

A Duplicate Resolved Paddlefish Genome Provides Insights into the Mechanisms of Rediploidisation and Hox Cluster Evolution

Dearbhaile Casey¹, Łukasz Niezabitowski¹, Manu Kumar Gundappa², Hanover Matz³, Arun Venugopalan⁴, Larry A. Hanson⁵, Helen Dooley³, Daniel J. Macqueen⁶, Anthony K. Redmond^{7*}, and Aoife McLysaght^{1*}

¹Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland

²Animal Breeding and Genomics, Wageningen University & Research, Wageningen, The Netherlands

³Department of Microbiology and Immunology, University of Maryland School of Medicine, Institute of Marine and Environmental Technology, Baltimore, MD, USA

⁴Department of Pathobiology and Population Medicine, College of Veterinary Medicine, Mississippi State University, Starkville, MS, USA

⁵Department of Comparative Biomedical Sciences, College of Veterinary Medicine, Mississippi State University, Starkville, MS, USA

⁶The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, Scotland

⁷School of Medicine, University College Dublin, Dublin, Ireland

*Correspondence to anthony.redmond@ucd.ie, aoife.mclysaght@tcd.ie

Abstract

Whole-genome duplication (WGD; or polyploidy) has played a major role in the evolution of many lineages however, our understanding of the processes that shape genome evolution following WGD remains incomplete. While polyploidy duplicates the entire genome sequence, it is rediploidisation that establishes independent duplicated genes. Rediploidisation proceeds through suppression of meiotic recombination across polysomic loci thus restoring disomic inheritance, a process that is not synchronised across the genome. Despite its importance, the mechanisms underlying this process remain poorly understood. The slowly evolving genomes of paleopolyploid Acipenseriformes paddlefish and sturgeon provide an invaluable system for investigating this, as rediploidisation was highly asynchronous in these lineages. In both genomes ohnologs tend to segregate into blocks on the chromosomes according to rediploidisation timing, a pattern that suggests links between chromosomal structure and rediploidisation. Here, we analyse a newly-produced duplicate-resolved paddlefish genome assembly and show a strong concordance between genome rearrangement and rediploidisation timing. We also find that topologically associated domain (TAD) boundaries are associated with rediploidisation block boundaries. Together these results indicate that rediploidisation in acipenseriformes occurred through a process of genome rearrangements that was subject to functional constraints imposed by 3D genome architecture. We investigate the evolution of Hox clusters in these lineages, revealing a previously overlooked duplicate HoxC region in paddlefish, and both ancestral and lineage-specific Hox cluster rediploidisation with substantially different timings. These findings highlight a complex evolutionary history following WGD in Acipenseriformes with implications for understanding short-term adaptations to polyploidy as well as longer-term diversification of lineages.

Introduction

Ancient whole genome duplications (WGD) have occurred in many eukaryotic lineages (1, 16, 37, 49, 50, 74, 81, 87, 88). These doubling events are thought to offer raw genetic material with the potential to facilitate evolutionary innovations in descendant species (83). The process of rediploidisation is crucial to unlocking this potential.

WGD can occur by self-doubling of the genome (autopolyploidisation) or through hybridisation of two different parent species (allopolyploidisation). Cytogenetically, in many cases, the latter may be diploid-on-arrival (i.e. no classical rediploidisation step is required) as the chromosomes are likely to be sufficiently diverged to form preferential bivalent pairs at meiosis. However, in autopolyploids there is no sequence differentiation and the chromosomes will experience polysomic inheritance due to the occurrence of meiotic recombination either multivalently, or in non-preferential bivalents (such that any of the four chromosomes may recombine with any of the others). This situation persists until recombination becomes suppressed. The suppression of recombination establishes independent duplicated loci from the previously polysomic locus; that is, it creates duplicated genes (ohnologs) where previously there were merely additional alleles (70) (Redmond et al. in review) (56). Suppression of meiotic recombination is the crux of rediploidisation, establishing independent duplicated loci, free to undergo uninterrupted sequence and functional divergence (70, 71).

Seen this way, rediploidisation is key to unlocking much of the potential of WGD for evolutionary novelty and diversification (14, 71). Despite being such a critical process in post-WGD eukaryotic evolution, the mechanisms of rediploidisation are poorly understood. The most popular hypothesis is that rediploidisation is brought about by genomic rearrangements and the accumulation of mutations along homologous chromosomes (25, 50). However, evidence in support of genome rearrangements as a mechanism of rediploidisation is scant, though a recent example implicates chromosomal fusions as instigators of rediploidisation in autotetraploid snow carp (90).

