a positive correlation between the presence of specific repetitive elements and the aquatic environments in which the considered species live. The chromosomal diversification and the consequent rapid speciation of the Antarctic teleosts belonging to the genus *Trematomus* have been attributed to the TE mobilization of *DIctyostelium* Repetitive Sequence 1 (DIRS1) [33]. Our research group performed a phylogenetic study evidencing an unexpected clusterization of the Rex3 retroelements in fish species living in cold environments independently from their taxonomic relationships [29]. Recently, we also published a paper concerning the possible correlation between the migratory behavior of catadromous teleost species and the amount of SINE retroelements [30]. In addition, analyzing RNA-seq data, we demonstrated for the

first time an activation of TE transcriptional activity in response to salinity variations in the marbled eel *Anguilla marmorata* [11]. Since most teleosts are poikilotherms, environmental stimuli affect physiological and metabolic activities in these organisms [34–37]. In particular, ambient temperature might be a critical factor for stenothermal fish in which thermal fluctuations are less tolerated. The tiger barb *Puntius tetrazona* is a freshwater fish belonging to the family Cyprinidae, living at 21–28 °C. In this paper, we analyzed the RNA-Seq data of tiger barb exposed at 27 °C and 13 °C [38] to investigate the transcriptional activity of TEs and genes encoding for proteins involved in their silencing in the brain, gill, and liver. Our results pointed out a clear response of TEs in the liver consistent with the idea of a potential role of these elements in the rapid adaptation, thus representing a promising molecular tool for species resilience. Furthermore, all TE silencing mechanisms were active in the hepatic tissue and *ago* genes also in the brain and gill tissues.

2. Materials and Methods

Tiger barb RNA-Seq raw data were obtained from the Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra, accessed on 15 June 2022) under the accession number SRP153005 [38]. Data were obtained sequencing the brain, gill, and liver of ninety fish (eight-month-old) divided into two groups, one used as the control group (27 °C) and another for testing the acute cold stress until 13 °C. The temperature was set to 27 ± 0.5 °C on the first day, followed by a decrease of 2 °C every 24 h. Fish were kept at 13 ± 0.5 °C for 24 h, and then, five fish, in triplicate, in each tank, were anesthetized before dissection. Raw paired-end reads were imported in the CLC Genomics Workbench v.12 (Qiagen, Hilden, Germany) (Table S1) and trimmed removing sequencing adapters, low-quality bases, and low-quality read ends using default parameters. Trimmed reads were then de novo assembled with the "De Novo Assembly" of CLC Genomics Workbench v.12 using the default parameters. Completeness of the de novo assembled transcriptome was evaluated through BUSCO v.5 using the Actinopterygii OrthoDB v.10 database as a reference [39].

2.1. Estimation of TE Transcriptional Activity

To estimate the TE transcriptional activity, we first identified TEs in the de novo assembled transcriptomes with RepeatMasker v.4.1.0 (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker, accessed on 15 July 2022) using a custom library created following the methodology described in our previous work [30]. After the RepeatMasker analysis, the trimmed reads related to 27 °C and 13 °C were mapped against the reference transcriptome to calculate the expression values using the proprietary RNA-Seq Analysis tool included in the CLC Genomics Workbench v.12 and setting the following mapping parameters: length fraction = 0.75 and similarity fraction = 0.98. To remove redundancies, the RepeatMasker output file was filtered, removing entries not classified as TEs and keeping those with the highest score and length values. The overall expression of each TE class was calculated by summing the expression values of each TE type: DNA transposons, LINE, LTR (also including endogenous retroviruses), non-LTR, Retro (retroelements that are not classified in any of the two main subclasses), and SINE. In our results, we also reported the fraction of unclear elements that are referred to repetitive elements without specific features to determine their attribution to a given TE class. We included this fraction to show the low percentage of the unclear fraction, representative of the goodness of our analyses. The expression values were then transformed into a percentage of mapped reads to achieve the comparability between species.

2.2. Identification and Expression Enalysis of Genes of Interest

Genes of interest were retrieved through TBLASTN [40] from the assembled transcriptome obtained from the brain, gill, and liver of *P. tetrazona*. In particular, the search was made for genes involved in heterochromatinization (*chromobox homolog 5* (*cbx5*), *chromobox homolog 1* (*cbx1*), *chromobox homolog 3* (*cbx3*), *DNA (cytosine-5-)-methyltransferase 1* (*dnmt1*), *DNA (cytosine-5-)-methyltransferase 3 alpha* (*dnmt3a*), and *SET domain bifurcated*

histone lysine methyltransferase 1 (*setdb1*)); genes related to the NuRD complex (*krab-like*, *trim33*, *chromodomain helicase DNA binding protein 3* (*chd3*), *chromodomain helicase DNA binding protein 4* (*chd4*), *histone deacetylase 1* (*hdac1*), *methyl-CpG binding domain protein 2* (*mbd2*), *methyl-CpG binding domain protein 3* (*mbd3*), *metastasis associated 1* (*mta1*), *metastasis associated 1 family, member 2* (*mta2*), *metastasis associated 1 family, member 3* (*mta3*), *GATA zinc finger domain containing 2* (*gatad2*), and *retinoblastoma binding protein 4* (*rbbp4*), *retinoblastoma binding protein 7* (*rbbp7*)); four genes of the *Argonaute* subfamily (*argonaute RISC component 1* (*ago1*), *argonaute RISC component 2* (*ago2*), *argonaute RISC component 3* (*ago3*), and *argonaute RISC component 4* (*ago4*)); and three genes of the GW182 family proteins (*tnrc6a*, *tnrc6b*, and *tnrc6c*). Transcripts were translated using the EMBOSS Transeq translation tool (https://www.ebi.ac.uk/Tools/st/emboss_transeq/, accessed on 12 September 2022), and the UTR and CDS regions were identified (Table S2). The aforementioned sequences were deposited in GenBank under the accession numbers listed in Table S2. To ensure the comparison between species, the gene expression values were computed using a scaling factor based on the cumulative expression of a dataset composed of 2124 orthologs derived from the Actinopterygii OrthoDB v.10 database [39]. In detail, the dataset was created as follows: for each transcriptome of the tissue considered in this study, expression levels of the genes attributed to the BUSCO analyses as "complete and single copy" and "fragmented" genes were kept as the number of mapped reads; expression values of genes classified as "duplicated" (most probably derived from transcriptional isoforms) were computed as the sum of each copy of single BUSCO, and the expression levels of "missing" genes were set to 0. This dataset was then used as a calibration set, computing a scaling factor that was applied to the original expression values of the genes of interest, as described in Biscotti et al. (2016) [41]. Transcriptional values of the genes analyzed in this study were reported as Transcripts Per Million (TPM).

2.3. Statistics

For each tissue, the data on gene expressions and TE transcriptional activity obtained from the three replicates for each tested condition were expressed as the mean ± standard error, and statistically significant differences were evaluated by one-way ANOVA. The symbol * indicates p-values < 0.05, ** for p-values < 0.01, and *** for p-values < 0.001.

3. Results

The transcriptional activity of the TEs was evaluated as a percentage of the mapped reads in the brain, gill, and liver of the tiger barb (Figure 1). The data reported in Figure 1A were referred to the total TE transcriptional contribution at the control (27 °C) and test (13 °C) conditions. Comparing the total TE transcriptional activity at 27 °C between the analyzed tissues, the highest level was showed in the brain, followed by the gill and then liver. At 13 °C, the highest TE transcriptional activity in the brain was confirmed, and the difference between the gill and liver was less remarkable. Moreover, analyzing the data obtained in each tissue, a statistically significant difference between the two conditions considered here, was detectable only in the liver (p-value: 1.65×10^{-5}), in which the total TE transcriptional activity was increased at 13 °C. Profiles of the TE relative abundances between the brain and gill samples were similar. Overall, a prevalence of DNA transposons followed by LINE, unclear, Retro, LTR, SINE, and non-LTR retroelements was shown. In the liver, LINE retroelements prevailed at 27 °C, while the TE distribution pattern was similar to that observed in the other two tissues at 13 °C. Despite the absence of an appreciable variation in the total contribution of TEs in the brain and gill, statistically significant differences emerged for the same tissues in single TE types between the two tested conditions. Indeed, in the brain, the LINE, LTR, and SINE retroelements showed a significant decrease passing from 27 °C to 13 °C (Figure 1B). In the gill, except for unclear elements and SINE retroelements, all TE types experienced significant changes, with DNA transposons and Retro having a higher activity at 13 °C, while LINE, LTR, and non-LTR retroelements have a lower activity in this condition than in the control one (Figure 1C). In

the hepatic tissue, with the exception of LINE retroelements that showed lower values at 13 °C, all the analyzed TE types significantly increased their transcriptional contribution (Figure 1D). These increases were ascribable to a few specific elements. Indeed, the ten elements having the highest values of variation in transcriptional activity between the two conditions represented 62.27%, 86.0%, and 73.35% for the DNA transposons, SINE, and LTR retroelements, respectively. In particular, two elements, DNA/hAT-Ac and LTR/DIRS, mostly contributed to the upregulation of their belonging TE type (32.67% and 51.49%, respectively). Moreover, in the other tissues, the ten elements having the highest values of variation in transcriptional activity between the two conditions were different from those identified in the liver, except for one LTR and four SINE retroelements.

