

## Article

# Quantitative Approach to Fish Cytogenetics in the Context of Vertebrate Genome Evolution

Veronika Borůvková<sup>1</sup>, W. Mike Howell<sup>2</sup>, Dominik Matoulek<sup>1</sup> and Radka Symonová<sup>3,\*</sup> 

<sup>1</sup> Faculty of Science, University of Hradec Kralove, 500 03 Hradec Kralove, Czech Republic; veronika.boruvkova@uhk.cz (V.B.); dominik.matoulek@uhk.cz (D.M.)

<sup>2</sup> Department of Biological and Environmental Sciences, Samford University, Birmingham, AL 35226, USA; wmlhowell@samford.edu

<sup>3</sup> Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technische Universität München, 85354 Freising, Germany

\* Correspondence: radka.symonova@gmail.com

**Abstract:** Our novel Python-based tool EVANGELIST allows the visualization of GC and repeats percentages along chromosomes in sequenced genomes and has enabled us to perform quantitative large-scale analyses on the chromosome level in fish and other vertebrates. This is a different approach from the prevailing analyses, i.e., analyses of GC% in the coding sequences that make up not more than 2% in human. We identified GC content (GC%) elevations in microchromosomes in ancient fish lineages similar to avian microchromosomes and a large variability in the relationship between the chromosome size and their GC% across fish lineages. This raises the question as to what extent does the chromosome size drive GC% as posited by the currently accepted explanation based on the recombination rate. We ascribe the differences found across fishes to varying GC% of repetitive sequences. Generally, our results suggest that the GC% of repeats and proportion of repeats are independent of the chromosome size. This leaves an open space for another mechanism driving the GC evolution in vertebrates.

**Keywords:** GC content; chromosome size; linkage group; microchromosomes; GC-biased gene conversion



**Citation:** Borůvková, V.; Howell, W.M.; Matoulek, D.; Symonová, R. Quantitative Approach to Fish Cytogenetics in the Context of Vertebrate Genome Evolution. *Genes* **2021**, *12*, 312. <https://doi.org/10.3390/genes12020312>

Academic Editors: Samuel Martin and Rossi Anna Rita

Received: 23 November 2020

Accepted: 17 February 2021

Published: 22 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The quantitative approach in fish cytogenetics, or more precisely in cytogenomics, has so far been rather neglected and the same can be said for vertebrates in general. This is due to the lack of available data. The traditional cytogenetics prevailing during the last decades has mostly used a qualitative approach, which was indispensable for the exploration of chromosomal traits generally. With the still accelerating progress of genome sequencing, a large body of evidence on base composition, i.e., the proportion of guanines and cytosines in DNA (GC%), became available also in non-model vertebrates. In parallel, the Animal Genome Size Database [1] is another valuable resource of information useful particularly to estimate the completeness of genome assemblies. Only this recent development enabled us to tackle compositional cytogenomics of vertebrates from a quantitative viewpoint.

There were attempts to quantify results of the mostly qualitative (molecular) cytogenetics during all the phases of its development. With the first human karyotypes being presented during the early 1960s, it was not surprising to see an explosion of karyotypes being rapidly published in other living groups of organisms, especially vertebrates. However, this was before chromosome banding techniques (G-, Q-, C-, R-, AgNOR, etc.), karyotypes were simply grouped according to their size and the position of the centromere. Quantitatively, one could still determine a number of chromosomal features: 1. The chromosome number of a species; 2. the position of the centromere; 3. often the position and number of the secondary constriction(s) = nucleolus organizer regions; 4. a measurement of the

chromosome long arm length; 5. measurement of the short arm length; 6. measurement of the total chromosome length (TCL); 7. the long arm/short arm ratio (this was important as it pinpointed the position of the centromere along the chromosome) [2]; 8. the total haploid chromosome length (THCL); 9. the fundamental number (FN) of a karyotype (determined by counting the chromosome arms). Therefore, the  $FN \leq 2 \times 2n$ , with the difference depending on the number of mono-armed chromosomes present (telocentric or acrocentric); 10. measurement of the centromeric index (CI), which is the length of the short arm divided by the total length of the chromosome  $\times 100$ . The CI proved to be a very valuable quantitative measure for describing the shape of a chromosome, especially before chromosome banding methods [2,3].

Since the chromosomes of the vast majority of fish species could not be banded, especially G-banded, most of the authors of early fish karyotype papers included from one to several of the above counts and measurements in efforts to quantify their findings. The difficulty in obtaining chromosome banding patterns in fishes is likely due to their small size relative to that of higher vertebrates, and that, unlike birds and mammals, whose genomes can be divided into multiple GC-rich and AT-rich segments, fishes have been suggested to have little compartmentalization of their genomes by base composition [4–7], etc. Other quantitative traits were developed and have often been investigated in fish separately: Chromosome numbers, particularly in a phylogenetic context [8]; genome size expressed as the C-value based on the Animal Genome Size Database [1], as related to the population size [9], and to the nucleotypic effect [10,11].

The base composition at the level of fish chromosomes has so far been addressed in two model species in relation to isochores [12]. In two more fish species, the base composition was analysed in the entire genome [13,14]. However, these quantitative cytogenomic traits, particularly at the chromosome level, are crucial inputs in the analyses of mechanisms driving the regional GC% and the AT/GC heterogeneity in mammals and birds vs. the AT/GC homogeneity in lower vertebrates. Namely, one of the currently most accepted concepts trying to elucidate the aforementioned questions is the GC-biased gene conversion, gBGC [15–17]. This concept explains the increased GC% and its heterogeneous organisation in mammals and birds by the recombination rate (RR) per megabase pair (Mb; recently reviewed by [18]). Here, the chromosome or chromosome arm size plays an important role as there is at least one crossing-over (i.e., one recombination event) per one chromosome arm in bi-armed chromosomes and per chromosome in mono-armed chromosomes [18,19]. Among vertebrates, the efficiency of gBGC on the GC% has been investigated in great detail mostly in the coding (i.e., exonic) regions and above all in mammals [20–22] and birds [23,24], but also in reptiles [25,26]. The same applies for the online available GCevobase, an evolution-based database for the GC content in eukaryotic genomes displaying GC contents for all the annotated coding sequences from Ensembl [27]. However, the coding sequences make up not more than 2% of the human genome and similarly small fractions in other vertebrates (Matoulek et Symonová, unpublished data). Therefore, these highly detailed and sophisticated analyses referenced above could not catch the compositional evolution along large genome fractions as, e.g., the far more abundant repetitive fraction. Another approach represented by Frenkel et al., employed a far higher resolution beyond the chromosome, thus omitting their importance. They also investigated a limited number of species available at that time [28]. A more relevant study by Li and Du already focused on the chromosome level, however, dividing animals into mammals and non-mammals blurred the lineage specific traits together with the low number of analysed species available at that time [29]. Later, fully irrespective of GC% and its organisation, fishes were demonstrated to have the highest recombination rate among vertebrates [19] even without including the information on chromosome arms (FN) numbers, which would further increase the acquired values. This shows an urgent need to revisit this issue from another, so far unexplored viewpoint.

GC% is linked to the genome size [30] and hence, chromosome counts are related to the genome size and chromosome size as well as their morphology (bi-armed vs. mono-

armed; [19]). The genome size shows a clear positive association with the proportion of repetitive elements, particularly of transposable elements (TEs [31]; etc.).

In terms of genome composition, we need to distinguish the overall genomic GC-richness or GC content, i.e., the percentage of G + C (GC%) from the avian or mammalian situation recognized as the AT/GC heterogeneity [32], known also in non-teleost gars [33] and further confirmed by [34] in this special issue. The latter situation is characterized by an alternation of GC-rich and GC-poor regions along chromosomes, thus forming banding patterns upon an AT- and GC-specific staining (for more details, see [34] in this special issue). This AT/GC heterogeneity had been ascribed to a higher diversity of the isochores and their families distinguished according to their GC% [35]. In the AT/GC heterogeneous (mammalian, avian, and gar) genomes, the overall GC% can, however, be even lower than it is in the AT/GC homogeneous (fish) genomes. Considering that all of the currently available vertebrate genome assemblies contain gaps due to either repeat-rich or GC-rich regions [36], fish with GC-rich genomes might actually be even more GC-rich than estimated, and potentially even more GC-rich than mammalian and avian genomes. However, this bias should be comparable for all the genomes along the vertebrate phylogenetic tree. The only difference can be expected in newer improved versions of genome assemblies employing the hybrid approach of sequencing as, e.g., [37] recently in bats.

Our goal in this study is to utilize the increasingly available vertebrate genomes assembled to the chromosome level to assess the role of chromosome size in the overall GC content and in the AT/GC genomic heterogeneity across vertebrates. To do so, we employed our novel Python-based tool EVANGELIST published in this special issue [34]. This tool uses the sliding window approach to visualize and quantify the percentage of repeats (rep%) and GC percentage (GC%) in both repeats and non-repetitive DNA simultaneously along the chromosomes [34]. Employing this tool revealed several interesting quantitative traits at the chromosomal level across the vertebrates. These inputs together with a large-scale meta-analysis of genomic data serve as a test of the concept of GC-biased gene conversion in fishes and in other vertebrates.