Investigating the patterns and mechanisms of rediploidisation requires suitable study

systems where the WGD is sufficiently old enough so that rediploidisation has progressed, but recent enough so that events can still be tracked through evolutionary time. The shared WGD at the base of the extant Acipenseriformes (sturgeons and paddlefish) provides an ideal system. Noted for their slow evolutionary rate (10, 29, 43) they have recently been established as a model of slow, asynchronous rediploidisation. Despite having shared an autopolyploidisation event well over 200Ma (million years ago), more than half of the genome remained tetraploid (i.e. under tetrasomic inheritance) until after the divergence of the two acipenseriform lineages (~170 Ma in the Jurassic) (70). Their genomes are thus a mosaic of ohnolog pairs that rediploidised either before (PreSpec; pre-speciation) or after (PostSpec; post-speciation) the speciation event separating sturgeons and paddlefish, known as AORe (Ancestral Ohnolog Resolution) and LORe (Lineage-specific Ohnolog resolution) ohnolog pairs respectively (70, 71). Importantly, ohnolog pairs with a shared rediploidisation history are not randomly distributed across the genome, instead, they form distinct syntenic blocks along uninterrupted sections of chromosomes (70). Based on this, Redmond et al. (2023) proposed a model for rediploidisation where the structural differences introduced by genome rearrangements interrupt meiotic pairing and homologous recombination in that genomic region, such that rediploidisation proceeds rearrangement by rearrangement along and between chromosomes (Fig 1 (a)). This model is conceptually similar to the establishment of mammalian sex chromosomes which formed by punctuated rearrangement events, each progressively suppressing X-Y recombination and establishing ‘evolutionary strata’ of chromosome divergence (31, 44). While rearrangements appear to be a plausible explanation for the stratified topologies of paddlefish and sturgeon given that blocks with shared histories are apparent in visual inspection in a circos plot (1(b)(i)), a formal investigation of this model of rediploidisation is still lacking.

Intriguingly, structuro-functional constraints imposed by 3D genome architecture have been proposed to influence chromosomal rearrangement possibilities, though this has never been linked to the pattern of asynchronous rediploidisation across the genome. For example, topologically associating domains (TADs) are regions of elevated intra-domain

chromatin contact that often contain genes with shared transcriptional programmes and are thus thought to play important roles in genome organisation and function (52). Importantly, disruption of TAD boundaries has in some cases been linked to negative outcomes including oncogenesis, embryonic lethality, and developmental disease (58, 82). Some TADs have been demonstrated to be evolutionarily conserved, and to often harbour age-matched genes, including those retained following WGD, and have vitally been shown to be resistant to disruption by genomic rearrangements (38, 64, 84). This raises the possibility that preserving normal 3D genome organisation may selectively constrain viable chromosomal rearrangement space and, consequently, influence the patterns and rates of rediploidisation following WGD.

To address key open questions highlighted above, we first improved the assembly of the American paddlefish (*Polyodon spathula*) genome by resolving mistakenly collapsed regions (17) and re-annotated it using newly generated RNA-seq data. Assembly enhancement was necessary because these genomes are exceptionally slow evolving, causing many ohnologous genes and genomic regions to be erroneously collapsed into single assembly loci. Ohnologs that underwent rediploidisation after the divergence of paddlefish and sturgeons –and therefore exhibit the lowest sequence divergence– were, as expected, the most severely affected (70). Using this enhanced assembly, we identified micro- and macro-synteny blocks of ohnologs as proxies for rearrangements and found that both rearrangement breakpoints and TAD boundaries coincide with changes in rediploidisation history along the chromosome (which we dub ‘topology breakpoints’; Fig 2 (a)). Patterns of 3D genome architecture appear to constrain the genomic locations at which rearrangements occur, and functional analyses further reveal distinct enrichment patterns among gene categories associated with early and late rediploidisation ohnologs. Building on this, we then investigated the rediploidisation patterns of the acipenseriform *Hox* clusters. Previous work has shown that the duplicate *HoxA*, *HoxB*, and *HoxD* gene clusters are largely retained in Acipenseriformes, whereas the duplicate *HoxC* cluster was inferred to be lost in the paddlefish but retained in sturgeon, echoing similar losses observed in some elasmobranch

lineages (12, 15, 41). By reconstructing Hox cluster evolution, we find major differences in rediploidisation histories among clusters, and reveal a previously misassembled, degenerate (rather than lost) duplicate *HoxC* region in paddlefish. We speculate that variation in the timing of *Hox* cluster duplication (that is, rediploidisation) may have contributed to phenotypic divergence between these sister lineages, perhaps consistent with hypotheses linking polyploidy and asynchronous rediploidisation to adaptive potential during periods of environmental instability (70).

Results

Characterising ohnologs with a new, duplicate-resolved paddlefish genome

Numerous ohnolog pairs appear to have been erroneously collapsed into into singleton assembly regions in the paddlefish genome (70). In line with this, the sturgeon genome is reported to have retained thousands of genes after WGD that were proposed to be lost in paddlefish, including an additional *HoxC* cluster (6, 15, 17), however, we considered that many of these may in fact be assembly error artefacts. As such we first sought to produce an improved paddlefish genome assembly in order to determine if the elevated rate of gene loss reported in paddlefish can be wholly or partially ascribed to error.