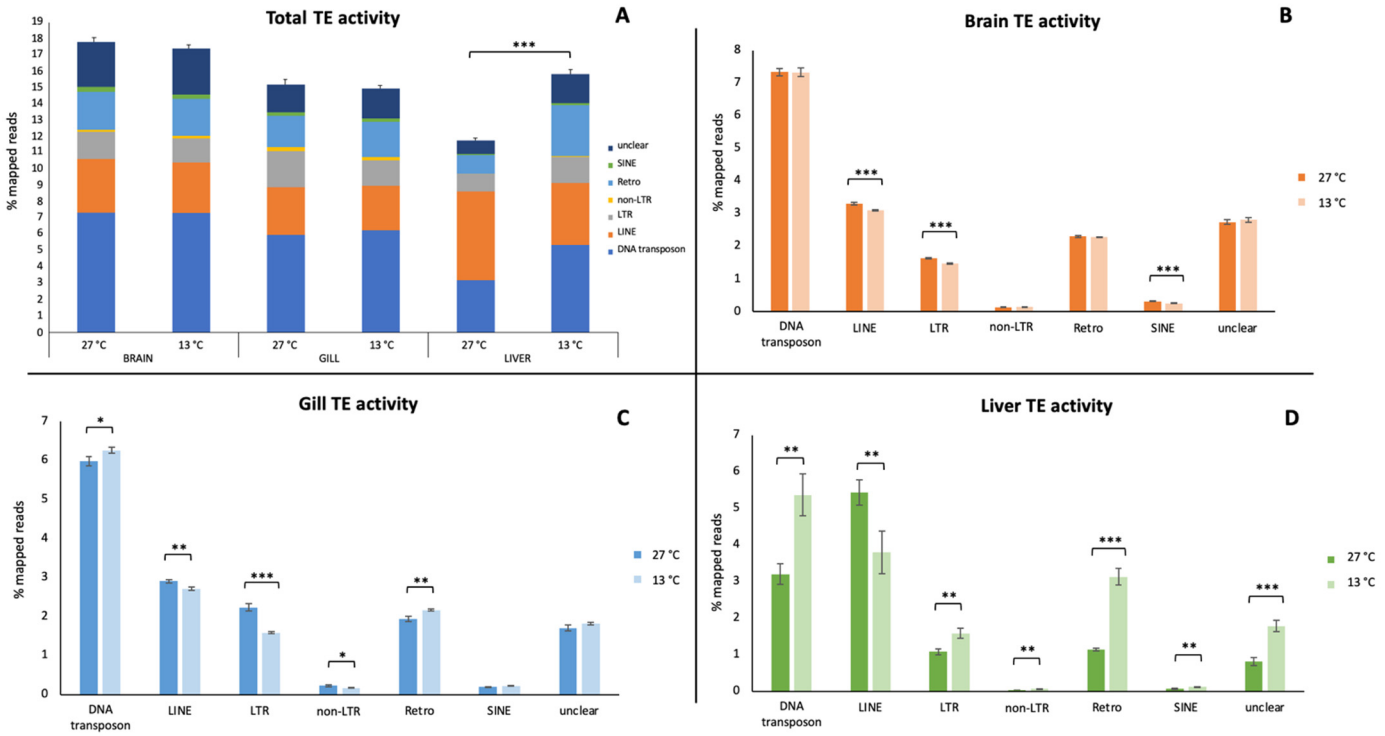

Figure 1. Transcriptional contribution of transposable elements in the brain, gill, and liver in *Puntius tetrazona*. (**A**) Total TE transcriptional contribution in the three analyzed tissues. (**B**) Transcriptional contribution of each TE type in the brain. (**C**) Transcriptional contribution of each TE type in the gill. (**D**) Transcriptional contribution of each TE type in the liver. Values expressed as the percentage of mapped reads are reported for both tested temperature conditions (27 °C and 13 °C) for each tissue. Statistically significant differences are presented as * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

The transcriptional activity of TEs is controlled by the host genome through silencing mechanisms that determine a stronger heterochromatinization status through the activity of heterochromatin proteins, DNA methyltransferases, and the NuRD complex, and AGO proteins also involved in mRNA cleavage and translation repression. In this study, a total of 34 genes encoding proteins involved in TE silencing mechanisms were retrieved: ten concerning heterochromatin formation (*cbx5*, *cbx1a*, *cbx1b*, *cbx3a*, *cbx3b*, *dnmt1*, *dnmt3aa*, *dnmt3ab*, *setdb1a*, and *setdb1b*); 16 related to the NuRD complex (*krab-like*, *trim33*, *chd3*, *chd4a*, *chd4b*, *hdac1*, *mbd2*, *mbd3a*, *mbd3b*, *mta1*, *mta2*, *mta3*, *gatad2ab*, *gatad2b*, *rbbp4*, and *rbbp7*); four to *ago* (*ago1*, *ago2*, *ago3*, and *ago4*); and four belonging to the GW182 family proteins (*tnrc6a*, *tnrc6b*, *tnrc6c1*, and *tnrc6c2*). Of these, 20 showed a complete CDS, 2 were incomplete, and 12 fragmented gene sequences were replaced with the complete version available in public repositories to further refine the assembled transcriptome (Table S2).

Regarding the transcriptional activity, our findings pointed out that all these mechanisms were active in *P. tetrazona* (Figures 2–4). In the case of the brain, a significant increase

in the transcription of the genes involved in the heterochromatin formation was detected for *cbx3a,* as well as the three *dnmt* (*dnmt1*, *dnmt3aa*, and *dnmt3ab*), while a significant decrease was shown for *cbx3b* (Figure 2A). No remarkable change was detected comparing the expression of the genes encoding for the proteins of the NuRD complex between the two considered temperatures (Figure 2B) differently from the *ago* genes and those of the GW182 family proteins that recorded a statistically significant increase (Figure 2C,D). Analyzing the RNA-Seq data of a gill, a general statistically significant decrease in the transcriptional level of the genes involved in heterochromatin formation (*cbx1a*, *cbx3b)* and the NuRD complex (*krab-like*, *chd3*, *chd4b*, and *mta3*) was observed, while, in the latter, a significant increase was observed for *mta1* (Figure 3A,B). Differences in the expression of *ago* genes and in those encoding the GW182 family proteins were identified also in the gill (Figure 3C,D). In the hepatic tissue, most of the analyzed genes showed higher values at 13 °C (Figure 4). In particular, *cbx5*, *cbx1b*, *cbx3b*, *dnmt1*, *dnmt3aa*, *dnmt3ab*, *trim33*, *chd4a*, *hdac1*, *gatad2ab*, *gatad2b*, *rbbp4*, the four *ago* genes, and two of the GW182 family proteins showed significant changes.

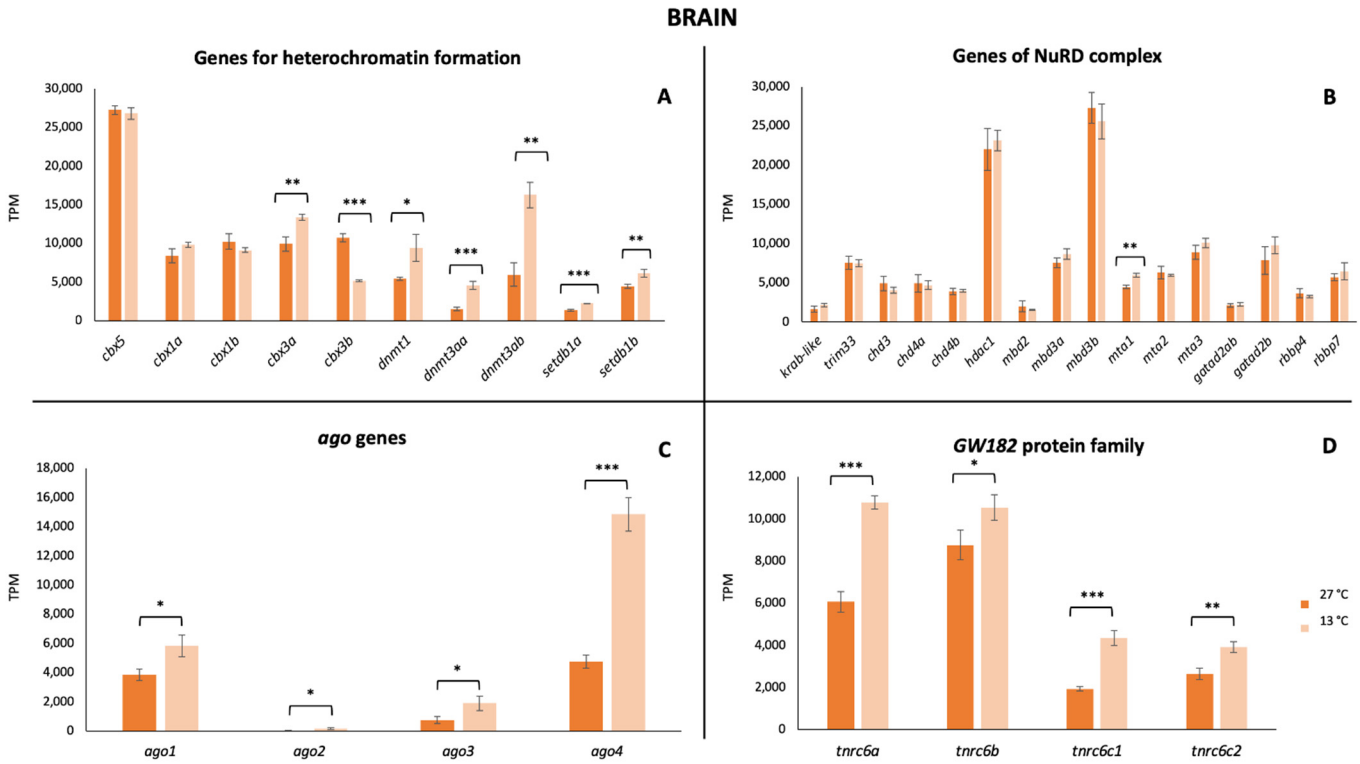

Figure 2. Expression of genes involved in TE silencing mechanisms in brain. (**A**) Expression values of genes encoding for proteins involved in heterochromatin formation. (**B**) Expression values of genes encoding for proteins forming the NuRD complex. (**C**) Expression values of members of the gene subfamily *Argonaute*. (**D**) Expression values of genes belonging to the GW182 family proteins. Values are reported for both tested temperature conditions (27 °C and 13 °C). Statistically significant differences are presented as * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

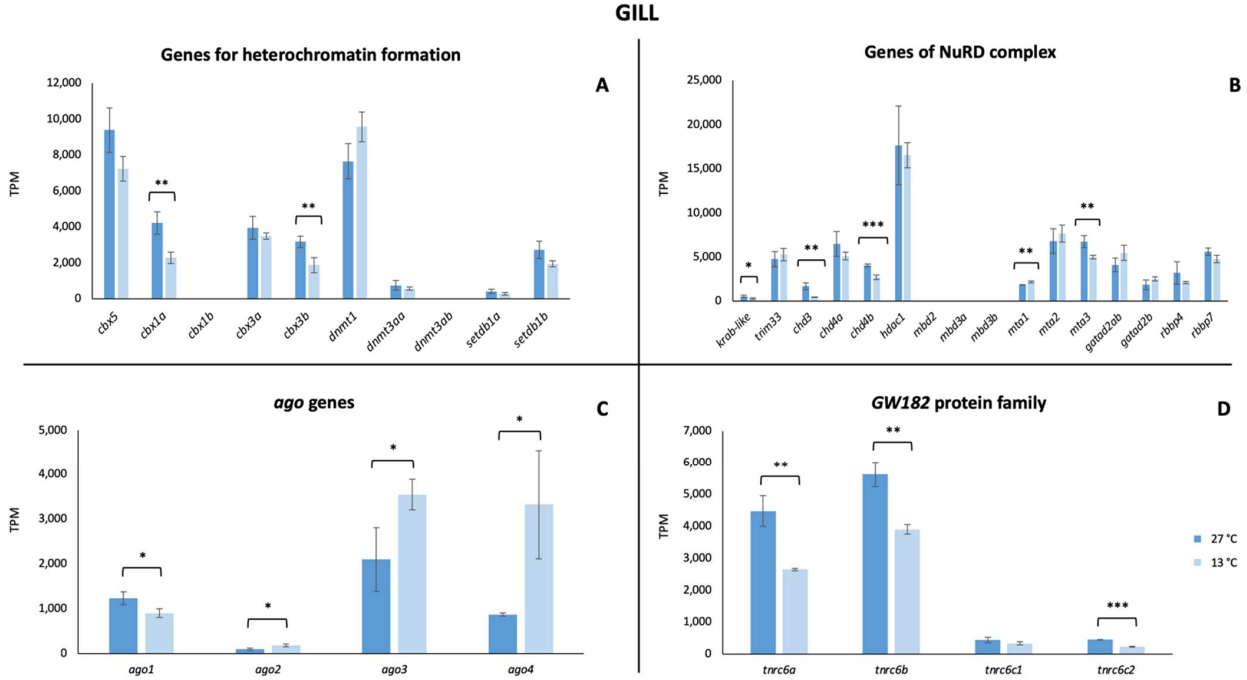

Figure 3. Expression genes involved in TE silencing mechanisms in the gill. (**A**) Expression values of genes encoding for the proteins involved in heterochromatin formation. (**B**) Expression values of genes encoding for proteins forming the NuRD complex. (**C**) Expression values of members of the gene subfamily *Argonaute*. (**D**) Expression values of genes belonging to the GW182 family proteins. Values are reported for both tested temperature conditions (27 °C and 13 °C). Statistically significant differences are presented as * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