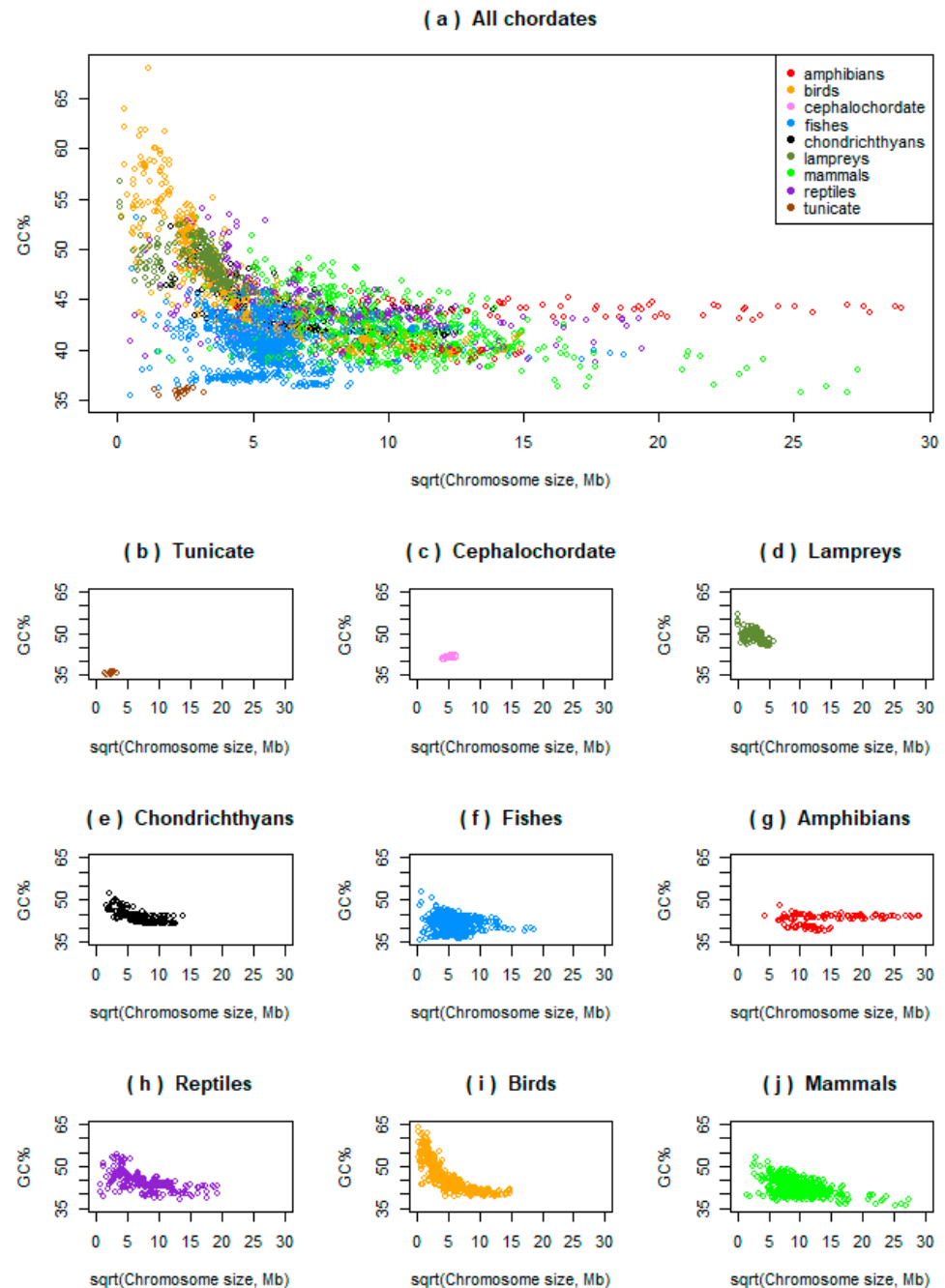
## 2. Materials and Methods

We meta-analysed NCBI/genome records (accessed in January 2021) on the genome size, chromosome size, and GC percentage (GC%) of entire genomes and their chromosomes (in the genome assemblies called linkage groups, LGs).

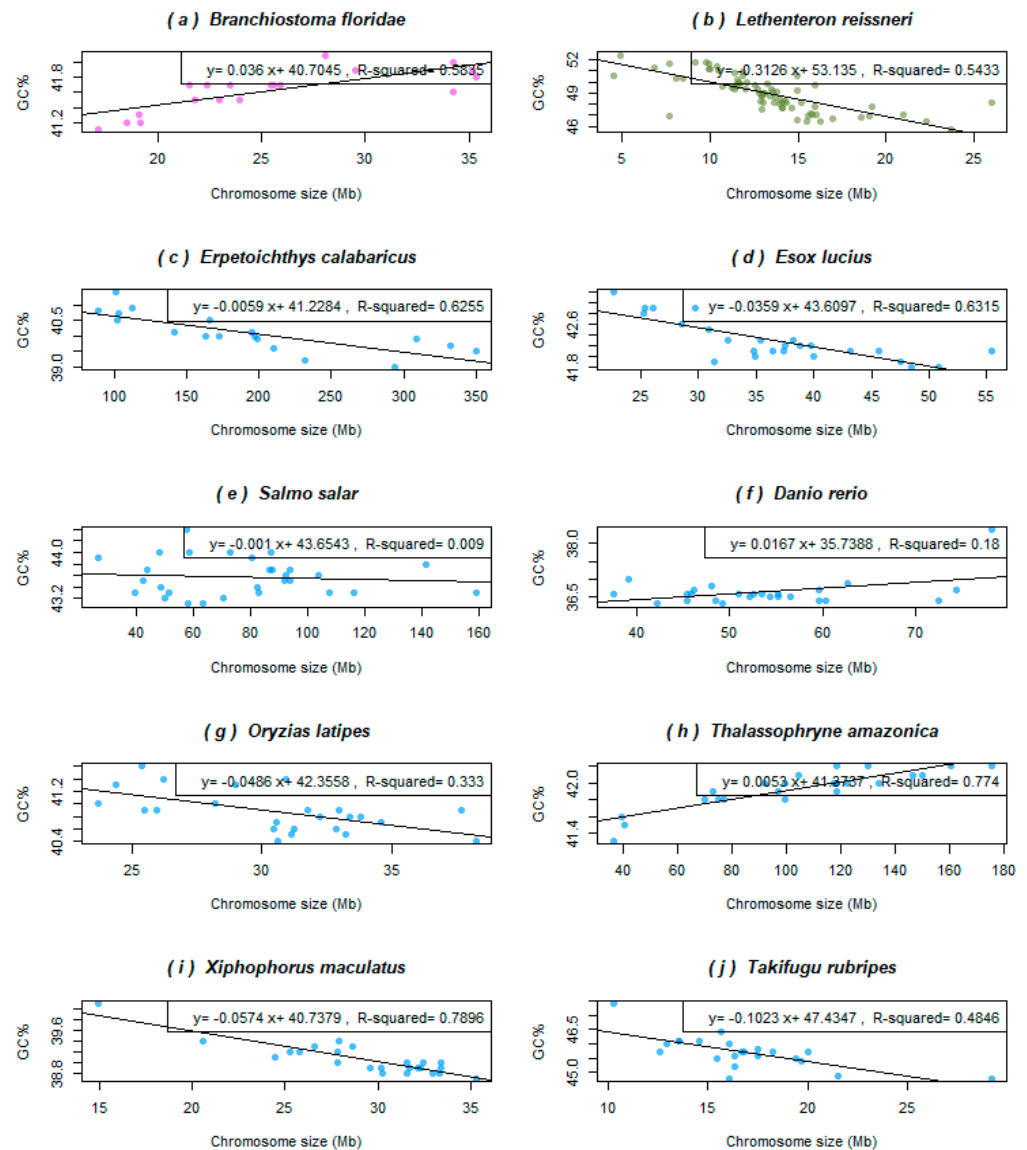
### 2.1. Data Acquisition, Filtering, and Manual Curation

At the NCBI webpage (NCBI, 2016) using the available genome selection filters, we selected chordates assembled to the chromosome level. We manually checked the obtained datasets for multiple records (particularly human, dog, primates, and other model species), incomplete assemblies and hybrids. The online filtering yielded 157 fishes, 10 amphibians, 13 reptiles, 66 birds, and 327 mammals. However, upon the manual inspections, the numbers of species usable in our study decreased to 85 mammals, 55 birds, 12 reptiles, nine amphibians, and 118 fishes. Among fishes, there were four chondrichthyans, three lampreys, one lungfish with an incomplete dataset, one sturgeon, one gar, one bichir, and 107 teleosts (lists of species are in the Supplementary). These data were used for Figures 1–4 and Figures S2 and S3. These figures were prepared using R [38] with the library scales included in the basic R and further with packages ggplot2, forcats, and ggpmisc, all belonging to tidyverse [39]. There was a bias in mammalian lineages represented among the species assembled to the chromosome level. Namely, some lineages were not represented at all (Afrotheria, Cingulata, Scandentia, Dermoptera, Eulipotyphla, and Pholidota), some were underrepresented (Chiroptera, the second largest mammalian order with only three species available), and some were overrepresented (primates, rodents, and artiodactyls). More details are available in the Supplementary Methods. Therefore, we have selected a subset of 26 mammals trying to compensate for this bias and to avoid the overrepresented