We found that approximately 2.3% of the paddlefish genome shows elevated read depth, consistent with collapsed regions (Fig 1(c)(ii)). Among genes that are duplicated in sturgeon but recorded as single-copy in paddlefish, there is a secondary read depth peak at double coverage suggesting that the collapsed genomic regions contain ohnolog pairs (Fig 1 (c)(ii); Supplementary Materials (22)). Adapting a reduplication strategy from (17) to avoid ohnologous region collapse we re-assembled the paddlefish genome, and indeed the secondary peak of supposed singletons with double read depth disappears (Fig 1 (c)(iii)). This improved assembly and genome annotation was significantly enhanced by supplementing with new RNA-seq data from 15 tissues from 4 individuals and integrating

three annotation strategies to predict 35,930 protein-coding genes with BRAKER3 (30).
This is a remarkable increase of 9,859 genes compared to the original assembly (12).
However, this still falls over 10,000 genes short of the predicted protein-coding gene
count in the sterlet sturgeon, implying that although previous estimates of paddlefish
protein-coding genes were erroneously low, the gene loss rate after their shared WGD was
indeed higher in paddlefish than in sturgeon.

Building on the work of Redmond et al. (70), we classified ohnolog gene trees based
on topology as either PreSpec (shared duplication node) or PostSpec (independent
duplication nodes). By integrating phylogenetic and syntenic evidence, we recovered
patterns consistent with those reported by Redmond et al. (70), while also identifying
approximately 423 additional ohnolog pairs. PostSpec topologies were more prevalent
than PreSpec topologies, with the remaining gene trees exhibiting intermediate histories
(PreSpec-like or PostSpec-like) (Supplementary Table S1), consistent with Redmond et al.
2023. The discovery of these additional genes and ohnolog pairs from our new assembly
should allow for better analysis of synteny and its link to rediploidisation in paddlefish (70).

Associating rearrangement breakpoints with rediploidisation events

To identify whether chromosomal rearrangements are associated with rediploidisation
events, we first needed to locate syntenic blocks of genes with consistent rediploidisation
histories. These blocks are noticeable by eye in the circos plots (Fig 1 (c)(i)), especially on
the larger chromosomes, where extended stretches of consecutive blue and red links (joining
prespec or postspec ohnolog pairs, respectively) occur sequentially along chromosomes.

To formally identify genomic locations where the syntenic chain of ohnolog pair gene
trees changes topology (henceforth referred to as topology breakpoints) we applied a
Bayesian segmentation algorithm (Niezabitowski et al., BIORXIV). Briefly, our approach
models each chromosome as a sequence of categorical observations (prespec or postspec)
with posterior probabilities, and a ‘breakpoint’ is defined as the first element in the

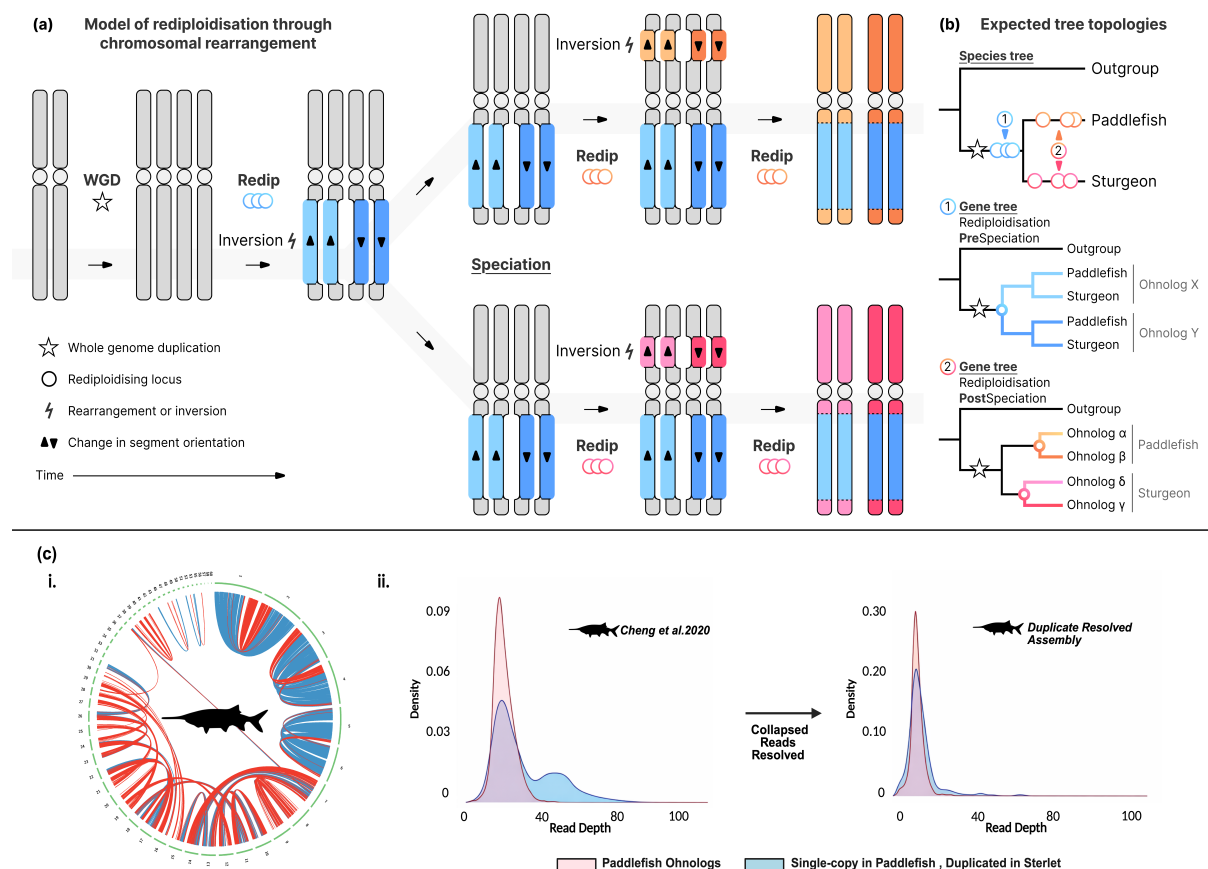


Figure 1. Model for rediploidisation through chromosomal rearrangement and resolving collapsed reads in the paddlefish genome.