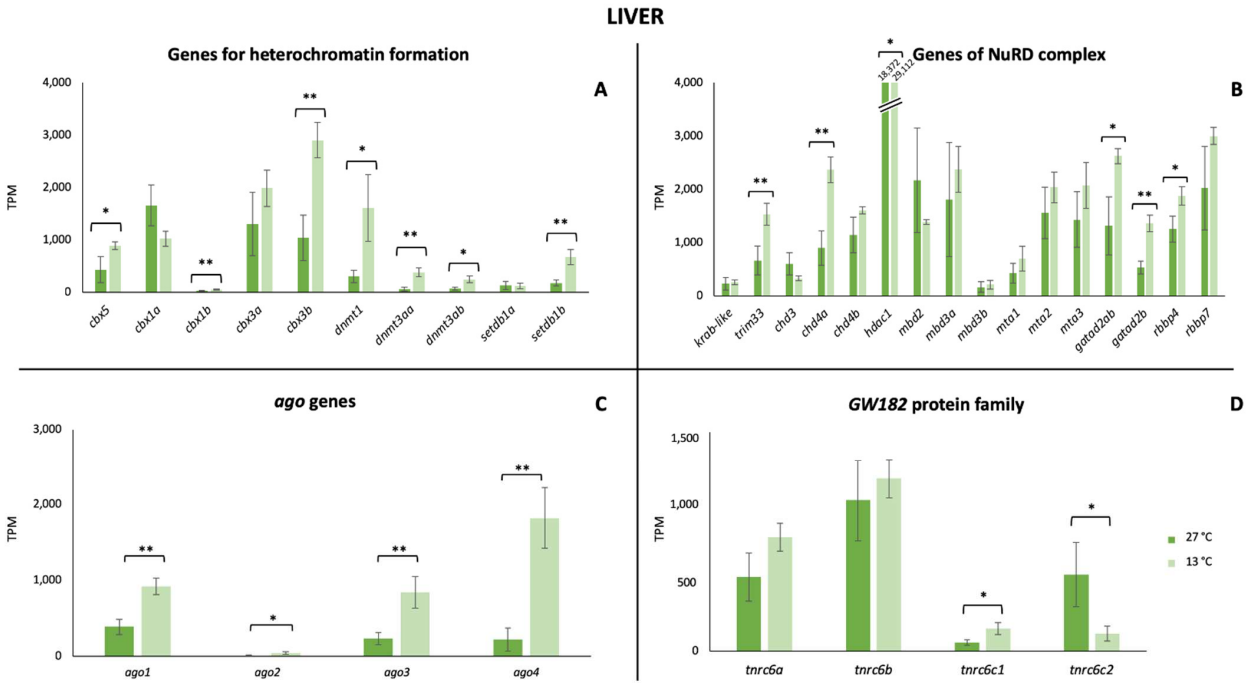

Figure 4. Expression genes involved in TE silencing mechanisms in the liver. (**A**) Expression values of genes encoding for proteins involved in heterochromatin formation. (**B**) Expression values of genes encoding for proteins forming the NuRD complex. (**C**) Expression values of members of the gene subfamily *Argonaute*. (**D**) Expression values of genes belonging to the GW182 family proteins. Values are reported for both the tested temperature conditions (27 °C and 13 °C). Statistically significant differences are presented as * for $p < 0.05$ and ** for $p < 0.01$.

4. Discussion

The ability of organisms to face changes of biotic and abiotic factors is mainly due to phenotypic plasticity that also requires modifications at the genome level. In particular, environmental changes can cause variations in the epigenetic status, leading to gene activation but also to the impairment of transposon silencing mechanisms inducing the activation of these mobile elements with the possible creation of genetic variability. Therefore, TEs represent a powerful adaptive response to environmental perturbations [7,8,27,42,43]. Although this issue has been largely studied in plants [20–22,25,44–47], information is still scarce in animals and, in particular, in vertebrates [27,43,48].

In this paper, we investigated the transcriptional TE response and the activity of genes involved in their silencing in the stenothermal fish *P. tetrazona*, a very popular ornamental freshwater fish native to Southeast Asia in places such as Malaysia and Borneo, places characterized by an equatorial climate; some of them are found in Thailand, Sumatra Island, and Cambodia [49]. This fish, living at 21–28 °C, tolerates small thermal fluctuations and experiences serious histopathological damages at low temperatures [38]. Overall, ambient temperature is an abiotic factor that acts on physiological and metabolic activities of poikilothermic teleosts, and in particular, on those of stenothermal fish [34–37]. Transcriptional analyses conducted by Liu and colleagues (2020) [38] on RNA-seq data obtained from *P. tetrazona* exposed at 27 °C (control) and 13 °C (test) have showed a high number of differentially expressed genes: in brain and liver the upregulated genes prevailed, differently from gill, in which the downregulated genes were the most represented. This is in line with the general idea that the genome epigenetic status varies in response to a stress, leading to changes in the expression levels. Our analysis of TE transcriptional contribution, performed on the same datasets, highlighted a different response in the three analyzed tissues. Apparently, the brain and gill showed the same behavior: in both the transcriptional activity of TEs did not vary between the two tested conditions. However, in the cerebral tissue, the activity of genes involved in heterochromatin formation, as well as *ago* genes and those related to GW182 family proteins, presented higher transcriptional levels at the stress condition. The lack of variation in TE transcriptional contribution in this tissue might be related to the silencing activity of these genes or to the short exposure period. In gill, although most of genes related to TE silencing mechanisms decreased their activity at 13 °C, no variation was observed in total TE transcriptional activity. However, for DNA transposons and Retro retroelements a slight increase of expression levels was appreciable at the stress condition. This might indicate that a chronic exposure is necessary to gain a remarkable TE variation.

Interestingly, a statistically significant increase of TE transcriptional activity was showed only in liver when fish were exposed at 13 °C. Our findings revealed that this increase was mainly due to few specific TEs that seem to undergo a remarkable transcriptional variation only in the hepatic tissue. This is in line with previous papers, reporting a relationship between a kind of stress and a specific TE type response. The mPing DNA transposon has been reported to be activated in response to salt stress in rice [44,46], the ONSEN retrotransposon as consequence of heat stress in *Arabidopsis* [45,47], and the nuclear import of Tam3 transposon related-transposase has been showed to increase after exposition to temperature decrease in *Antirrhinum* [20–22,25]. In addition, the different TE behavior might be related to their tissue-specific response. Indeed, Hunter and colleagues (2012) [50] have reported a stress-induced activation of particular TE families as ERV intracisternal-A particle (IAP), B2_RN SINE, and L1_RN in the rat hippocampus.

It is well known that the transcriptional and transpositional activity of TEs can have positive effects increasing genetic variability and leading to advantageous phenotypes. On the other side, TEs can provoke negative effects, deleterious for the host genome that in turn activates TE silencing mechanisms. In the tiger barb *P. tetrazona*, the high TE transcriptional levels at 13 °C in liver were accompanied by an increase in the transcriptional activity of genes related to heterochromatinization, the NuRD complex, *ago*, and the GW182 family proteins. This picture might suggest that the hepatic tissue has a quicker responsiveness

than other tissues. Indeed, it is a target organ for this kind of stress since several metabolic activities, also linked to body thermoregulation, occur in liver. Species respond to thermal conditions by modulating the metabolic rate to satisfy energy demand [51]. In *P. tetrazona* liver, Liu and colleagues (2020) [38] have reported an upregulation of cold-induced genes belonging to metabolism and biosynthesis pathways such as biosynthesis of steroid and fatty acids and tryptophan metabolism. The role of the hepatic tissue is extremely important since the tiger barb is a poikilothermic organism and also a stenothermal fish, tolerating small thermal fluctuations. The activation of pathways associated with organism response to stress conditions might be related to TE transcriptional activity. It is well-known that these genetic elements contribute in the up- or downregulation of nearby genes through *cis*- or *trans*-regulatory sequences located within TEs or TE-derived noncoding RNAs [26,52,53]. Alternatively, the decrease of methylation levels of genes involved in stress response might passively affect the heterochromatic state of nearby TEs leading to an increase of their transcriptional activity and consequently genes involved in their silencing are activated to counteract this epigenetic deregulation [54]. On the other side, liver might better tolerate the negative mutagenic effects related to TE activation thanks to its ability to regenerate after cell damage.

5. Conclusions

In conclusion, the analysis of the transcriptional activity of TEs and genes related to their silencing highlighted a clear response of these mobile elements in liver. Overall, the tissue-specific activation of TEs might represent a promising molecular tool, favoring species adaptation and resilience. Knowledge in this field could be useful in the perspective of wildlife species conservation since climate changes are threatening their biodiversity.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ani13010001/s1, Supplementary Table S1. Accession numbers of raw data downloaded from NCBI related to tested temperature conditions for brain, gill, and liver in *P. tetrazona*.; and Supplementary Table S2. Gene sequences analyzed in *P. tetrazona*. Details about the transcript length and its parts (5′ UTR, CDS, and 3′ UTR); aminoacidic length (aa); and information about its completeness and related accession numbers is reported.