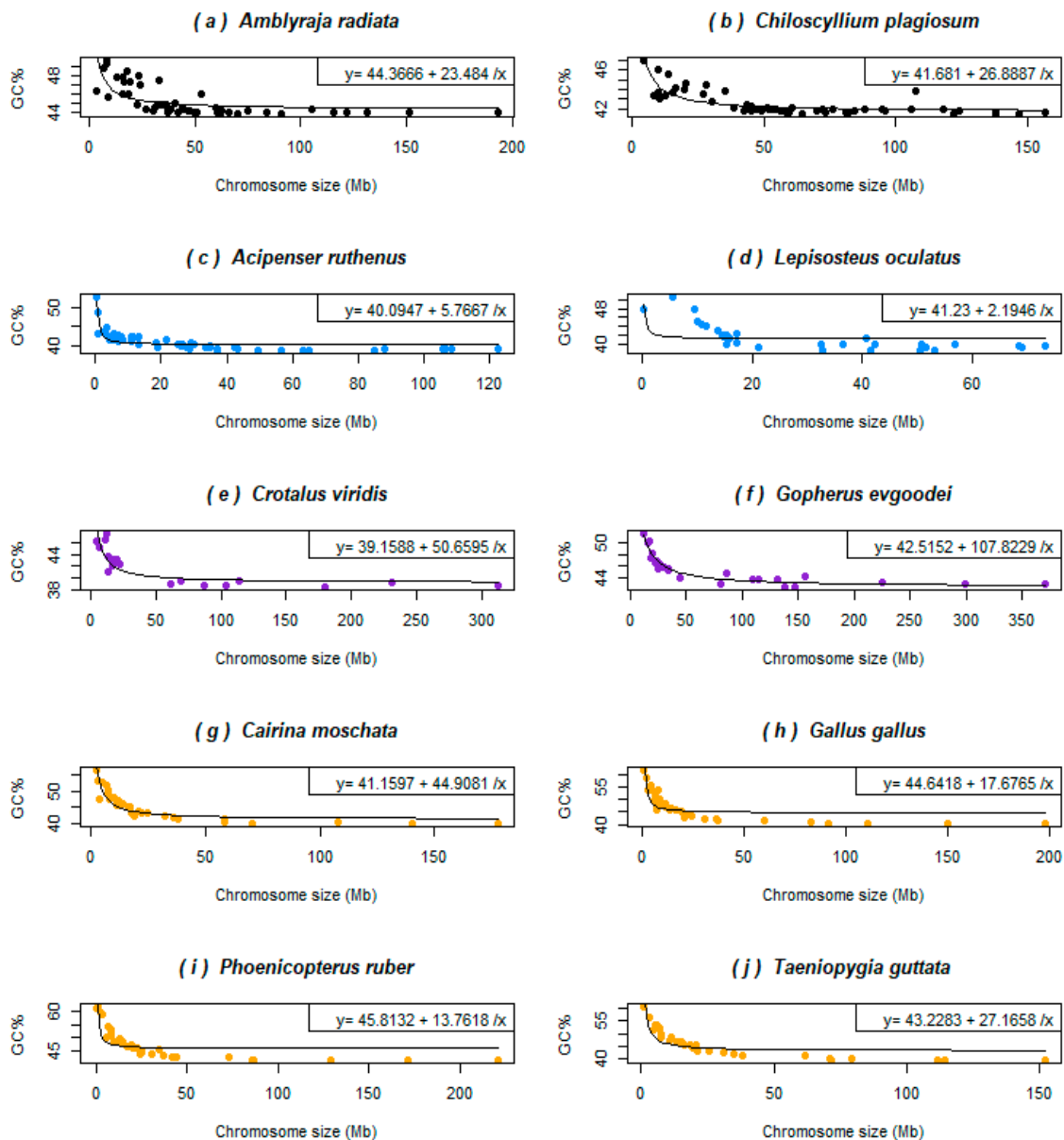
species skewing the results. Similarly, in birds, the most abundant order available were Passeriformes and to a less extent also Anseriformes. Hence, we again selected a subset of 18 avian species to avoid any bias towards these orders. However, some avian species, despite being filtered, turned out not to have the required chromosome level assembly available. Finally, we performed a comparable species selection in fish and produced a subset of 49 species representing the orders, as much as possible, and not overrepresenting the better explored ones.



**Figure 1.** Large-scale overview of the relationships between chromosome size ( $x$ -axis, square root transformed for a better visualization) and chromosome GC% ( $y$ -axis) across selected chordates on equidistant axes  $x$  and  $y$ . (a) All chordates together; (b–j) major chordate lineages separately ordered phylogenetically. The graph (f), fishes, includes two basal ray-finned species (sturgeon and gar) and teleosts. In (g), the largest genome of axolotl is not visualized, see Figure 4 for amphibians. Each point represents a single chromosome of a species within the given group.

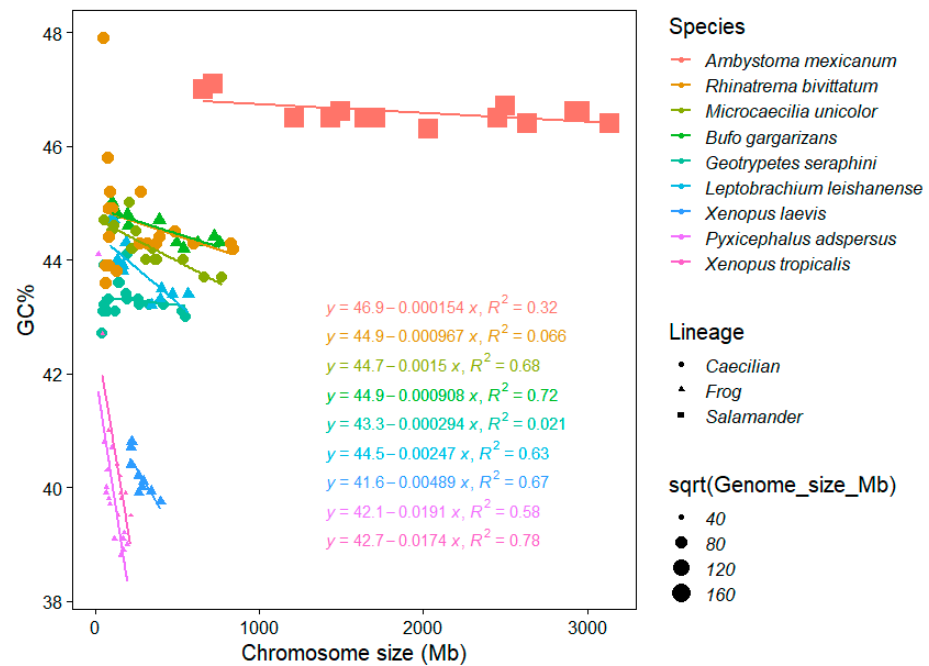


**Figure 2.** Relationships between GC% and the chromosome size in Mb in selected fish and fish-like species representing different patterns of this association. (a) Lancelet as the basal chordate considered the ancestral condition shows a significantly positive association between the chromosome size and GC%; (b) lamprey, the extremely GC-rich jawless fish with numerous mostly tiny chromosomes, and reedfish (c) with its giant non-teleost genome and extremely large chromosomes both show a significantly negative association; (d) the Northern Pike, an important pre-duplication model species for salmonids, shows a significantly negative association, as well; (e) in the Atlantic salmon, there is no association between these two measures; (f) zebrafish, although the mostly investigated fish model with a larger teleost genome, shows a rather weak positive association; (g) medaka, another important model species, shows a weaker negative association; (h) toadfish with a surprisingly large teleost genome shows a significantly positive association and the largest teleost chromosome, although no whole genome duplication is known here as, e.g., in salmonids with comparable genome and chromosome size but higher chromosome numbers; (i) the Southern platyfish shows the most significant negative association and the typical teleost chromosome sizes; (j) fugu represents the compact teleost and vertebrate genomes shows a slightly less prominent but still significantly negative association. The colour code corresponds to Figure 1.