(a) The schematic shows diploid chromosomes that duplicate via WGD, resulting in a tetraploid. A meiotic rearrangement (inversion) suppresses recombination and promotes their rediploidisation (blue loci). These ancestral duplication events give rise to *PreSpec* gene tree topologies in phylogenetic analyses (far right, (b)(1)). Following speciation, independent rearrangements occur in each species, shown in shades of orange for paddlefish and shades of pink for sturgeon. These generate lineage-specific ohnologs and *PostSpec* gene tree topologies (far right, (b)(2)). **(c)(i)** Circos plot showing *PreSpec* ohnolog pairs as blue links and *PostSpec* ohnolog pairs as red links in the Paddlefish Genome **(ii)** Read-depth distributions for paddlefish ohnolog pairs (red) and putative ohnologs (blue). In the published genome assembly (left), genes present as single copies in paddlefish but duplicated in sterlet show a peak at approximately $\sim 2\times$ the average coverage, indicating that duplicated loci were collapsed into a single genomic region during assembly. In the duplicate-resolved assembly (right), this peak is substantially reduced. Read depth is expressed as fold coverage ($\times$), where $1\times$ represents the genome-wide average sequencing coverage per base.

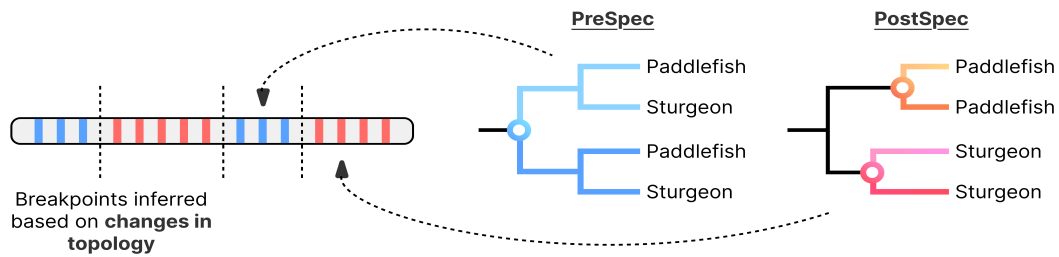
sequence where the probability changes between topology category (see NIEZABITOWSKI 138
BIORXIV for a full explanation of this approach). Using this method, we identified a 139
total of 383 topology breakpoints across the 60 largest chromosomes of paddlefish and 140
293 breakpoints across the 60 largest sterlet chromosomes (Supplementary Table S2; and 141
Figshare (22)). 142

We used synteny block co-ordinates as a proxy for genome rearrangement breakpoints. 143
We identified pairwise micro-synteny blocks between the genomes using *i-ADHoRe* (67) 144
(Fig 2 (b)(i)). These blocks represent regions where both gene order and orientation have 145
been retained or jointly modified. We also identified macro-synteny blocks shared by both 146
species using a method developed in (61, 62) and adapted in Niezabitowski et al., 2026 147
(Fig 2 (a)(i&iii)). 148

We next related these synteny blocks to the topology breakpoints (Fig 2 (b) (ii), 149
(70); Supplementary Materials; Table 1 (22)) to assess whether genome rearrangement 150
events were spatially associated with transitions in rediploidisation history. Specifically, 151
we quantified the genomic distance between each synteny block boundary and the nearest 152
topology breakpoint, and compared this to a null distribution generated from random 153
genomic coordinates. This allowed us to test whether shifts between early (PreSpec) and 154
late (PostSpec) rediploidisation topologies along a chromosome, which we take to mark the 155
boundaries of rediploidising units, occur significantly closer to rearrangement boundaries 156
than expected by chance (Fig 2 (a)). We found that our rearrangement blocks were in 157
closer proximity, both in base pair distance and gene distance, to topology breakpoints 158
than to randomly chosen locations in the genomes of both fish (Fig 2 (b) (ii)and(iii), 159
Supplementary Materials (22).) 160