Author Contributions: E.C. and F.C.: methodology and transcriptomic data analysis; A.C. and M.B.: project supervision; and M.A.B.: conceptualization and project administration. All authors discussed the results, wrote the manuscript, and commented the final version of the manuscript prior to submission. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article or Supplementary Materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. West-Eberhard, M.J. *Developmental Plasticity and Evolution*; Oxford University Press: New York, NY, USA, 2003.
2. Pigliucci, M.; Murren, C.J.; Schlichting, C.D. Phenotypic plasticity and evolution by genetic assimilation. *J. Exp. Biol.* **2006**, *209*, 2362–2367. [CrossRef] [PubMed]
3. Feil, R.; Fraga, M.F. Epigenetics and the environment: Emerging patterns and implications. *Nat. Rev. Genet.* **2012**, *13*, 97–109. [CrossRef] [PubMed]
4. Almojil, D.; Bourgeois, Y.; Falis, M.; Hariyani, I.; Wilcox, J.; Boissinot, S. The structural, functional and evolutionary impact of transposable elements in eukaryotes. *Genes* **2021**, *12*, 918. [CrossRef] [PubMed]
5. Kidwell, M.G.; Lisch, D.R. Perspective: Transposable elements, parasitic DNA, and genome evolution. *Evolution* **2001**, *55*, 1–24.
6. Kazazian, H.H. Mobile elements: Drivers of genome evolution. *Science* **2004**, *303*, 1626–1632. [CrossRef]
7. Chénais, B.; Caruso, A.; Hiard, S.; Casse, N. The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene* **2012**, *509*, 7–15. [CrossRef]

8. Casacuberta, E.; González, J. The impact of transposable elements in environmental adaptation. *Mol. Ecol.* **2013**, *6*, 1503–1517. [CrossRef]
9. McClintock, B. The significance of responses of the genome to challenge. *Science* **1984**, *226*, 792–801. [CrossRef]
10. Schultz, D.C.; Friedman, J.R.; Rauscher, F.J. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: The PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2a subunit of NuRD. *Genes Dev.* **2001**, *15*, 428–443. [CrossRef]
11. Carotti, E.; Carducci, F.; Greco, S.; Gerdol, M.; Di Marino, D.; Perta, N.; La Teana, A.; Canapa, A.; Barucca, M.; Biscotti, M.A. Transcriptional contribution of transposable elements in relation to salinity conditions in teleosts and silencing mechanisms involved. *Int. J. Mol. Sci.* **2022**, *23*, 5215. [CrossRef]
12. Czech, B.; Hannon, G.J. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends Biochem. Sci.* **2016**, *41*, 324–337. [CrossRef] [PubMed]
13. Dechaud, C.; Volff, J.N.; Schartl, M.; Naville, M. Sex and the TEs: Transposable elements in sexual development and function in animals. *Mob. DNA* **2019**, *10*, 42. [CrossRef] [PubMed]
14. Jakymiw, A.; Lian, S.; Eystathioy, T.; Li, S.; Satoh, M.; Hamel, J.C.; Fritzler, M.J.; Chan, E.K. Disruption of GW bodies impairs mammalian RNA interference. *Nat. Cell. Biol.* **2005**, *7*, 1267–1274. [CrossRef] [PubMed]
15. Liu, J.; Rivas, F.V.; Wohlschlegel, J.; Yates, J.R., III; Parker, R.; Hannon, G.J. A role for the P-body component GW182 in microRNA function. *Nat. Cell. Biol.* **2005**, *7*, 1261–1266. [CrossRef] [PubMed]
16. Meister, G.; Landthaler, M.; Peters, L.; Chen, P.Y.; Urlaub, H.; Lu¨hrmann, R.; Tuschl, T. Identification of novel Argonaute-associated proteins. *Curr. Biol.* **2005**, *15*, 2149–2155. [CrossRef]
17. Chekulaeva, M.; Mathys, H.; Zipprich, J.T.; Attig, J.; Colic, M.; Parker, R.; Filipowicz, W. miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1218–1226. [CrossRef]
18. Fabian, M.R.; Cieplak, M.K.; Frank, F.; Morita, M.; Green, J.; Srikumar, T.; Nagar, B.; Yamamoto, T.; Raught, B.; Duchaine, T.F.; et al. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1211–1217. [CrossRef]
19. Grandbastien, M.A.; Audeon, C.; Bonnivard, E.; Casacuberta, J.M.; Chalhoub, B.; Costa, A.P.; Le, Q.H.; Melayah, D.; Petit, M.; Poncet, C.; et al. Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet. Genome Res.* **2005**, *110*, 229–241. [CrossRef]
20. Hashida, S.N.; Kitamura, K.; Mikami, T.; Kishima, Y. Temperature shift coordinately changes the activity and the methylation state of transposon Tam3 in *Antirrhinum majus*. *Plant Physiol.* **2003**, *132*, 1207–1216. [CrossRef]
21. Hashida, S.N.; Kishima, Y.; Mikami, T. DNA methylation is not necessary for the inactivation of the Tam3 transposon at non-permissive temperature in *Antirrhinum*. *J. Plant. Physiol.* **2005**, *162*, 1292–1296. [CrossRef]
22. Hashida, S.N.; Uchiyama, T.; Martin, C.; Kishima, Y.; Sano, Y.; Mikami, T. The temperature-dependent change in methylation of the *Antirrhinum* transposon Tam3 is controlled by the activity of its transposase. *Plant Cell* **2006**, *18*, 104–118. [CrossRef] [PubMed]
23. Zeller, G.; Henz, S.R.; Widmer, C.K.; Sachsenberg, T.; Ratsch, G.; Weigel, D.; Laubinger, S. Stress-induced changes in the Arabidopsis thaliana transcriptome analyzed using whole-genome tiling arrays. *Plant J.* **2009**, *58*, 1068–1082. [CrossRef] [PubMed]
24. Tittel-Elmer, M.; Bucher, E.; Broger, L.; Mathieu, O.; Paszkowski, J.; Vaillant, I. Stress-induced activation of heterochromatic transcription. *PLoS Genet.* **2010**, *6*, e1001175. [CrossRef]
25. Fujino, K.; Hashida, S.N.; Ogawa, T.; Natsume, T.; Uchiyama, T.; Mikami, T.; Kishima, Y. Temperature controls nuclear import of Tam3 transposase in *Antirrhinum*. *Plant J.* **2011**, *65*, 146–155. [CrossRef] [PubMed]
26. Makarevitch, I.; Waters, A.J.; West, P.T.; Stitzer, M.; Hirsch, C.N.; Ross-Ibarra, J.; Springer, N.M. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* **2015**, *11*, e1004915.
27. Piacentini, L.; Fanti, L.; Specchia, V.; Bozzetti, M.P.; Berloco, M.; Palumbo, G.; Pimpinelli, S. Transposons, environmental changes, and heritable induced phenotypic variability. *Chromosoma* **2014**, *123*, 345–354. [CrossRef]
28. Feiner, N. Accumulation of transposable elements in Hox gene clusters during adaptive radiation of Anolis lizards. *Proc. Biol. Sci.* **2016**, *283*, 20161555.
29. Carducci, F.; Biscotti, M.A.; Forconi, M.; Barucca, M.; Canapa, A. An intriguing relationship between teleost *Rex3* retroelement and environmental temperature. *Biol. Lett.* **2019**, *15*, 20190279. [CrossRef]
30. Carotti, E.; Carducci, F.; Canapa, A.; Barucca, M.; Greco, S.; Gerdol, M.; Biscotti, M.A. Transposable elements and teleost migratory behaviour. *Int. J. Mol. Sci.* **2021**, *22*, 602. [CrossRef]
31. Rhee, J.S.; Choi, B.S.; Kim, J.; Kim, B.M.; Lee, Y.M.; Kim, I.C.; Kanamori, A.; Choi, I.Y.; Schartl, M.; Lee, J.S. Diversity, distribution, and significance of transposable elements in the genome of the only selfing hermaphroditic vertebrate *Kryptolebias marmoratus*. *Sci. Rep.* **2017**, *7*, 40121. [CrossRef]
32. Yuan, Z.; Liu, S.; Zhou, T.; Tian, C.; Bao, L.; Dunham, R.; Liu, Z. Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. *BMC Genom.* **2018**, *19*, 141. [CrossRef] [PubMed]
33. Auvinet, J.; Graça, P.; Ghigliotti, L.; Pisano, E.; Dettaï, A.; Ozouf-Costaz, C.; Higuet, D. Insertion hot spots of DIRS1 retrotransposon and chromosomal diversifications among the Antarctic teleosts Nototheniidae. *Int. J. Mol. Sci.* **2019**, *20*, 701. [CrossRef] [PubMed]