**Figure 3.** Relationships between GC% and the chromosome size (Mb) in basal fish lineages that all possess microchromosomes show an inverse relationship as do birds and reptiles with microchromosomes. (a) The thorny skate, representing the superorder Batoidea, is the GC-richest species in this group; (b) the whitespotted bamboo shark represents the superorder Selachimorpha; (c) sterlet, one of the only two sequenced sturgeons, with the GC-poorest genome; (d) the spotted gar belongs to the infraclass Holostei and is known for its mammalian-like AT/GC compartmentalization; (e) the rattlesnake represents a smaller and GC-poorer reptilian genome; (f) gopher tortoise represents a larger and more GC-rich reptilian genome; (g) the Muscovy duck is the first avian genome compared here; (h) chicken represents one of the best assembled avian genomes; (i) the American flamingo represents large wading birds; finally, (j) the zebra finch with its fifteen genome assemblies available represents songbirds. Violin plots of these species are available in the Supplementary Figure S3. The colour code corresponds to Figure 1.



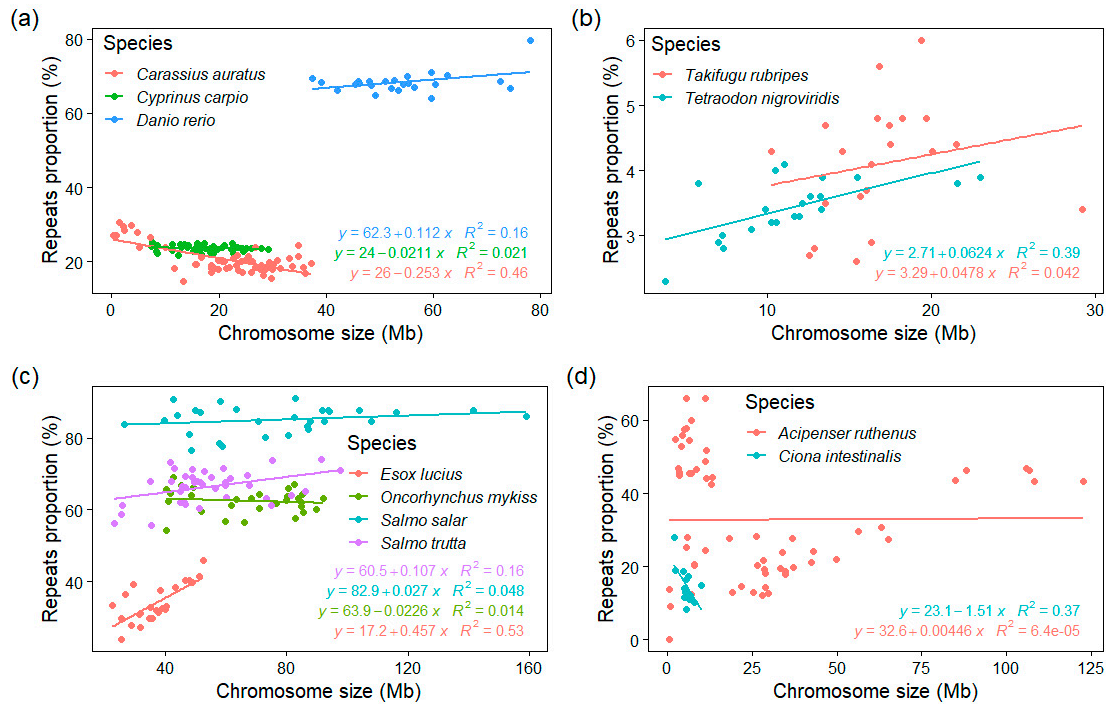


**Figure 4.** The chromosome size and GC% of chromosomes in currently available amphibians. Nine species ordered according to the size of their entire genomes in the legend and equations, which is also reflected in the point size. Point shapes reflect the main amphibian lineages. Patterns of the associations between the chromosome size and GC% forms three distinct groups partly corresponding to the main lineages and fully corresponding to genome sizes.

## 2.2. Repeats Analyses and Genome Size Data

To employ the Python-based tool EVANGELIST [34], we have used repeat-masked data from Ensembl to analyse a potential relationship between the GC% and repeats proportion (rep%). This approach was utilized also to address the issue of a potential effect of the decreasing chromosome size (particularly in microchromosomes) on the GC% and rep%. These data were used for Figure 5 and Figures S4–S7.

Genome size data obtained from the Animal Genome Size Database [1] was utilized during the manual curation of the completeness of genomic data from NCBI and to the comparison of genome size and GC% between mammals and fish (Figure S1).



**Figure 5.** Relationships between the repeats proportion and chromosome size in selected fish species with relevant traits—genome size, chromosome numbers and sizes, and whole-genome duplication or genome compaction underwent. (a) In three cyprinids with highly different cytogenomic traits; (b) two of the smallest teleost (and vertebrate) genomes show different associations; (c) the Northern pike as the pre-duplication ancestor of salmonids shows a clear positive association, while three salmonids show no association with a substantial repeats enrichment; (d) the tunicate sea squirt has a chromosome size comparable with microchromosomes of sturgeon, however, differs greatly in proportions of repeats showing a negative association.

### 3. Results

Initially, we analysed the potential relationships between the chromosome size and its GC% across ray-finned fishes and compared them to one tunicate, one cephalochordate, all the three available lampreys, all the three available chondrichthyans, and to other vertebrates (amphibians, reptiles, birds, and mammals; Figure 1a, each point represents a single chromosome; graphs for each group are shown separately in Figure 1b–j). This shows that most fishes occupy the area with the lowest GC% and intermediate chromosome sizes (blue points). The single tunicate represented, the sea squirt (*Ciona intestinalis*, brown points in Figure 1b), occupies the actual lowest values of GC% and chromosome sizes and shows no association between these measures. The single cephalochordate represented, the Florida lancelet (*Branchiostoma floridae*, pink points in Figure 1c) is embedded approximately in the middle of the “fish area” surrounded by fish and shows a significant positive association between the GC% and chromosome size (better details in Figure 2a). In lampreys, the GC% and chromosome sizes are significantly positively associated (black points, Figure 1d, Figure 2b in better details). Chondrichthyans (black points, Figure 1e) but also sturgeon and gar, show a negative association between the GC% and chromosomes forming a curve resembling a hyperbole similarly to avian chromosomes (more details in Figure 3). In fishes (Figure 1a,f, blue), the linear approximation is mostly a suitable model representing their relationship between the chromosome size and GC%, however, all negative, positive, and no associations occur (more details in Figure 2 and below). Only eight of the nine currently available amphibian species are visualized here (Figure 1a,g, red) due to the extremely huge chromosome size in axolotl (*Ambystoma mexicanum*). All the amphibians are visualized in Figure 4 in proper details showing different levels of the negative association between the GC% and chromosome size not apparent at the resolution in Figure 1. Reptiles (Figure 1a,h, violet) with macro- and microchromosomes show a similar association of



GC% with the chromosome size to chondrichthyans, basal fish lineages (sturgeon and gar), and birds. This means a negative relationship of GC% vs. chromosome size, behaving as if there was no association in large(r) (macro)chromosomes and a steep descent in GC% in microchromosomes. Birds (Figure 1a,i, orange) show a clear hyperbolic curve. Mammals (Figure 1a,j, green) show a negative relationship between GC% and the chromosome size with a varying strength similar to fishes. Violin plots of compositional differences between the selected fishes and mammals are in Figure S2.