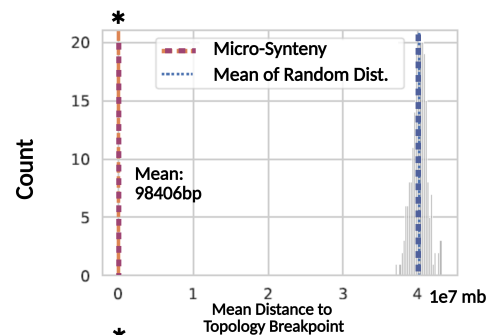
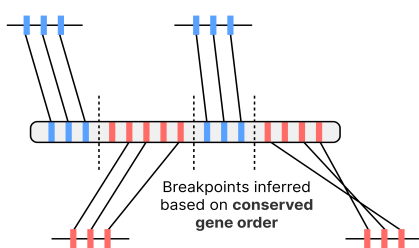
Finally, in addition to rearrangement block associations we also assessed whether TAD 161
boundaries were associated with topology breakpoints (Fig 2 (b) (iii); Supplementary 162
Materials; Table 1 (22); Supplementary Table S5). Using the logic above, we assessed 163
whether TAD boundaries were associated with topology breakpoints. We find that TAD 164
boundaries are indeed spatially associated with topology breakpoints and the observed 165

(a) Topology Breakpoints

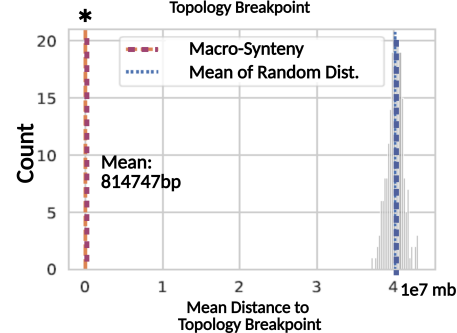
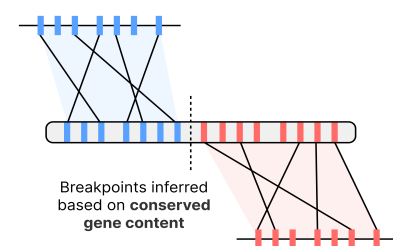


(b) Distance from topology breakpoint to:

(i) Micro-synteny breakpoints



(ii) Macro-synteny breakpoints



(iii) TAD-Boundaries

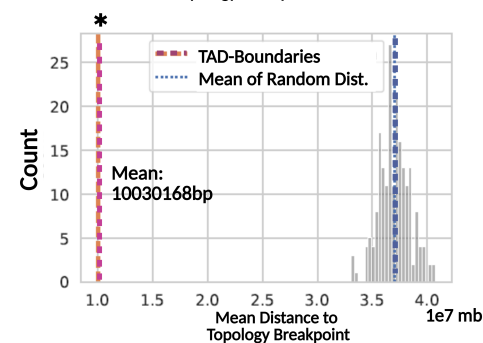
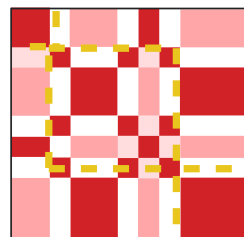


Figure 2. Rearrangements are associated rediploidisation in Acipenseriformes genomes (a) We compared the distance from topology breakpoints to: (b)(i) micro-synteny blocks, representing smaller and potentially more recent rearrangements; (ii) macro-synteny blocks, representing older and larger scale rearrangements and (iii) topologically associating domain (TAD) boundaries. Distances from these genomic features to topology breakpoints were compared with distances from randomly sampled genomic locations (only paddlefish shown). Statistical significance was assessed using a one-sided Mann-Whitney U test, and the observed mean distance was evaluated against a null distribution generated from random simulations. Asterisk indicates significant differences in mean distance.

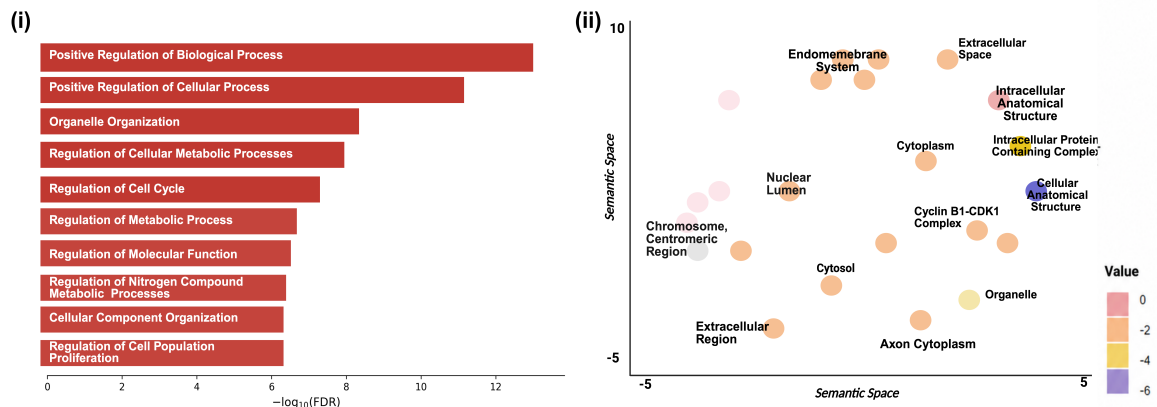
proximity of the two therefore suggests that entire chromatin domains, rather than isolated individual genes, tended to rediploidise together, possibly as functional units.