34. Fry, F.J. Responses of vertebrate poikilotherms to temperature. In *Thermobiology*; Rose, A.H., Ed.; Academic Press: London, UK, 1967; pp. 375–409.
35. Huey, R.B.; Kingsolver, J.G. Evolution of thermal sensitivity of ectotherm performance. *Trends Ecol. Evol.* **1989**, *4*, 131–135. [CrossRef]
36. Cossins, A. *Temperature Biology of Animals*; Springer Science & Business Media: Berlin, Germany, 2012.
37. Bernal, M.A.; Schmidt, E.; Donelson, J.M.; Munday, P.L.; Ravasi, T. Molecular response of the brain to cross-generational warming in a coral reef fish. *Front. Mar. Sci.* **2022**, *9*, 784418. [CrossRef]
38. Liu, L.; Zhang, R.; Wang, X.; Zhu, H.; Tian, Z. Transcriptome analysis reveals molecular mechanisms responsive to acute cold stress in the tropical stenothermal fish tiger barb (*Puntius tetrazona*). *BMC Genom.* **2020**, *21*, 737. [CrossRef]
39. Manni, M.; Berkeley, M.R.; Seppey, M.; Zdobnov, E.M. BUSCO: Assessing Genomic Data Quality and Beyond. *Curr. Protoc.* **2021**, *1*, e323. [CrossRef]
40. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef] [PubMed]
41. Biscotti, M.A.; Gerdol, M.; Canapa, A.; Forconi, M.; Olmo, E.; Pallavicini, A.; Barucca, M.; Schartl, M. The lungfish transcriptome: A glimpse into molecular evolution events at the transition from water to land. *Sci. Rep.* **2016**, *6*, 21571. [CrossRef]
42. Lisch, D. How important are transposons for plant evolution? *Nat. Rev. Genet.* **2013**, *14*, 49–61. [CrossRef]
43. Schrader, L.; Schmitz, J. The impact of transposable elements in adaptive evolution. *Mol Ecol.* **2019**, *28*, 1537–1549. [CrossRef]
44. Naito, K.; Zhang, F.; Tsukiyama, T.; Saito, H.; Hancock, C.N.; Richardson, A.O.; Okumoto, Y.; Tanisaka, T.; Wessler, S.R. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* **2009**, *461*, 1130–1134. [CrossRef] [PubMed]
45. Pecinka, A.; Dinh, H.Q.; Baubec, T.; Rosa, M.; Lettner, N.; Mittelsten Scheid, O. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *Plant Cell* **2010**, *22*, 3118–3129. [CrossRef]
46. Yasuda, K.; Ito, M.; Sugita, T.; Tsukiyama, T.; Saito, H.; Naito, K.; Teraishi, M.; Tanisaka, T.; Okumoto, Y. Utilization of transposable element as a novel genetic tool for modification of the stress response in rice. *Mol. Breed.* **2013**, *32*, 505–516. [CrossRef] [PubMed]
47. Cavrak, V.V.; Lettner, N.; Jamge, S.; Kosarewicz, A.; Bayer, L.M.; Mittelsten Scheid, O. How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet.* **2014**, *10*, e1004115. [CrossRef] [PubMed]
48. Pappalardo, A.M.; Ferrito, V.; Biscotti, M.A.; Canapa, A.; Capriglione, T. Transposable elements and stress in vertebrates: An overview. *Int. J. Mol. Sci.* **2021**, *22*, 1970. [CrossRef] [PubMed]
49. Barik, M.; Bhattacharjee, I.; Ghosh, A.; Chandra, G. Larvivorous potentiality of *Puntius tetrazona* and *Hyphessobrycon* rosaceus against *Culex vishnui* subgroup in laboratory and field based bioassay. *BMC Res. Notes* **2018**, *11*, 804. [CrossRef]
50. Hunter, R.G.; Murakami, G.; Dewell, S.; Seligsohn, M.; Baker, M.E.; Datson, N.A.; McEwen, B.S.; Pfaff, D.W. Acute stress and hippocampal histone H3 lysine 9 trimethylation, a retrotransposon silencing response. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17657–17662. [CrossRef]
51. Alberto-Payet, F.; Lassus, R.; Isla, A.; Daufresne, M.; Sentis, A. Nine years of experimental warming did not influence the thermal sensitivity of metabolic rate in the medaka fish *Oryzias latipes*. *Freshw. Biol.* **2022**, *67*, 577–585. [CrossRef]
52. Horváth, V.; Merenciano, M.; González, J. Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends Genet.* **2017**, *11*, 832–841. [CrossRef] [PubMed]
53. Zovoilis, A.; Cifuentes-Rojas, C.; Chu, H.P.; Hernandez, A.J.; Lee, J.T. Destabilization of B2 RNA by EZH2 Activates the Stress Response. *Cell* **2016**, *167*, 1788–1802. [CrossRef]
54. Slotkin, R.K.; Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **2007**, *8*, 272–285. [CrossRef] [PubMed]

4.Analyses on proteins involved in NuRD complex recruitment in fish

The activity of transposable elements is finely regulated through the involvement of proteins that act at the chromatin level to generate a heterochromatic status to silence TE transcription. This is certainly one of the most fascinating research fields that still presents several gaps. The findings presented in the papers 5 and 6 suggested that the NuRD complex is involved in TE regulation also in fish. In sarcopterygians, this complex is recruited by the binding between KAP1 and KRAB-ZNF proteins that are part of the largest family of transcriptional repressors in mammals. In actinopterygians, both KAP1 and KRAB-ZFPs are absent in their genomes (Baker et al. 2017). Analysing the members of TRIM family to which KAP1 belongs to, I proposed the protein TRIM33 as possible substitute of KAP1 in the NuRD fish system. Indeed, TRIM33 has the same protein domain architecture of KAP1 and is widely present in ray-finned fish. Moreover, for the first time in fish I identified a zinc finger protein having a KRAB-like domain at N-ter. The results of docking simulation support the interaction between TRIM33 and KRAB-like in *Anguilla anguilla* corroborating the hypothesis of these two proteins as possible substitutes of KAP1/KRAB-ZFP of sarcopterygians. Therefore, to test this result also at experimental level, I spent part of my PhD project at the Virology and Genetics Laboratory lead by Prof. Didier Trono at École Polytechnique Fédérale de Lausanne (EPFL) in Switzerland.

In order to perform co-Immunoprecipitation (co-IP) analyses, the sequences encoding TRIM33 and KRAB-like proteins were inserted in the pTRE lentivector using Gateway cloning system (Thermo Fisher Scientific Inc.). The target gene recombines close to the TAG sequence (HA and FLAG for KRAB-like and TRIM33, respectively), useful for co-IP steps, and upstream presents an inducible promoter for Doxycycline.

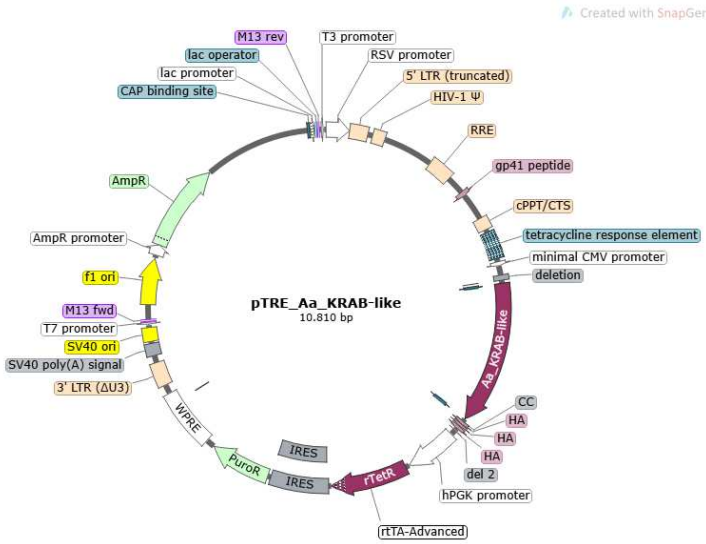

Figure 4. Plamid map of pTRE_Aa_KRAB-like lentivector.

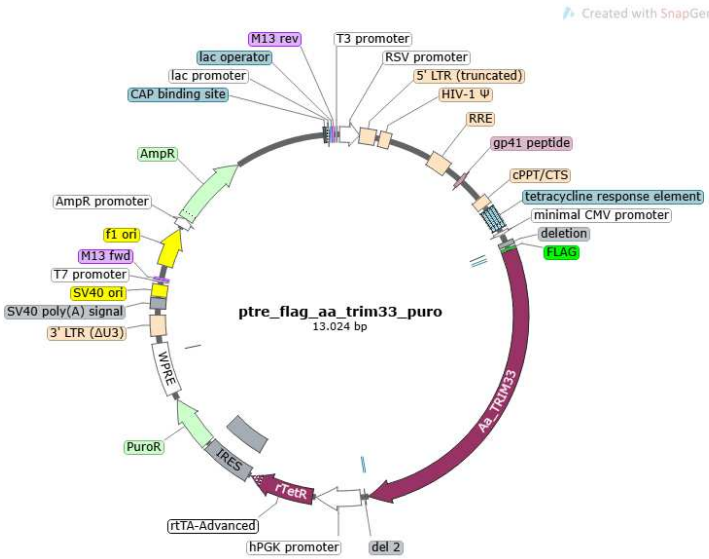

Figure 5. Plamid map of pTRE_Aa_KRAB-like lentivector

The obtained lentivector were transfected in 293T cells. The transfection was performed using FuGene reagent (Promega), a nonliposomal reagent that transfects DNA into a wide variety of cell lines with high efficiency and low toxicity. After having ensured the success of the transfection by using cells transfected with a plasmid expressing GFP as a control sample, protein extraction and then co-IP analyses were performed. For this purpose, antiFLAG magnetic beads were used to precipitate TRIM33 together with the KRAB-like having HA as tag.

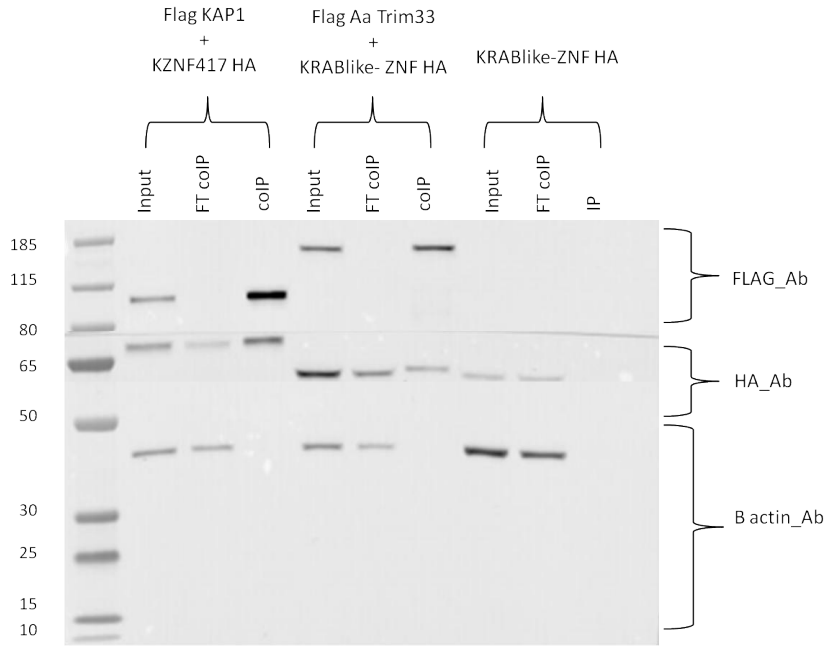

Figure 6. Western Blot analysis of protein co-immunoprecipitation (input: total protein extraction; FT co-IP: flow through co-IP; IP: immunoprecipitation; FLAG_Ab: antibody anti FLAG; HA_Ab: antibody anti HA; β actin_Ab: antibody anti β actin).

The co-IP result confirms the link between TRIM33 and KRAB-like. Because of false positives, positive and negative controls were used in parallel. In particular, the positive control was represented by the known interaction between KAP1 and KRAB-ZFP. The negative controls highlighted that the positive result obtained in the co-IP of TRIM33 and KRAB-like is due to the binding between the two proteins of interest and not to technical issues as beads employed.

Recently, the research team led by Prof. Cedric Feschotte has identified specific ZNF proteins in zebrafish, named Fish N-terminal Zinc-finger associated (FiNZ) domain, and widely and abundantly distributed in cyprinids. Indeed, Chang and colleagues (2022) have performed a

characterization of the genomic landscape and embryonic expression of zebrafish TEs and have highlighted their high diversity not only at sequence level but also and the transcriptional level during early developmental stages. Using the same dataset, Wells and colleagues (2022) have reported that zebrafish FiNZ-ZNFs are expressed at the onset of zygotic genome activation as in mammals. Moreover, performing the blocking of FiNZ-ZNF translation using morpholinos, they have observed a global de-repression of young and transcriptionally active TEs during early zebrafish embryogenesis. Overall, these data suggest a connection between FiNZ-ZNFs and TE expansion throughout animal evolution and an independent distribution of ZNFs in fish and mammals to repress TE during early embryogenesis (Wells et al. 2022).