### 3.1. Variability in Relationships between the Chromosome Size and GC% in Ray-Finned Fishes

Figure 2 shows representative species with diverse trends in their relationship between the chromosome size and GC%. The cephalochordate outgroup of all vertebrates, the lancelet (*B. floridae*), shows a positive correlation between GC% and the chromosome size that fits the linear model well (Figure 2a). Its haploid genome assembly size 513.461 Mb corresponds to the smallest teleosts and the size of its nineteen haploid chromosomes ranging between 17.12 and 35.34 Mb. The jawless outgroup of other vertebrates, the lamprey (*Lethenteron reissneri*) shows a negative correlation between GC% and the chromosomes size fitting the linear model well (Figure 2b). Its haploid genome assembly size is 1063.01 Mb and corresponds to the average teleosts, however, the size of its 72 haploid chromosomes ranges between 4.5 and 26 Mb. The remaining fish species depicted show diverse associations of diverse strength or no association at all. The non-teleost reedfish (*Erpetoichthys calabaricus*, Figure 2c) has a huge (haploid) genome of 3209 Mb separated into only eighteen chromosomes ranging in their size between 88.37 and 350.1 Mb (the largest chromosome is almost as large as the entire genome of pufferfishes, which have a reduced genome size). Its regression line is clearly decreasing, however, due to the extremely different scales of the X and Y axes, the line slope is close to zero. The relationship fits the linear model well (Figure 2c). Following are two teleosts belonging to the superorder Protacanthopterygii [40] but showing different trends and evolutionary history. Namely, in the Northern pike (*Esox Lucius*, Figure 2d), GC% decreases almost linearly with the increasing chromosome size, thus fitting the linear model well. This species has a typical teleost haploid genome size and chromosome number (940 Mb and 25 chromosomes ranging between 22.6 and 55.4 Mb). Whereas in the Atlantic salmon (*Salmo salar*, Figure 2e), that underwent the salmonid specific whole genome duplication points in the graph, form a cloud that does not fit the linear model and with no obvious simple relationship between the variables. This species has the haploid genome size 2966.89 Mb in 29 chromosomes of size between 40 and 160 Mb. The same situation exists in two other salmonids in NCBI assembled to the chromosome level (*Salmo*, *Oncorhynchus*, but not in *Thymallus* and *Salvelinus* that both show a slightly negative association between GC% and the chromosome size, not shown). The next species analysed is the model organism zebrafish (*Danio rerio*, Figure 2f) of the basal teleost lineage Ostariophysi [40], known to have a rather larger [1,41] and AT-enriched genome, e.g., [33]. Its haploid genome assembly of about 1408 Mb is divided into 25 chromosomes ( $2n = 50$ ) ranging in size between 37.5 and 59.58 Mb. Excluding the extreme, rightmost point of the graph the GC% oscillates around the mean independently of the chromosome size and the relationship does not fit a linear model. Other cypriniform fishes analysed (*Cyprinus* and *Carassius*, not shown) show similar trends, i.e., no association between the chromosome size and GC%. Similarly, in the Japanese rice fish (*Oryzias latipes*, Figure 2g), Beloniformes, the lower GC% appears to occur in larger chromosomes, but the relationship cannot be described in simple terms and does not fit the linear model well. So far, a unique situation among teleosts occurs in toadfish (*Thalassophryne amazonica*, Figure 2h) of the order Batrachoidiformes [40]. This species shows a significantly positive association between GC% and the chromosome size, similarly to the cephalochordate and a large haploid genome for a teleost, 2446.59 Mb with 23 large chromosomes ranging in size between 36.25 and 175.46 Mb. These traits together with its GC-richness (42.02%) resemble salmonids. Furthermore, the Southern platyfish (*Xiphophorus maculatus*, Figure 2i) of the order Cyprinodontiformes shows a tight linear

relationship between GC% and the chromosome size ( $R^2 = 0.7896$ ) with points distributed visibly along the decreasing line. Finally, we selected pufferfish (*Takifugu rubripes*, Figure 2j) representing the order Tetraodontiformes with the smallest vertebrate genomes. It might appear that the shape of the line is influenced mainly by the two outer points. However, the decreasing tendency holds (although not that strong) without them and their negative association fits the linear model well.

### 3.2. Basal Fish Lineages Show a Similar Relationship between the Chromosome Size and GC% to Birds and Some Reptiles

In the two non-teleost ray-finned fishes and three chondrichthyans with available genomes assembled to the chromosome level (four shown here, Figure 3a–d), the GC% is inversely related to their chromosome size as in some reptiles (a snake and a turtle shown here, Figure 3e–f) and in birds (four shown here, Figure 3g–j). This relationship results in a hyperbolic curve in all the species analysed here. The curve is “smoother” for birds than in these basal fish lineages and the selected reptiles, where points representing single chromosomes are more scattered. In chondrichthyans, reptiles, and birds the chromosome size around 50 Mb represents a breakpoint behind the fact that the GC% becomes independent of the chromosome size. Sterlet and gar have this breakpoint around 20 Mb.

### 3.3. Genome Size Drives GC% in Amphibians but Inversely than Expected

Amphibians greatly differ in their genome size and salamanders are known to reach the second highest vertebrate genome sizes (116 and 118 Gb [42]). This fact highly influences the availability and also quality of amphibian genome assemblies in the NCBI database [43], whereas cytological data (C-values in pg) are very rich and provide robust evidence on the genome size in anurans and urodels [1]. The mean C-value of anurans and urodels is 3.59 pg and 35.21, respectively [1]. Our results show interesting consequences of this dynamic genome size evolution in amphibians for their GC%. Here, even such a low sample size (nine species) enabled to distinguish three distinct groups, partly phylogenetically independent, according to their GC%, chromosome size, and associations thereof (Figure 4):

1. A single salamander species with its giant and extremely GC-rich chromosomes (and genome) showing a weak negative association between GC% and the chromosome size ( $R^2 = 0.32$ ; genome size 32,396.4 Mb, GC ~ 46.5%).
2. Three caecilians and two frogs (pelobatid and bufonid) with an intermediate GC% and chromosome sizes (and intermediate genome sizes between 3779.43 and 5319.24 Mb, GC ~ 43%–44%) showing no association in two caecilians ( $R^2 = 0.066$  in *Rhinatrema*,  $R^2 = 0.021$  in *Geotrypetes*) to a significant negative association in the remaining caecilian *Microcaecilia* and two anurans ( $R^2 = 0.63$ – $0.72$ ) between GC% and the chromosome size.
3. Three remaining frogs (two pipids and a pyxicephalid) with a significantly ( $R^2 = 0.5837$ ,  $0.6689$ , and  $0.7824$ ) negative association between GC% and the chromosome size (the smallest genomes between 1451.3 and 2718.43 Mb, GC ~ 39%–40.5%).

This indicates that GC% increases together with the genome size and the genome size appears to be the determinant of this peculiar situation.

### 3.4. GC% vs. Repeats Proportion (%) and Chromosome Size

To explore any potential relationship between the chromosome size and their corresponding GC%, rep%, and GC% of repeats, we calculated these measures for each chromosome separately in 43 fish and fish-like species and in eleven mammals. These data were extracted from our tool EVANGELIST [34] and represent individual sliding windows along chromosomes and their GC% and rep% values. The results are pictured as scatter plots for each chromosome organized as tiles in Supplementary online data S4–S7 and on the GitHub online repository of our tool EVANGELIST ([https://github.com/bioinfohk/evangelist\\_plots/tree/master/rep%25\\_vs\\_GC%25](https://github.com/bioinfohk/evangelist_plots/tree/master/rep%25_vs_GC%25) (accessed on 31 January 2021)). These plots show several patterns in the relationships between GC%

and repeats%, but no obvious association of GC%, rep%, and GC% of repeats with the chromosome size. The distribution of repeats percentage is asymmetric for each chromosome, while typically unchanging throughout the species and independent of the chromosome size. This can be seen from the position of the median of rep% (vertical line) in each tile. The same holds for the position of regression lines (similar in shape and position within the species). We also analysed the relationship between the repeats proportion at the level of entire chromosomes extracting the median rep% for each chromosome and its size (Figure 5). This is first shown in three cyprinids (goldfish, carp, and zebrafish, Figure 5a) differing in chromosome counts but similar genome sizes, where a significant ( $R^2 = 0.46$ ) negative association exists in goldfish, a very slight positive association in zebrafish, and no association at all in carp. Two compact tetraodontiform genomes show a moderate positive association in tetraodon and no association in fugu (Figure 5b). Another complex situation is shown in the Northern pike and three salmonids, where whole-genome duplication might have played an important role (Figure 5c). Interestingly, there is no clear association between the chromosome size and rep% even in microchromosomes (Figure 5d), where GC% is otherwise negatively associated with the chromosome size. Here, however, the availability of repeats libraries and hence, the quality of the repeat-masking process needs to be considered particularly in basal fish lineages (sturgeon and gar).

#### 4. Discussion

Chromosome sizes greatly differ among currently sequenced chordates. In parallel to chromosome sizes, genome sizes and proportions of repeat also differ considerably. Despite some mostly superficial similarities, each major lineage of vertebrates occupies its own specific cytogenomic “niche” composed of a unique combination of cytogenomic traits.

##### 4.1. Genome Size and Chromosome Numbers Do Not Entirely Explain the Difference in GC Evolution between Fish and Mammals

The meta-analysis of currently available data on genome size and GC% in vertebrates [1,43] shows that fish genomes are, despite their additional teleost-specific genome duplication [44], about two to three times smaller than mammalian genomes (Figure S1). However, both fish and mammals have converged to very similar diploid chromosome numbers ( $2n$ ) despite this difference in their genome evolution. Namely, the means and the medians of  $2n$  are 48–50 for fish [8] and 46–47 for mammals [1]. With these chromosome counts and fish genome size two to three times smaller than in mammals, this would mean roughly double the recombination rate per megabase (Mb) of DNA in fish in comparison with mammals [19]. This should result in higher GC contents in fish according to the concept of GC-biased gene conversion (gBGC) suggesting that the GC-richness of the mammalian genome is positively associated with higher recombination rates [18]. However, this is not the case—fish genomes are GC-poorer than the mammalian ones (Figure S1). Moreover, fish genomes are AT/GC homogenous in comparison with AT/GC heterogeneity in mammals as we show here in Supplementary, in our previous study in this special issue [34], as well as in [35]. This implies that the basic quantitative traits (genome size and  $2n$ ) are not the only clues to the differences in nucleotide composition between fish and mammals and that another component must be involved. This component might be the repeatome, i.e., all repeated DNA sequences including also the simple tandem repeats, etc. [45] or specifically the mobilome, i.e., the sum of all mobile elements *sensu* [46]. The large-scale view of associations between the chromosome size vs. GC is similar in both mammals and fishes (Figure 1).