Functional organisation of early- and late-rediploidising regions

Given this enhanced information on the origin of these rediploidising units, we examined their chromosomal distribution in greater detail. We classified blocks as PostSpec-type or PreSpec-type topology blocks when they contained a significant enrichment of PostSpec or PreSpec genes relative to the genomic background (one-sided Fisher's exact test, $p < 0.05$; Supplementary Materials; Table 2 S2). The majority of PostSpec-type blocks were located on Paddlefish chromosomes 8, 7, 16, 18 and 22, and Sterlet chromosomes 20, 16, 12, 15, and 19. By contrast, PreSpec-type blocks were predominantly found on the 6 largest chromosomes in both species consistent with findings from Redmond et al. (2023) (Supplementary Materials; Table 2 (22)).

We next examined the biological functions enriched within the genes contained in these blocks by performing functional enrichment analyses using STRING on the PreSpec and PostSpec ohnolog sets (78). PreSpec gene sets showed strong and consistent enrichment for pathways associated with chromosome segregation, cell-cycle control, centrosome maturation, homologous recombination, and G2/M checkpoint regulation, precisely the types of functions which may be advantageous in stabilising meiosis, mitosis, and cell biology in the immediate aftermath of WGD (all enriched terms from STRING can be found in the Supplementary Materials (22)). Notably, pathways such as AURKA activation by TPX2, which is a critical allosteric mechanism that drives mitotic spindle assembly, centrosome maturation, and chromosome segregation was one of the most significantly enriched. Progesterone mediated oocyte maturation and the MAPK (Mitogen-Activated Protein Kinase) pathway were also enriched. The MAPK pathway is a crucial cell signaling network that relays external signals (like growth factors, stress) from the cell surface to the nucleus, controlling fundamental processes such as cell proliferation, differentiation, apoptosis, and stress response. In contrast, PostSpec gene sets showed no clear pattern

(a) PreSpec (BP)



(b) PostSpec (BP)

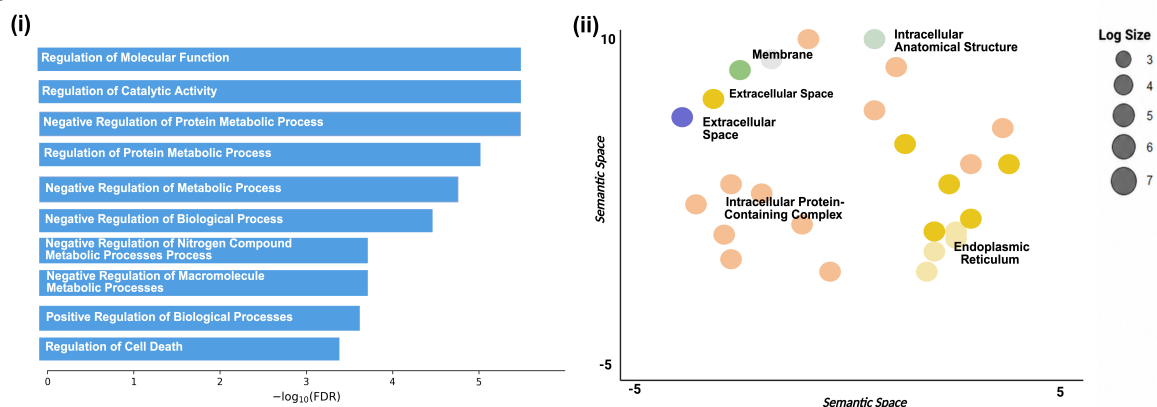


Figure 3. (a) (i) Top ten enriched Gene Ontology (GO) terms for PreSpec ohnologs. Biological process terms are shown as a bar plot, and (ii) cellular component terms are shown as a bubble plot (b) (i) Top ten enriched GO terms for PostSpec ohnologs in paddlefish. (ii) Biological process terms are shown as a bar plot, and cellular component terms are shown as a bubble plot. Semantic similarity was calculated using REVIGO, and statistical significance was assessed using two-sided permutation tests ($p < 0.01$).

of enrichment: PPI networks were not significantly enriched and functional terms were varied and showed no clear consistency (Fig3(c) and (d)).

Uncovering a missing duplicate *HoxC* cluster in the paddlefish assembly

The sterlet sturgeon genome has retained a particularly complete set of duplicate *Hox* clusters following the shared acipenseriform WGD. Paddlefish, while also having high retention of duplicate *Hox* clusters, have, according to two previous genome sequencing studies, only a single *HoxC* cluster, proposing that the other was secondarily lost after speciation (6, 12). These studies also inferred different total numbers of *Hox* genes in paddlefish, respectively 65 (6) and 75 (12), and as such we reassessed the *Hox* cluster and gene repertoire in our new duplicate resolved paddlefish assembly. We found a total of 77 *Hox* genes (Fig. 4 (c-d); Supplementary Materials (22)). Strikingly, we identified the missing *HoxC* cluster on chromosome 47 in our assembly which retained only one gene, *HoxC4*, suggesting that this cluster is highly degenerate. Notably, *HoxC4* is absent from the more intact *HoxC* cluster on chromosome 44 which is consistent with the previous chromosome-level assembly (12). Interestingly, *HoxC4* was identified in a previous scaffold-level assembly (6), but placed within the non-degenerate cluster, meaning that as well as being at odds with our inferred *Hox* complement, the previous studies are at odds with each other (6). Furthermore, the *HoxC4* sequences identified here and in the scaffold-level assembly (6) are identical, indicating that there are likely assembly and/or annotation errors in at least one of the three genomes.