During the period abroad, I also tested the interaction between *D. rerio* TRIM33 and FiNZ-ZNF proteins both using docking simulation (Figure 7) and the same experimental approach carried out for *A. anguilla*.

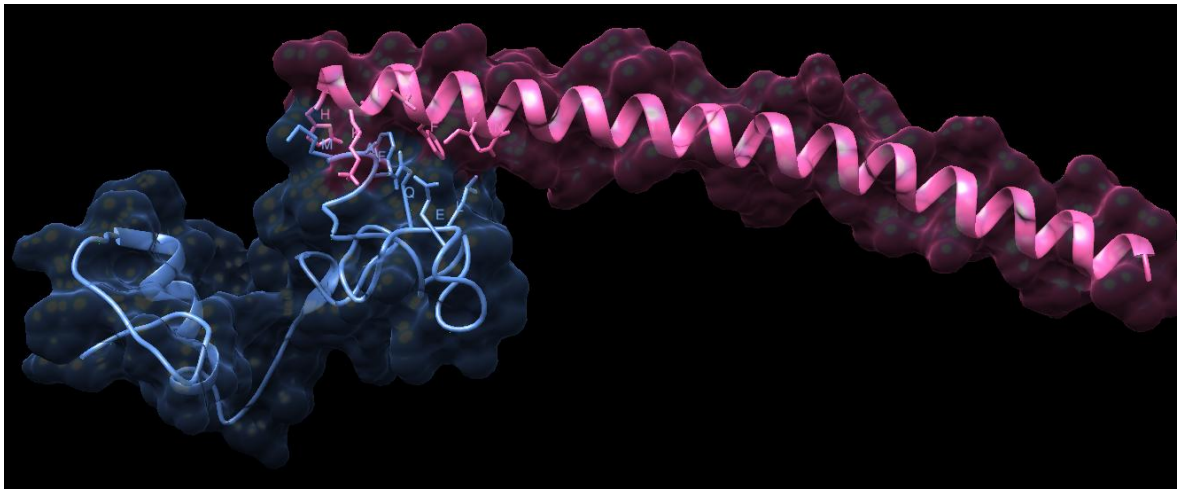

Figure 7. Docking simulation of RBCC domain of *D. rerio* TRIM33 and FiNZ domain (without ZNF tail). Amino acids selected in RBCC domain of *D. rerio* TRIM33Dr: 364(H), 367(K), 368(I), 370(V), 371(F), 374(I), 375(N). Amino acids selected for FiNZ domain: 1(M), 2(A), 3(F), 22(E), 24(L), 25(Q). Pink: RBCC domain of *D. rerio* TRIM33; blue: FiNZ domain.

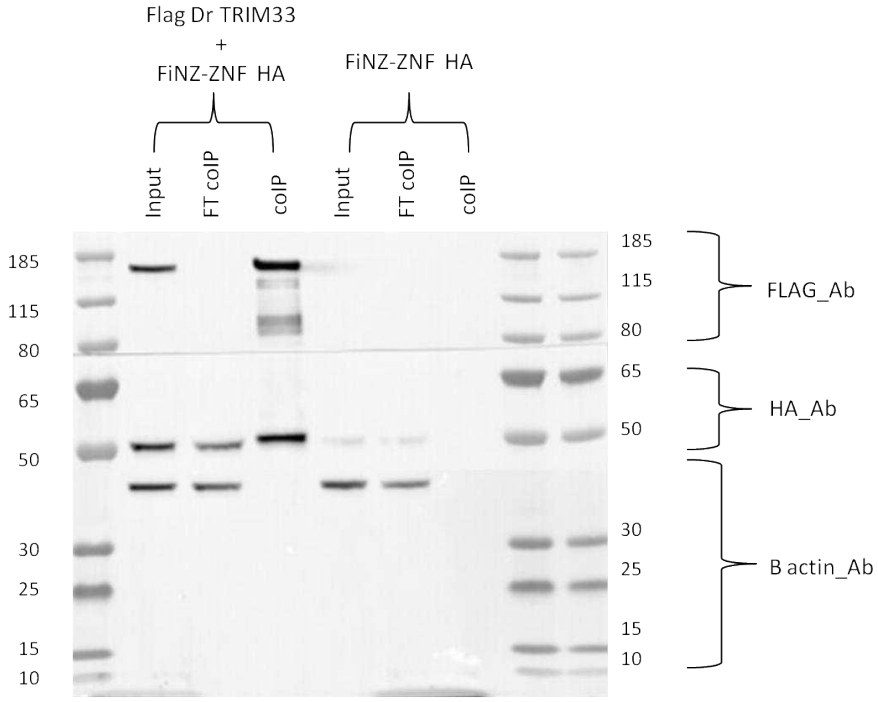

Figure 8. Western Blot analysis of protein co-immunoprecipitation (input: total protein extraction; FT co-IP: flow through co-IP; FLAG_Ab: antibody anti FLAG; HA_Ab: antibody anti HA; β actin_Ab: antibody anti β actin).

Also in this case the performed analyses highlighted the interaction between TRIM33 and FiNZ-ZNF proteins supporting the functioning of this system in zebrafish (Figure 8).

5. Discussion and conclusions

Genome size varies considerably across eukaryotes and it is correlated neither with the number of genes, nor with the morpho-functional complexity of a species. It is now recognized that a significant impact is due to the relative abundance and activity of TEs. Indeed, these genetic elements constitute a large fraction of the repetitive genomic DNA of eukaryotes. Among vertebrates, species belonging to the amphibian order Caudata display a genome size ranging from 13.89 to 120.60 Gb (Gregory 2020). The highest values belong to salamanders, which together with lungfish, are the record holders for the largest genomes among extant vertebrates. Evidence suggest that the gigantism of urodele genome is not attributable to polyploidy but rather to a reduced DNA loss rate and to the accumulation of TEs, in particular of LTR retrotransposons (Sun et al. 2012a, b; Sun and Mueller 2014). During my PhD, I analysed the activity of transposable elements and that of 33 genes encoding proteins involved in the TE host silencing mechanisms in the transcriptomes of the Chinese fire-bellied newt *Cynops orientalis*. Moreover, I compared these data with those obtained in the coelacanth *Latimeria menadoensis* (Pallavicini et al. 2013) and the lungfish *Protopterus annectens* (Biscotti et al. 2016), two species characterized by significant differences in the activity of TEs and in genome size (Wang et al. 2021; Amemiya et al. 2013). The data obtained showed a higher TE activity

in the newt, with a major impact of non-LTR retroelements. The transcriptional activity of genes involved in TE silencing mechanisms highlighted that, although these mechanisms are active also in adults, differences in TE activity evidenced between the species compared in this study may be due to the presence of old and inactive copies. Moreover, the high transcriptional levels of genes involved in TE silencing as those of NuRD complex in gonads suggested a protective role of these mechanisms to preserve genome integrity in germlines.

Actinopterygii is another taxon of vertebrates extremely intriguing to investigate the role of TEs in shaping genomes. Indeed, ray-finned fish are characterized by a high diversity of species that have adapted to live in a wide range of environments. Therefore, the evaluation of the impact of mobile elements may contribute not only to unravel the biodiversity of this class, but also to unveil the reason of their extreme adaptability to widely different aquatic environmental conditions. In this context, migrating fish represent a particularly interesting case study. In particular, diadromous organisms have always attracted scientific attention due to their spectacular migratory routes between SW and FW during their life cycle that have requested the evolution of an extraordinary physiological plasticity to cope the substantial differences in salinity and temperature encountered. I analysed the genomes of 24 fish

species, of which 15 having a migratory behaviour, highlighting variable TE content both in the bony and non-bony fish genomes investigated and a positive correlation between TE content and the assembled genome size. In the TE relative contribution analyses, LINE retroelements showed the highest impact in sea lamprey and elephant shark while DNA transposons presented the higher relative abundance in ray-finned fish with some exceptions in the sturgeon *Acipenser ruthenus*, the non-teleost *Lepisosteus oculatus*, and seven teleost species in which retroelements prevailed. Overall, the lack of correlation between the patters of TE relative abundance and phylogeny of considered species suggested that the accumulation of specific TEs might be related to the evolutionary adaptation of species to specific environmental niches. Indeed, variation partitioning analyses evidenced a statistically significant influence of migratory behaviour on quantitative differences of DNA transposons between catadromous and amphidromous species. The effect of migration was more pronounced on the quantitative difference reported for SINE retroelements in the comparison between anadromous and catadromous fish species independently from their phylogenetic position. This aspect is likely due to the substantial environmental changes faced by diadromous species during their migratory routes.

In line with these results, I investigated the transcriptional activity of TEs in response to abiotic factors such as salinity and temperature. Indeed, an increasing number of papers reported changes in TE activity to regulate mechanisms responsible for species adaptation to specific environments (Fujino et al. 2021; Carducci et al. 2019b; Carducci et al. 2020; Carotti et al. 2021; Pappalardo et al. 2021; Liu et al. 2020). Salinity is an abiotic factor that influences the adaptation of marine and freshwater organisms, and its changes can also represent a possible stress that threatens their survival. The giant marbled eel *Anguilla marmorata* is a catadromous species widely distributed in tropical, subtropical, and temperate areas that moves from freshwater to salt water for spawning (Steendam et al. 2020). The chum salmon *Oncorhynchus keta* is an anadromous species that spends much of its life in salt water and goes up rivers to spawn (Lee et al. 2020; Beechie et al. 2021). The marine medaka *Oryzias melastigma* is an euryhaline species able to survive in environments with different salinities. The analyses of TE activity were performed in gills the primary organs that are highly sensitive to salinity level and play crucial roles in physical processes such as gas exchange, nitrogenous waste excretion, and acid-base balance (Lee et al. 2020; Cao et al. 2021). These data revealed an interesting variation in the case of juvenile eels, commonly adapted to salty water, when exposed to brackish and freshwater

conditions. Moreover, the expression assessment of genes involved in TE silencing mechanisms (six in heterochromatin formation, fourteen known to be part of the nucleosome remodeling deacetylase (NuRD) complex, and four of the *Argonaute* subfamily) unveiled that they are active. Interestingly, I reported for the first time the transcriptional activity of genes involved in NuRD complex in fish, known only in sarcopterygians so far. Given the absence of TRIM28/KAP1 and KRAB-ZFPs in ray-finned fish, I identified a krüppel-associated box (KRAB)-like domain specific to actinopterygians that, together with TRIM33 (another protein of TRIM gene family), might allow the functioning of NuRD complex also in this evolutionary lineage. The possible interaction between these two proteins was supported by structural prediction analyses.