##### 4.2. The Role of Chromosome Size Combines with the Influence of Repeats

The situation in chondrichthyans, sturgeon, and gar demonstrated here in comparison with birds indeed indicates an important role of the chromosome size on GC% of the entire genome. This is impacted by the presence of microchromosomes in all these species. The larger genome size in chondrichthyans and sturgeon can explain, in line with the concept of gBGC, their overall lower GC% (Figure 3). Despite this similar karyotype organization

(numerous tiny microchromosomes and clearly larger macrochromosomes), the AT/GC heterogeneity of macrochromosomes has evolved only in gars [33] and not in sturgeons [47] and paddlefishes [48]. The closest living relatives of the AT/GC heterogeneous gars, the bowfin, possess the typical teleost-like karyotype in terms of AT/GC homogeneity and the absence of the microchromosomes [49]. On the other hand, both these ancient fish and birds show the independence of GC% from the chromosome size in their macrochromosomes (Figure 3), although the GC% of the mammalian chromosomes larger than 50–100 Mb is still negatively associated with the respective chromosome size (not shown here). This finding in both basal fish lineages and birds is contradictory to the concept of GC-biased gene conversion (gBGC) [18].

Salmonids and amphibians may provide another clue to understanding the relationship between the genome size and GC%. As shown here at the chromosomal level (Figure 5c), in another paper in this special issue [50] and earlier [30] at the genome level, there is no simple association between these two measures in salmonids. This can be at least partly explained by their extreme repeat content: 58%–60% in the Atlantic salmon [51], ~52% transposable elements in *Coregonus* sp. “Balchen” from the Lake Thun [52], 57.1% of the rainbow trout genome [53], RepeatMasker associated 53.07% of the genome with interspersed repeats, and masked 56.48% of the genome as repeat-derived in the Chinook salmon [54], etc. These high proportions of repeats belong to the highest ones among fishes and found in any vertebrate [51,55], similar to the only salamander species with a sequenced genome to date, the axolotl (*A. mexicanum*) with its 65.6% of repetitive sequences [56]. The impact of repeats on the GC% of the entire genome might be similar to that in zebrafish and some other cyprinids, where repeats occur in high proportions as well, 52.2% in Zv9 [41]. However, repeats in most cyprinids are GC-poor [30], whereas those in salmonids might be GC-rich, a possibility that needs to be explored. Both Class I and II transposons are almost equally represented in the sequenced salmonids. Therefore, it is currently impossible to ascribe the potential GC-enrichment to one of them as the dominant one *sensu* [30]. In axolotl, however, distinct long terminal repeat (LTR) retroelement classes and endogenous retroviruses made up the largest portion of the repetitive sequences and included elements of more than 10 kb in length [56]. This shows the broad repertoire in repeats composition in lineages with giant genomes and the need to explore them. The entire genomics of polyploid animals is still in its infancy and hence, important new facts can be expected soon [57] since new genomes have been sequenced in the meanwhile and others are on the way.

#### 4.3. Nucleotide Composition Investigations in Fish

The currently most plausible findings link the genome GC content with the chromosome size and with life history traits, where the effective population size ( $N_e$ ) is the most important indicator of the strength of the gBGC [18,20]. For fish, there is no systematic analysis of GC% based on purely genomic data as are available for mammals [20–22], birds [23,24], and reptiles [25,26]. One study assessing GC% based on non-genomic determination showed a slightly higher GC% in marine and migrating marine fish species [58]. Hence, we can merely indirectly assess the above mentioned factors with correlatives with GC% known for fish. Genome size is one of them and it is negatively correlated with the genomic GC% in fish except for salmonids [30]. Genome size was shown to negatively correlate also with the  $N_e$  and this association was demonstrated to have resulted in a larger genome in freshwater ray-finned fish [59]. This is in line with the aforementioned results on the effect of environment and lifestyle in fish [58]. However, the reduced  $N_e$  presumably resulting in larger and more complex genomes in freshwater fish [59] are contradictory to the situation in chondrichthyans, which are marine [40] and possess large genomes [43], but were not included in the study [59]. Another recent study indicated that the fish species distributed at higher latitude might have a small long-term  $N_e$  [60]. A link between  $N_e$  and the transposon evolution/transposition intensity was also indicated [61,62]. Moreover,



there is a highly complex and ambiguous relationship between the transposons insertion and accumulation and the recombination rate (summarized in [63]).

Crucial results come from the compositional biology of mammalian transposons. There are indications that the base composition is evolving under selection and may be reflective of the long-term co-evolution between non-LTR retrotransposons and their host [64]. These authors also hypothesise that the coexistence of elements with drastically different base compositions suggests that these elements may be using different strategies to persist and multiply in the genome of their host. Similarly, other authors proposed that in addition to gBGC, there may be additional, still uncharacterized molecular mechanisms that either preserve genomic regions with biased nucleotide compositions from mutational degradation or fail to degrade such inhomogeneities in specific chromosomal regions [65].

## 5. Conclusions

To sum up, fishes show an immense diversity in their genome organisation and in the mutual relationships between their major genomic traits (GC% and genome size). On the other hand, their diploid chromosome numbers are relatively conserved [8], which means that their chromosome sizes need to decrease with the genome compaction observed in modern fish lineages. This resulted in an increased recombination rate as evidenced in the literature [19] and in a GC-enrichment, however, without AT/GC heterogeneity. Hence, modern fish lineages reached the smallest genome sizes among vertebrates despite their teleost-specific whole genome duplication [44]. Birds increased their absolute GC% and acquired or retained their AT/GC heterogeneity and substantially decreased their genome size, at least partly thanks to their microchromosomes (they increase the recombination rate that leads to genome compaction [66]). In contrast, basal fish lineages with microchromosomes do show a certain GC-enrichment, but they retained larger genome sizes despite the increased recombination rate mediated by microchromosomes that should have led to genome compaction similarly as in birds. Finally, mammals substantially expanded their genomes without any additional whole genome duplication and acquired or retained the AT/GC heterogeneity without any significant increase in their GC%. Therefore, mammals rather redistributed their genomes to GC-rich and GC-poor regions and accumulated them separately based on their GC%. The exact role and extent of GC-biased gene conversion in shaping the genome composition across vertebrates needs to be clarified and further investigated with the newly available resources. The quantitative approach to this issue, presented here, opens new possibilities to address related questions.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/2073-4425/12/2/312/s1>. Figure S1: Meta-analysis of GC% (A) and genome size (B) in fish (red) and mammals (green). Data from NCBI/genome and [www.genomesize.com](http://www.genomesize.com). Each point represents an entire genome; Figure S2: GC% of individual chromosomes in one lancelet (left, pink) and six ray-finned fishes (six blue violin plots) with diversified associations between the chromosome size and GC% compared with seven mammals (green violin plots). The overall comparison of GC% and genome size in fish and mammals can be found in Figure 1. The colour-code corresponds to Figure 1; Figure S3: GC% of individual chromosomes in lineages with microchromosomes—four ancient fish and in reptiles and birds. The left four violin plots are two chondrichthyans (thorny skate and bamboo shark in black), two basal ray-finned fishes (sterlet and spotted gar in blue), two violin plots in the middle are reptiles (rattlesnake and tortoise purple), and four right violin plots are birds (the Muscovy duck, chicken, the American flamingo, and zebra finch in yellow). The colour code corresponds to Figure 1; Figure S4: *Clupea harengus* GC% vs. repeats% in each chromosome. The chromosome size decreases from 33 to 12.4 Mb. Neither substantial changes in repeats% nor GC% of repeats occur with the decreasing chromosome size as predicted by the gBGC concept. A similar pattern was identified in all the three salmonids analysed here. The vertical line represents the median of rep%; Figure S5: *Danio rerio* GC% vs. repeats% in each chromosome. The chromosome size decreases from 78 to 37.5 Mb. Such a pattern has so far been identified only in this species. The vertical line represents the median of rep%; Figure S6: *Lepisosteus oculatus* GC% vs. repeats% in each chromosome. The chromosome size decreases from 73.2 to 5.4 Mb (we excluded the smallest chromosome with size 294 kb). Such a

pattern occurs in numerous species; Figure S7: *Erpetoichthys calabaricus* GC% vs. repeats% in each chromosome. The chromosome size decreases from 350 to 88 Mb. The vertical line represents the median of rep%.

**Author Contributions:** Conceptualization, R.S., W.M.H. and V.B.; methodology, R.S., V.B. and D.M.; software, D.M.; validation, V.B.; resources, R.S.; data curation, D.M.; writing—original draft preparation, R.S. and W.M.H.; writing—review and editing, R.S., W.M.H. and V.B.; visualization, V.B. and R.S.; project administration, R.S.; funding acquisition, R.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 754462. The APC was funded by the Erasmus + programme of the European Union with contract No. 2019-1-CZ01-KA203-061433.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** [https://github.com/bioinfohk/evangelist\\_plots](https://github.com/bioinfohk/evangelist_plots) (accessed on 31 January 2021).