To rule out the possibility that the lone, degenerate-cluster-forming *HoxC4* gene on chromosome 47 of our new assembly was an artefact of misassembly or otherwise we first searched for genes commonly found to flank *HoxC* clusters; *tspan31*, *ddx23*, *rnd1*, *adcy6*, *smug1*, and *copz1* upstream and downstream of the intact *HoxC* cluster and the isolated *HoxC4* gene (91). None of these genes were found adjacent to the intact *HoxC* cluster but all expected flanking genes were found upstream or downstream of the single *HoxC4* on

chromosome 47, and nowhere else in the genome (Fig 4 (a(ii)), Supplementary Materials (22)).

To exclude the possibility of misassembly, we also examined the read depth along chromosomes 47 and 44, identifying no abnormal signals in the regions containing the Hox clusters. Additionally, *HoxC4* was found in the scaffold-level assembly (6), located on scaffold JAAWVQ010136386.1, but all other *HoxC* genes were found as pairs on different scaffolds (*HoxC5* & *HoxC8* & *HoxC9* on scaffold JAAWVQ010005640.1, *HoxC11* & *HoxC10* on scaffold JAAWVQ010044606.1, *HoxC12* & *HoxC13* on scaffold JAAWVQ010102576.1). Furthermore, flanking single genes *smug1* and *copz1* were located on the same scaffold as *HoxC4* in the scaffold assembly. As such, the scaffold-level assembly validates *HoxC4* as a lone remaining member of the the *HoxC* on chromosome 47 and does not support the inclusion of *HoxC4* within the more complete *HoxC* cluster. It appears likely that, lacking chromosomal context, the gene was mistakenly assumed to belong to the intact cluster. Importantly, sequence alignment shows that *HoxC4* retains a conserved homeodomain, the canonical two-exon structure, and an intact YPWM hexapeptide motif, suggesting that it remains functional despite its isolated genomic location (Supplementary Fig. S1 (B)).

Collectively, these findings provide strong evidence for the loss of *HoxC4* from the cluster on chromosome 44 and suggest that the copy located on chromosome 47 represents the sole remnant of the otherwise missing duplicate cluster. Notably, *HoxC4* was one of many genes recovered through the genome enhancement performed here (Supplementary Materials; table 3(22)), highlighted the value added by this approach. These findings suggest that, despite being singletons distributed across two distinct chromosomes in paddlefish, both the full complement of *HoxC* and flanking genes may still function as they do in genomes where they remain clustered, a phenomenon that, to our knowledge, has not previously been described in vertebrate *Hox* cluster evolution.

Rediploidisation of duplicate *Hox* clusters

246

Hox clusters have been key in validating several WGD events including in the acipenseriformes (15, 77), in teleosts, as well as in ancient WGD events at the base of vertebrates (16, 25, 37). We reconstructed gene trees for each *Hox* ohnolog pair gene family but found that clear signal for their rediploidisation histories was difficult to resolve (Supplementary Materials (22)). *Hox* genes are short, highly conserved sequences that can often yield weak phylogenetic signal (80). To combat this, the concatenation method has been useful in resolving phylogenetic problems where weak signal obscures underlying evolutionary relationships, including to resolve the rediploidisation of salmonid *Hox* clusters (71). As *Hox* clusters are not interrupted by rearrangement events, and based on our above findings regarding the link with rediploidisation, we infer that the clusters rediploidised as complete units and can safely and validly be analysed collectively using concatenation. By concatenating individual *Hox* genes along a cluster from species that span the vertebrate phylogeny and inferring phylogenetic trees we recover PostSpec topologies for *Hox* cluster pairs A and D, indicating that they rediploidised independently in sturgeon and paddlefish, and a PreSpec topology for *HoxB* clusters, indicating that they underwent shared rediploidisation prior to the divergence of sturgeon and paddlefish (Supplementary Fig S1). Rediploidisation timing of the concatenated *HoxC* duplicate clusters was less straightforward, as only one clear, intact cluster is present in paddlefish, weakening the robustness of the timing estimate for this cluster. Nonetheless, we assessed the topology and found that the sterlet duplicates grouped together, consistent with a shared, independent duplication (rediploidisation), with the single paddlefish cluster sister to the sterlet pair.

To augment our results, we used the Approximately Unbiased (AU) test to assess for a statistically significant in the likelihood of the PreSpec and PostSpec topologies for each *Hox* cluster (72) (Supplementary Fig S1 (A) and Supplementary Materials; Table 4 (22)). Since AU tests can be applied to fixed topologies, they were also performed on individual *Hox* gene alignments, testing whether each alternative topology could be rejected. This

supported the concatenation analysis: for *HoxA*, both PreSpec topologies were rejected in favour of PostSpec in 7/11 gene trees ($p < 0.05$), while the remaining 4/11 failed to reject any topology (inconclusive). For *HoxD*, both PreSpec topologies were rejected in 8/10 gene trees. *HoxB* showed the opposite pattern: the PostSpec topology was rejected in favour of PreSpec in 4/11 gene trees ($p < 0.05$), with the remaining 5/11 inconclusive (Supplementary Materials (22)). These findings indicate that *Hox* clusters rediploidised at markedly different time points, with *HoxB* rediploidising in the common ancestor of sterlet and paddlefish, whereas *HoxA*, *HoxD* and *HoxC* rediploidised independently in sterlet and paddlefish.