In addition to salinity, temperature is another abiotic factor affecting metabolic and physiological activities in poikilothermic organisms as fish. In particular, ambient temperature might be a critical factor for stenothermal fish in which thermal fluctuations are less tolerated. The tiger barb *P. tetrazona* is a freshwater fish belonging to the family Cyprinidae, living at 21–28 °C. I investigated the transcriptional activity of TEs and genes encoding proteins involved in their silencing in the brain, gill, and liver. The results obtained pointed out a quicker responsiveness of TEs in liver. Indeed, this organ might

represent a target for this kind of stress since species respond to thermal conditions by modulating the body thermoregulation and the metabolic rate to satisfy energy demand (Alberto-Payet et al 2022). This aspect is extremely important in *P. tetrazona* that is a stenothermal fish able to tolerate small thermal fluctuations. Moreover, the data obtained highlighted the activity of genes encoding proteins of NuRD complex as well as TRIM33 and KRAB-like also in this fish. The involvement of proteins encoded by these two genes in the recruitment of NuRD complex, instead of the sarcopterygian KAP1 and KRAB-ZFPs, was supported by the CoIP analyses that evidenced an interaction between them in vitro in agreement with the docking simulations.

In conclusion, the results of this PhD thesis evidenced that TEs represent an important functional fraction of the genome. The amount of particular TE types might be responsible for adaptation of species to specific environments. Moreover, TEs can have a potential role in the rapid adaptation of species, thus representing a promising molecular tool for their resilience. Indeed, environmental changes can cause variation in the epigenetic status of TEs determining changes in their transcriptional and transpositional activity. This might provoke an increase in the genetic variability that allows the creation of advantageous phenotypes. On the other side, TEs can provoke negative

effects, deleterious for the host genome and thus TE silencing mechanisms are activated by the host genome to control their activity.

Therefore, TEs undoubtedly have had a key role in determining the high diversity of ray-finned fish and their evolutionary success. Moreover, they are important molecular tools that would allow fish to survive and to cope with changes in environmental conditions.

6. References

1. Ahmed, M. and Liang, P. Transposable elements are a significant contributor to tandem repeats in the human genome. *Comp. Funct. Genomics*, **2012**, 2012, 947089.
2. Alberto-Payet, F.; Lassus, R.; Isla, A. et al. Nine years of experimental warming did not influence the thermal sensitivity of metabolic rate in the medaka fish *Oryzias latipes*. *Freshw. Biol*., **2022**, 67, 577–585.
3. Alexandrou, M. A.; Swartz, B. A.; Matzke, N. J. et al. Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Mol. Phylogenet. Evol*., **2013**, 69, 514–523.;
4. Amemiya, C.; Alföldi, J.; Lee, A. et al. The African coelacanth genome provides insights into tetrapod evolution. *Nature*, **2013**, 496, 311–316.
5. Auvinet, J.; Graça, P.; Belkadi, L. et al. Mobilization of retrotransposons as a cause of chromosomal diversification and rapid speciation: The case for the Antarctic teleost genus *Trematomus*. *BMC Genom.*, **2018**, 19, 339.
6. Baker, Z.; Schumer, M.; Haba, Y. et al. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife*, **2017**, 6, e24133.
7. Bao, Z. and Eddy, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res*., **2002**, 12, 1269–1276.
8. Beechie, T. J.; Nicol, C.; Fogel, C. et al. Chum Salmon Life-Cycle Model Description and Results for the Chehalis River Basin. *NOAA Contract Report NMFS-NWFSC-CR.* **2021**, U.S. Department of Commerce: Washington, DC, USA.
9. Bernardi, G. Structural and evolutionary genomics. Lausanne, New York, Oxford, Shannon, Singapore, Tokyo: *Elsevier* Amsterdam. **2004**, pp. 434+XV.
10. Biscotti, M. A.; Olmo, E. and Heslop-Harrison, J. S. Repetitive DNA in eukaryotic genomes. *Chromosome Res*., **2015**, 23(3), 415–420.
11. Biscotti, M. A.; Gerdol, M.; Canapa, A. et al. The lungfish transcriptome: a glimpse into molecular evolution events at the transition from water to land. *Sci. Rep*., **2016**, 6, 21571.

12. Biscotti, M. A.; Canapa, A.; Forconi, M. et al. The small non-coding RNA processing machinery of two living fossil species, lungfish and coelacanth, gives new insights into the evolution of the Argonaute protein family. *Genome Biol. Evol*., **2017**, 9, 438–453.
13. Biscotti, M. A.; Carducci, F.; Olmo, E. et al. Vertebrate genome size and the impact of transposable elements in genome evolution. In Evolution, Origin of Life, Concepts and Methods. *Springer*, Cham. **2019**, (pp. 233–251).
14. Biscotti, M. A.; Carducci, F.; Barucca, M. et al. The transcriptome of the newt *Cynops orientalis* provides new insights into evolution and function of sexual gene networks in sarcopterygians. *Sci. Rep.*, **2020**, 10(1), 5445.
15. Bourque, G.; Burns, K. H.; Gehring, M. et al. Ten things you should know about transposable elements. *Genome Biol.*, **2018**, 19, 199.
16. Braun, J. E.; Huntzinger, E. and Izaurralde, E. The role of GW182 proteins in miRNA-mediated gene silencing. *Adv. Exp. Med. Biol*., **2013**, 768, 147–163.
17. Canapa, A.; Barucca, M.; Biscotti, M. A. et al. Transposons, Genome size, and evolutionary insights in animals. *Cytogenet. Genome Res.,* **2015**, 147, 217–239.
18. Canapa, A.; Biscotti, M. A.; Barucca, M. et al. Shedding light upon the complex net of genome size, genome composition and environment in chordates. *Eur. Zool. J*., **2020**, 87, 192–202.
19. Cao, Q.; Wang, H.; Fan, C. et al. Environmental salinity influences the branchial expression of TCR pathway related genes based on transcriptome of a catadromous fish. *Comp. Biochem. Physiol. Part D Genomics Proteomics*, **2021**, 38, 100815.
20. Capy, P.; Gasperi, G.; Biémont, C. et al. Stress and transposable elements: Co-evolution or useful parasites? *Heredity*, **2000**, 85, 101–106.
21. Carducci, F.; Biscotti, M. A.; Barucca, M. et al. Transposable elements in vertebrates: Species evolution and environmental adaptation. *Eur. Zool. J*., **2019a**, 86, 497–503.

22. Carducci, F.; Biscotti, M. A.; Forconi, M. et al. An intriguing relationship between teleost Rex3 retroelement and environmental temperature. *Biol. Lett.*, **2019b**, 15, 20190279.
23. Carducci, F.; Barucca, M.; Canapa, A. et al. Mobile elements in ray-finned fish genomes. *Life*, **2020**, 10, 221.
24. Carducci, F.; Carotti, E.; Gerdol, M. et al. Investigation of the activity of transposable elements and genes involved in their silencing in the newt *Cynops orientalis*, a species with a giant genome. *Sci. Rep.*, **2021**, 11, 14743.
25. Carotti, E.; Carducci, F.; Canapa, A. et al. Transposable Elements and Teleost Migratory Behaviour. *Int. J. Mol. Sci.*, **2021**, 22, 602.
26. Carotti, E.; Carducci, F.; Greco, S. et al. Transcriptional contribution of transposable elements in relation to salinity conditions in teleosts and silencing mechanisms involved. *Int. J. Mol. Sci.* **2022**, 23, 5215.
27. Carotti, E.; Carducci, F.; Canapa, A. et al. Transposable element tissue-specific response to temperature stress in the stenothermal fish *Puntius tetrazona*. *Animals,* **2023**, 13, 1.
28. Casacuberta, E. and González, J. The impact of transposable elements in environmental adaptation. *Mol. Ecol.*, **2013**, 6, 1503–1517.
29. Chalopin, D.; Naville, M.; Plard, F. et al. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.*, **2015**, 7(2), 567–580.
30. Chang, N. C.; Rovira, Q.; Wells, J. N. et al. Zebrafish transposable elements show extensive diversification in age, genomic distribution, and developmental expression. *Genome Res.*, **2022**, 32(7), 1408–1423.
31. Chapman, F. A. Evaluation of commercially-formulated diets for feeding Tiger barb, *Puntius tetrazona*. *J. Appl. Aquaculture*, **1997**, 1(7), 69–74.
32. Christoffels, A.; Koh, E. G. L.; Chia, J. M. et al. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.*, **2004**, 21, 1146–1151.
33. Chu, C.; Nielsen, R. and Wu, Y. REPdenovo: inferring de novo repeat motifs from short sequence reads. *PLOS ONE*, **2016**, 11, e0150719.
34. Cloutier, R. and Ahlberg, P. E. Morphology, characters, and the interrelationships of basal sarcopterygians. *In Interrelationship of*

fishes. Edited by Stiassny LJ, Parenti LR, Johson G. San Diego, CA: Academic; **1996**, 445–479.

35. Dehal, P. and Boore, J. L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.*, **2005**, 3, e314.
36. Dos Santos, R. S.; Galina, A. and Da-Silva W. S. Cold acclimation increases mitochondrial oxidative capacity without inducing mitochondrial uncoupling in goldfish white skeletal muscle. *Biol Open.*, **2013**, 2(1), 82–87.
37. Ecco, G.; Cassano, M.; Kauzlaric A. et al. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. *Dev Cell.*, **2016**, 36(6), 611–623.
38. Emerson, R. O. and Thomas, J. H. Adaptive evolution in zinc finger transcription factors. *PLoS Genet.*, **2009**, 5, e1000325.
39. Evans, D. H.; Piermarini, P. M. and Choe, K. P. The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. *Physiol Rev.*, **2005**, 85, 97–177.
40. Friedman, J. R.; Condon, N. E. and Drazen J. C. Gill surface area and metabolic enzyme activities of demersal fishes associated with the oxygen minimum zone off California. *Limnol Oceanogr.*, **2012**, 57, 1701–1710.
41. Fujino, K.; Hashida, S. N.; Ogawa, T. et al. Temperature controls nuclear import of Tam3 transposase in Antirrhinum. *Plant Journal*, **2021**, 65,146–155.
42. Garcia Guerreiro, M. P. What makes transposable elements move in the Drosophila genome? *Heredity*, **2012**, 108, 461–468.
43. Garrido-Ramos M. A. Satellite DNA: An Evolving Topic. *Genes (Basel).*, **2017**, 8(9), 230.
44. Goerner-Potvin, P. and Bourque, G. Computational tools to unmask transposable elements. *Nat. Rev. Genet.*, **2018**, 19, 688–704.
45. Goubert, C.; Modolo, L.; Vieira, C. et al. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito *(Aedes aegypti*). *Genome Biol. Evol.*, **2015**, 7, 1192–1205.