**Acknowledgments:** Computational resources were supplied by the project “e-Infrastruktura CZ” (e-INFRA LM2018140) provided within the program Projects of Large Research, Development, and Innovations Infrastructures.

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

## References

- Gregory, T.R. Animal Genome Size Database. Available online: <http://www.genomesize.com> (accessed on 31 January 2021).
- Levan, A.; Fredga, K.; Sandberg, A.A. Nomenclature for Centromeric Position on Chromosomes. *Hereditas* **1964**, *52*, 201–220. [[CrossRef](#)]
- Mark, H.F.; Mark, R.; Pan, T.; Mark, Y. Centromere Index Derivation by a Novel and Convenient Approach. *Ann. Clin. Lab. Sci.* **1993**, *23*, 267–274.
- Comings, D.E. Mechanisms of Chromosome Banding and Implications for Chromosome Structure. *Annu. Rev. Genet.* **1978**, *12*, 25–46. [[CrossRef](#)] [[PubMed](#)]
- Luo, C. Multiple Chromosomal Banding in Grass Carp, *Ctenopharyngodon idellus*. *Heredity* **1998**, *81*, 481–485. [[CrossRef](#)]
- Medrano, L.; Bernardi, G.; Couturier, J.; Dutrillaux, B.; Bernardi, G. Chromosome Banding and Genome Compartmentalization in Fishes. *Chromosoma* **1988**, *96*, 178–183. [[CrossRef](#)]
- Mayr, B.; Kalat, M.; Ráb, P.; Lambrou, M. Band Karyotypes and Specific Types of Heterochromatins in Several Species of European Percid Fishes (Percidea, Pisces). *Genetica* **1987**, *75*, 199–205. [[CrossRef](#)]
- Mank, J.E.; Avise, J.C. Phylogenetic Conservation of Chromosome Numbers in Actinopterygian Fishes. *Genetica* **2006**, *127*, 321–327. [[CrossRef](#)] [[PubMed](#)]
- Gregory, T.R.; Witt, J.D.S. Population Size and Genome Size in Fishes: A Closer Look. *Genome* **2008**, *51*, 309–313. [[CrossRef](#)]
- Hardie, D.C.; Hebert, P.D.N. The Nucleotypic Effects of Cellular DNA Content in Cartilaginous and Ray-Finned Fishes. *Genome* **2003**, *46*, 683–706. [[CrossRef](#)]
- Hardie, D.C.; Hebert, P.D. Genome-Size Evolution in Fishes. *Can. J. Fish. Aquat. Sci.* **2004**, *61*, 1636–1646. [[CrossRef](#)]
- Melodelima, C.; Gautier, C. The GC-Heterogeneity of Teleost Fishes. *BMC Genom.* **2008**, *9*, 632. [[CrossRef](#)] [[PubMed](#)]
- Costantini, M.; Auletta, F.; Bernardi, G. Isochore Patterns and Gene Distributions in Fish Genomes. *Genomics* **2007**, *90*, 364–371. [[CrossRef](#)]
- Costantini, M.; Cammarano, R.; Bernardi, G. The Evolution of Isochore Patterns in Vertebrate Genomes. *BMC Genom.* **2009**, *10*, 146. [[CrossRef](#)] [[PubMed](#)]
- Eyre-Walker, A. Recombination and Mammalian Genome Evolution. *Proc. R. Soc. Lond. B Biol. Sci.* **1993**, *252*, 237–243. [[CrossRef](#)]
- Fullerton, S.M.; Bernardo Carvalho, A.; Clark, A.G. Local Rates of Recombination Are Positively Correlated with GC Content in the Human Genome. *Mol. Biol. Evol.* **2001**, *18*, 1139–1142. [[CrossRef](#)] [[PubMed](#)]
- Montoya-Burgos, J.I.; Boursot, P.; Galtier, N. Recombination Explains Isochores in Mammalian Genomes. *Trends Genet.* **2003**, *19*, 128–130. [[CrossRef](#)]
- Mugal, C.F.; Weber, C.C.; Ellegren, H. GC-Biased Gene Conversion Links the Recombination Landscape and Demography to Genomic Base Composition: GC-Biased Gene Conversion Drives Genomic Base Composition across a Wide Range of Species. *BioEssays* **2015**, *37*, 1317–1326. [[CrossRef](#)] [[PubMed](#)]
- Stapley, J.; Feulner, P.G.D.; Johnston, S.E.; Santure, A.W.; Smadja, C.M. Variation in Recombination Frequency and Distribution across Eukaryotes: Patterns and Processes. *Philos. Trans. R. Soc. B Biol. Sci.* **2017**, *372*, 20160455. [[CrossRef](#)]



20. Romiguier, J.; Ranwez, V.; Douzery, E.J.P.; Galtier, N. Contrasting GC-Content Dynamics across 33 Mammalian Genomes: Relationship with Life-History Traits and Chromosome Sizes. *Genome Res.* **2010**, *20*, 1001–1009. [[CrossRef](#)]
21. Huttener, R.; Thorrez, L.; In't Veld, T.; Granvik, M.; Snoeck, L.; Van Lommel, L.; Schuit, F. GC Content of Vertebrate Exome Landscapes Reveal Areas of Accelerated Protein Evolution. *BMC Evol. Biol.* **2019**, *19*. [[CrossRef](#)] [[PubMed](#)]
22. Shen, W.; Wang, D.; Ye, B.; Shi, M.; Ma, L.; Zhang, Y.; Zhao, Z. GC3-Biased Gene Domains in Mammalian Genomes. *Bioinformatics* **2015**, *31*, 3081–3084. [[CrossRef](#)] [[PubMed](#)]
23. Weber, C.C.; Boussau, B.; Romiguier, J.; Jarvis, E.D.; Ellegren, H. Evidence for GC-Biased Gene Conversion as a Driver of between-Lineage Differences in Avian Base Composition. *Genome Biol.* **2014**, *15*. [[CrossRef](#)]
24. Bolívar, P.; Mugal, C.F.; Nater, A.; Ellegren, H. Recombination Rate Variation Modulates Gene Sequence Evolution Mainly via GC-Biased Gene Conversion, Not Hill–Robertson Interference, in an Avian System. *Mol. Biol. Evol.* **2016**, *33*, 216–227. [[CrossRef](#)] [[PubMed](#)]
25. Matsubara, K.; Kuraku, S.; Tarui, H.; Nishimura, O.; Nishida, C.; Agata, K.; Kumazawa, Y.; Matsuda, Y. Intra-Genomic GC Heterogeneity in Sauropsids: Evolutionary Insights from cDNA Mapping and GC3 Profiling in Snake. *BMC Genom.* **2012**, *13*, 604. [[CrossRef](#)]
26. Figuet, E.; Ballenghien, M.; Romiguier, J.; Galtier, N. Biased Gene Conversion and GC-Content Evolution in the Coding Sequences of Reptiles and Vertebrates. *Genome Biol. Evol.* **2015**, *7*, 240–250. [[CrossRef](#)]
27. Wang, D. GCevobase: An Evolution-Based Database for GC Content in Eukaryotic Genomes. *Bioinformatics* **2018**, *34*, 2129–2131. [[CrossRef](#)] [[PubMed](#)]
28. Frenkel, S.; Kirzhner, V.; Korol, A. Organizational Heterogeneity of Vertebrate Genomes. *PLoS ONE* **2012**, *7*, e32076. [[CrossRef](#)]
29. Li, X.-Q.; Du, D. Variation, Evolution, and Correlation Analysis of C+G Content and Genome or Chromosome Size in Different Kingdoms and Phyla. *PLoS ONE* **2014**, *9*, e88339. [[CrossRef](#)] [[PubMed](#)]
30. Symonová, R.; Suh, A. Nucleotide Composition of Transposable Elements Likely Contributes to AT/GC Compositional Homogeneity of Teleost Fish Genomes. *Mob. DNA* **2019**, *10*. [[CrossRef](#)]
31. Carducci, F.; Barucca, M.; Canapa, A.; Carotti, E.; Biscotti, M.A. Mobile Elements in Ray-Finned Fish Genomes. *Life* **2020**, *10*, 221. [[CrossRef](#)] [[PubMed](#)]
32. Bernardi, G. The Vertebrate Genome: Isochores and Evolution. *Mol. Biol. Evol.* **1993**. [[CrossRef](#)]
33. Symonová, R.; Majtánová, Z.; Arias-Rodriguez, L.; Mořkovský, L.; Kořínková, T.; Cavin, L.; Pokorná, M.J.; Doležalková, M.; Flajšhans, M.; Normandeau, E.; et al. Genome Compositional Organization in Gars Shows More Similarities to Mammals than to Other Ray-Finned Fish: Cytogenomics of Gars. *J. Exp. Zool. Part B Mol. Dev. Evol.* **2017**, *328*, 607–619. [[CrossRef](#)] [[PubMed](#)]
34. Matoulek, D.; Borůvková, V.; Ocalewicz, K.; Symonová, R. GC and Repeats Profiling along Chromosomes—The Future of Fish Compositional Cytogenomics. *Genes* **2021**, *12*, 50. [[CrossRef](#)] [[PubMed](#)]
35. Bernardi, G. *Structural and Evolutionary Genomics Natural Selection in Genome Evolution*; Elsevier: Amsterdam, The Netherlands, 2005.
36. Peona, V.; Weissensteiner, M.H.; Suh, A. How Complete Are “Complete” Genome Assemblies?—An Avian Perspective. *Mol. Ecol. Resour.* **2018**, *18*, 1188–1195. [[CrossRef](#)] [[PubMed](#)]
37. Jebb, D.; Huang, Z.; Pippel, M.; Hughes, G.M.; Lavrichenko, K.; Devanna, P.; Winkler, S.; Jermin, L.S.; Skirmuntt, E.C.; Katzourakis, A.; et al. Six Reference-Quality Genomes Reveal Evolution of Bat Adaptations. *Nature* **2020**, *583*, 578–584. [[CrossRef](#)] [[PubMed](#)]
38. R Core Team. *R: A Language and Environment for Statistical Computing*; Version 2.6.2; R Core Team: Vienna, Austria, 2013.
39. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*, 2nd ed.; Use R! Springer International Publishing: Cham, Switzerland, 2016; ISBN 978-3-319-24277-4.
40. Nelson, J.S.; Grande, T.; Wilson, M.V.H. *Fishes of the World*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016; ISBN 978-1-118-34233-6.
41. Howe, K.; Clark, M.D.; Torroja, C.F.; Torrance, J.; Berthelot, C.; Muffato, M.; Collins, J.E.; Humphray, S.; McLaren, K.; Matthews, L.; et al. The Zebrafish Reference Genome Sequence and Its Relationship to the Human Genome. *Nature* **2013**, *496*, 498–503. [[CrossRef](#)] [[PubMed](#)]
42. Hidalgo, O.; Pellicer, J.; Christenhusz, M.; Schneider, H.; Leitch, A.R.; Leitch, I.J. Is There an Upper Limit to Genome Size? *Trends Plant Sci.* **2017**, *22*, 567–573. [[CrossRef](#)] [[PubMed](#)]
43. NCBI Genome Browser. Available online: <https://www.ncbi.nlm.nih.gov/genome/browse> (accessed on 31 January 2021).
44. Meyer, A.; Schartl, M. Gene and Genome Duplications in Vertebrates: The One-to-Four (-to-Eight in Fish) Rule and the Evolution of Novel Gene Functions. *Curr. Opin. Cell Biol.* **1999**, *11*, 699–704. [[CrossRef](#)]
45. Hannan, A.J. Tandem Repeats and Repeatomes: Delving Deeper into the ‘Dark Matter’ of Genomes. *EBioMedicine* **2018**, *31*, 3–4. [[CrossRef](#)] [[PubMed](#)]
46. Arkhipova, I.R.; Yushenova, I.A. Giant Transposons in Eukaryotes: Is Bigger Better? *Genome Biol. Evol.* **2019**, *11*, 906–918. [[CrossRef](#)]
47. Fontana, F.; Bruch, R.M.; Binkowski, F.P.; Lanfredi, M.; Chicca, M.; Beltrami, N.; Congiu, L. Karyotype Characterization of the Lake Sturgeon, *Acipenser fulvescens* (Rafinesque 1817) by Chromosome Banding and Fluorescent in Situ Hybridization. *Genome* **2004**, *47*, 742–746. [[CrossRef](#)]

48. Symonová, R.; Havelka, M.; Amemiya, C.T.; Howell, W.M.; Kořínková, T.; Flajšhans, M.; Gela, D.; Ráb, P. Molecular Cytogenetic Differentiation of Paralogs of Hox Paralogs in Duplicated and Re-Diploidized Genome of the North American Paddlefish (*Polyodon spathula*). *BMC Genet.* **2017**, *18*. [[CrossRef](#)]
49. Majtánová, Z.; Symonová, R.; Arias-Rodriguez, L.; Sallan, L.; Ráb, P. “Holostei versus Halecostomi” Problem: Insight from Cytogenetics of Ancient Nonteleost Actinopterygian Fish, Bowfin *Amia calva*: Molecular Cytogenetics of *Amia calva*. *J. Exp. Zool. Part B Mol. Dev. Evol.* **2017**, *328*, 620–628. [[CrossRef](#)]
50. Gaffaroglu, M.; Majtánová, Z.; Symonová, R.; Pelikánová, Š.; Unal, S.; Lajbner, Z.; Ráb, P. Present and Future Salmonid Cytogenetics. *Genes* **2020**, *11*, 1462. [[CrossRef](#)]
51. Lien, S.; Koop, B.F.; 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. [[CrossRef](#)]
52. De-Kayne, R.; Zoller, S.; Feulner, P.G.D. A de Novo Chromosome-level Genome Assembly of *Coregonus* sp. “Balchen”: One Representative of the Swiss Alpine Whitefish Radiation. *Mol. Ecol. Resour.* **2020**, *20*, 1093–1109. [[CrossRef](#)]
53. Pearse, D.E.; Barson, N.J.; Nome, T.; Gao, G.; Campbell, M.A.; Abadía-Cardoso, A.; Anderson, E.C.; Rundio, D.E.; Williams, T.H.; Naish, K.A.; et al. Sex-Dependent Dominance Maintains Migration Supergene in Rainbow Trout. *Nat. Ecol. Evol.* **2019**, *3*, 1731–1742. [[CrossRef](#)] [[PubMed](#)]
54. Christensen, K.A.; Leong, J.S.; Sakhrani, D.; Biagi, C.A.; Minkley, D.R.; Withler, R.E.; Rondeau, E.B.; Koop, B.F.; Devlin, R.H. Chinook Salmon (*Oncorhynchus tshawytscha*) Genome and Transcriptome. *PLoS ONE* **2018**, *13*, e0195461. [[CrossRef](#)]
55. 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](#)] [[PubMed](#)]
56. Nowoshilow, S.; Schloissnig, S.; Fei, J.-F.; Dahl, A.; Pang, A.W.C.; Pippel, M.; Winkler, S.; Hastie, A.R.; Young, G.; Roscito, J.G.; et al. The Axolotl Genome and the Evolution of Key Tissue Formation Regulators. *Nature* **2018**, *554*, 50–55. [[CrossRef](#)]
57. Rodriguez, F.; Arkhipova, I.R. Transposable Elements and Polyploid Evolution in Animals. *Curr. Opin. Genet. Dev.* **2018**, *49*, 115–123. [[CrossRef](#)] [[PubMed](#)]
58. Tarallo, A.; Angelini, C.; Sanges, R.; Yagi, M.; Agnisola, C.; D’Onofrio, G. On the Genome Base Composition of Teleosts: The Effect of Environment and Lifestyle. *BMC Genom.* **2016**, *17*. [[CrossRef](#)]
59. Yi, S.; Streelman, J.T. Genome Size Is Negatively Correlated with Effective Population Size in Ray-Finned Fish. *Trends Genet.* **2005**, *21*, 643–646. [[CrossRef](#)] [[PubMed](#)]
60. Rolland, J.; Schluter, D.; Romiguier, J. Vulnerability to Fishing and Life History Traits Correlate with the Load of Deleterious Mutations in Teleosts. *Mol. Biol. Evol.* **2020**, *37*, 2192–2196. [[CrossRef](#)] [[PubMed](#)]
61. Tollis, M.; Boissinot, S. The Evolutionary Dynamics of Transposable Elements in Eukaryote Genomes. In *Genome Dynamics*; Garrido-Ramos, M.A., Ed.; S. KARGER AG: Basel, Switzerland, 2012; Volume 7, pp. 68–91. ISBN 978-3-318-02149-3.
62. Bourgeois, Y.; Boissinot, S. On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. *Genes* **2019**, *10*, 419. [[CrossRef](#)] [[PubMed](#)]
63. Kent, T.V.; Uzunović, J.; Wright, S.I. Coevolution between Transposable Elements and Recombination. *Philos. Trans. R. Soc. B Biol. Sci.* **2017**, *372*, 20160458. [[CrossRef](#)]
64. Ruggiero, R.P.; Boissinot, S. Variation in Base Composition Underlies Functional and Evolutionary Divergence in Non-LTR Retrotransposons. *Mob. DNA* **2020**, *11*. [[CrossRef](#)]
65. Paudel, R.; Fedorova, L.; Fedorov, A. Adapting Biased Gene Conversion Theory to Account for Intensive GC-Content Deterioration in the Human Genome by Novel Mutations. *PLoS ONE* **2020**, *15*, e0232167. [[CrossRef](#)]
66. Nam, K.; Ellegren, H. Recombination Drives Vertebrate Genome Contraction. *PLoS Genet.* **2012**, *8*, e1002680. [[CrossRef](#)] [[PubMed](#)]