46. Gracey, A. Y.; Fraser, E. J.; Li, W. et al. Coping with cold: an integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. *Proc. Natl. Acad. Sci. USA.*, **2004**, 101(48), 16970–16975.
47. Gregory, T. R. Animal Genome Size Database. http://www.genomesize.com. (**2020**).
48. Guo, R.; Li, Y. R.; He, S. RepLong: de novo repeat identification using long read sequencing data. *Bioinformatics*, **2018**, 34, 1099–1107.
49. Helleboid, P. Y.; Heusel, M.; Duc, J. et al. The interactome of KRAB zinc finger proteins reveals the evolutionary history of their functional diversification. *The EMBO journal*, **2019**, 38(18), e101220.
50. Heslop-Harrison, J. S. and Schwarzacher, T. Nucleosomes and centromeric DNA packaging. *Proc. Natl. Acad. Sci. USA*, **2013**, 110(50), 19974–19975.
51. Jacobs, F.M.; Greenberg, D.; Nguyen, N. et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*, **2014**, 516, 242–245.
52. Jakymiw, A.; Lian, S.; Eystathioy, T. et al. Disruption of GW bodies impairs mammalian RNA interference. *Nat. Cell Biol.*, **2005**, 7(12), 1267–1274.
53. Kassahn, K. S.; Caley, M. J.; Ward, A. C. et al. Heterologous microarray experiments used to identify the early gene response to heat stress in a coral reef fish. *Mol. Ecol.*, **2007**, 16(8), 1749–1763.
54. Kirankumar, S. and Pandian, T.J. Production of androgenetic tiger barb, *Puntius tetrazona*. *Aquaculture*, **2003**, 228(1–4), 37–51.
55. Laurin, M.; Canoville, A.; Struble, M. et al. Early genome size increase in urodeles. *C.R. Palevol.*, **2016**, 15, 74–82.
56. Lee, S. Y.; Lee, H. J. and Kim, Y. K. Comparative transcriptome profiling of selected osmotic regulatory proteins in the gill during seawater acclimation of chum salmon (*Oncorhynchus keta*) fry. *Sci. Rep.*, **2020**, 10, 1987.
57. Leknes, I. L. Goblet cell types in intestine of tiger barb and black tetra (Cyprinidae, Characidae: Teleostei). *Anat. Histol. Embryol.*, **2014**, 43(5), 352–360.

58. Liedtke, H. C.; Müller, H.; Rödel, M. O. et al. No ecological opportunity signal on a continental scale? Diversification and life-history evolution of African true toads (anura: bufonidae). *Evol.*, **2016**, 70, 1717–1733.
59. Liedtke, H. C.; Gower, D. J.; Wilkinson, M. et al. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. *Nat. Ecol. Evol.*, **2018**, 2, 1792–1799.
60. Liu, L.; Zhang, R.; Wang, X. et al. Transcriptome analysis reveals molecular mechanisms responsive to acute cold stress in the tropical stenothermal fish tiger barb (*Puntius tetrazona*). *BMC Genom.*, **2020**, 21, 737.
61. Logan, C. A and Buckley, B. A. Transcriptomic responses to environmental temperature in eurythermal and stenothermal fishes. *J. Exp. Biol.*, **2015**, 12(218), 1915–1924.
62. López-Flores, I. and Garrido-Ramos, M. A. The repetitive DNA content of eukaryotic genomes. *Genome dyn.*, **2012**, 7, 1–28.
63. Louzada, S.; Lopes, M.; Ferreira, D. et al. Decoding the Role of Satellite DNA in Genome Architecture and Plasticity—An Evolutionary and Clinical Affair. *Genes,* **2020**, 11, 72.
64. Macqueen, D. J. and Johnston, I. A. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. R. Soc. B*, **2014**, 281, 2013–2881.
65. Makarevitch, I.; Waters, A.J.; West, P.T. et al. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.*, **2015**, 11, e1004915.
66. Marburger, S.; Alexandrou, M.A.; Taggart, J.B. et al. Whole genome duplication and transposable element proliferation drive genome expansion in Corydoradinae catfishes. *Proc. R. Soc. B Biol. Sci.*, **2018**, 285, 20172732.
67. Metcalfe, C. J.; File´e, J.; Germon, I. et al. Evolution of the Australian lungfsh (*Neoceratodus forsteri*) genome: A major role for CR1 and L2 LINE elements. *Mol. Biol. Evol.*, **2012**, 29, 3529–3539.

68. Metcalfe, C. J. and Casane, D. Accommodating the load: Te transposable element content of very large genomes. *Mob. Genet. Elements*, **2013**, 3, e24775.
69. Meyer, A. and Van de Peer, Y. From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). *Bioessays*, **2005**, 27, 937–945.
70. Meyer, A.; Schloissnig, S.; Franchini, P. et al. Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature,* **2021**, 590, 284–289.
71. Newton, J. R.; Zenger, K. R. and Jerry, D. R. Next-generation transcriptome profiling reveals insights into genetic factors contributing to growth differences and temperature adaptation in Australian populations of barramundi (*Lates calcarifer*). *Mar. Genom.*, **2013**, 11, 45–52.
72. Ng, P. K. L. and Tan, H. H. Freshwater fishes of Southeast Asia: potential for the aquarium fish trade and conservation issues. *Aquar. Sci. Conserv.*, **1997**, 1, 79–90.
73. Novák, P.; Neumann, P.; Pech, J. et al. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, **2013**, 29, 792–793.
74. Olmo E. Nucleotype and cell size in vertebrates: A review. *Basic Appl. Histochem.*, **1983**, 27, 227–256.
75. Organ, C. L.; Struble, M.; Canoville, A. et al. Macroevolution of genome size in sarcopterygians during the water-land transition. *C.R. Palevol.*, **2016**, 15, 65–73.
76. Pallavicini, A.; Canapa, A.; Barucca, M. et al. Analysis of the transcriptome of the Indonesian coelacanth *Latimeria menadoensis*. *BMC Gen.*, **2013**, 14, 538.
77. Palzenberger, M. and Pohla, H. Gill surface area of water-breathing freshwater fish. *Rev. Fish. Biol. Fish.*, **1992**, 2, 187–216.
78. Pappalardo, A. M.; Ferrito, V.; Biscotti, M. A. et al. Transposable elements and stress in vertebrates: An overview. *Int. J. Mol. Sci.*, **2021**, 22, 1970.
79. Plohl, M.; Meštrović, N. and Mravinac, B. Centromere identity from the DNA point of view. *Chromosoma*, **2014**, 123, 313–325.

80. Postlethwait, J. H.; Woods, I. G.; Ngo-Hazelett, P. et al. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res.*, **2000**, 10, 1890–1902.
81. Pray, L. and Zhaurova, K. Barbara McClintock and the discovery of jumping genes (transposons). *Nature Education*, **2008**, 1(1):169.
82. Price, A. L.; Jones, N. C. and Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics*, **2005**, (Suppl. 1), 351–358.
83. Rebollo, R.; Horard, B.; Hubert, B. et al. Jumping genesand epigenetics: Towards new species. *Gene*, **2010**, 454, 1–7.
84. Russo, R.; Curtis E. W. and Yanong, R. P. E. Preliminary investigations of hydrogen peroxide treatment of selected ornamental fishes and efficacy against external Bacteria and parasites in Green swordtails. *J. Aquat. Anim. Health.*, **2007**, 19(2), 121–127.
85. Schaeffer, C. E.; Figueroa, N. D.; Liu, X. et al. phRAIDER: pattern-hunter based rapid ab initio detection of elementary repeats. *Bioinformatics*, **2016**, i209–i215.
86. Schmidt, A. L. and Anderson, L. M. Repetitive DNA elements as mediators of genomic change in response to environmental cues. *Biol. Rev. Camb. Philos. Soc.*, **2006**, 81, 531–543.
87. Sirijovsky, N.; Woolnough, C.; Rock, J. et al. NfCR1, the frst non-LTR retrotransposon characterized in the Australian lungfsh genome, *Neoceratodus forsteri* shows similarities to CR1-like elements. *J. Exp. Zool.*, **2005**, 304, 40–49.
88. Smit, A. and Hubley, R. RepeatModeler 1.0.11. **2018,** RepeatModeler http://www.repeatmasker.org/ RepeatModeler/.
89. Steendam, C.; Verhelst, P.; Van Wassenbergh, S. et al. Burrowing behaviour of the European eel (*Anguilla anguilla*): Effects of life stage. *J. Fish Biol.*, **2020**, 97, 1332–1342.
90. Stoll, G. A.; Oda, S. I.; Chong, Z. S. et al. Structure of KAP1 tripartite motif identifies molecular interfaces required for retroelement silencing. *Proc. Natl. Acad. Sci. USA.*, **2019**, 116(30), 15042–15051.
91. Stoll, G. A.; Pandiloski, N.; Douse, C. H. et al. Structure and functional mapping of the KRAB-KAP1 repressor complex. *The EMBO journal*, **2022**, 41(24), e111179.

92. Sun, C.; Shepard, D. B.; Chong, R. A. et al. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol. Evol.*, **2012a**, 4(2), 168–183.
93. Sun, C.; Arriaza, J. R. L. and Mueller, R. L. Slow DNA loss in the gigantic genomes of salamanders. *Genome Biol. Evol.*, **2012b**, 4, 1340–1348.
94. Sun, C. and Mueller, R. L. Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. *Genome Biol. Evol.,* **2014,** 6, 1818–1829.
95. Takeda, S.; Sugimoto, K.; Otsuki, H. et al. A 13-pb cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *Plant Journal*., **1999**, 18, 383–393.
96. Tarallo, A.; Angelini, C.; Sanges, R. et al. On the genome base composition of teleosts: the effect of environment and lifestyle. *BMC Genom.,* **2016**, 17, 173.
97. Vandepoele, K.; de Vos, W.; Taylor, J. S. et al. 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*, **2004**, 101, 1638–1643.
98. Vinogradov, A. E. Genome in toto. *Genome*, **1999**, 42, 361–362.
99. Wang, J. T.; Li, J. T.; Zhang, X. F. et al. Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (*Cyprinus carpio*). *BMC Genom.*, **2012**, 13, 96.
100. Wang, K.; Wang, J.; Zhu, C. et al. African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell*, **2021**, 184(5), 1362–1376.e18.
101. Wells, J. N.; Chang, N. C.; John McCormick, J. et al. Transposable elements drive the evolution of metazoan zinc finger genes. *Biorxiv*, **2022**.
102. Yao, B.; Li, S. and Chan, E. K. Function of GW182 and GW bodies in siRNA and miRNA pathways. *Adv. Exp. Med. Biol.*, **2013**, 768, 71–96.

103. Zhang, J.; Yu, C.; Krishnaswamy, L. et al. Transposableelements as catalysts for chromosome rearrangements. *Methodsin mol. biol.*, **2011**, 701, 315–326.