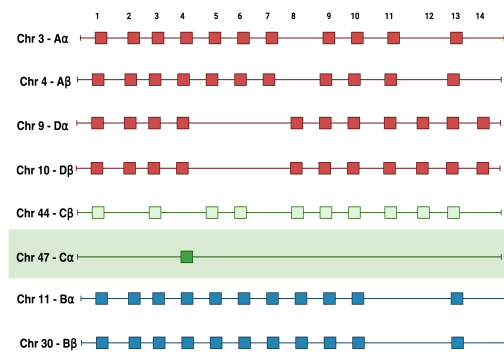
Phylogenomic dating of the *Hox* cluster rediploidisation

Redmond et al (2023), estimated the timing of the shared sturgeon–paddlefish WGD using ohnologs that rediploidised prior to species divergence, providing mean lower-bound estimates for the WGD event at approximately ~ 254.7 Ma and ~ 241.8 Ma (with joint 95% confidence intervals spanning 202.9–289 Ma), depending on which fossil calibrations were applied. They also estimated the speciation event splitting sturgeons and paddlefish at ~ 171.6 Ma or ~ 167.5 Ma depending on calibration strategy, with joint 95% confidence intervals spanning 123.4–203.3 Ma, suggesting a Jurassic or perhaps early Cretaceous divergence consistent with many other studies (70). This estimate is substantially different to previous estimates which lacked the asynchronous rediploidisation context (12, 15). Here we estimate divergence times of the concatenated *Hox* gene datasets, expecting a long-lag time post-WGD (given the time between the WGD and sturgeon-paddlefish lineage divergence) and anticipating the possibility of vastly different times in each species for the lineage-specific rediploidisation of the *HoxA* and *HoxD* clusters.

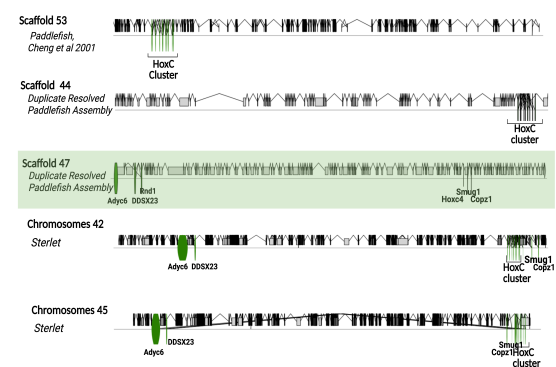
For *HoxA*, which rediploidised after speciation, we infer the timing of independent rediploidisation events: in paddlefish at ~ 98.74 ma (95% CI: 137.47–60.01 Ma), and in sturgeon at ~ 143.01 ma (95% CI: 182.93–103.10 Ma), suggesting these events may have occurred at very different times, although the confidence intervals overlap. *HoxD*

(a) Putative remnants of a second degenerate *HoxC* cluster

i. Duplicate Resolved Paddlefish assembly

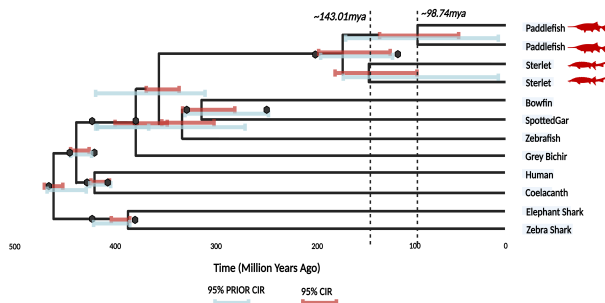


ii. Flanking genes and conserved synteny of the *HoxC* cluster

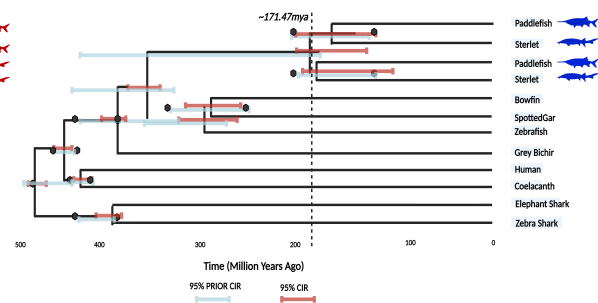


(b)

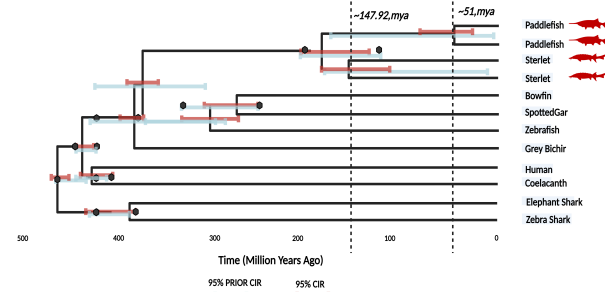
i. *HoxA* PostSpec



ii. *HoxB* PreSpec



iii. *HoxD* PostSpec



iv.

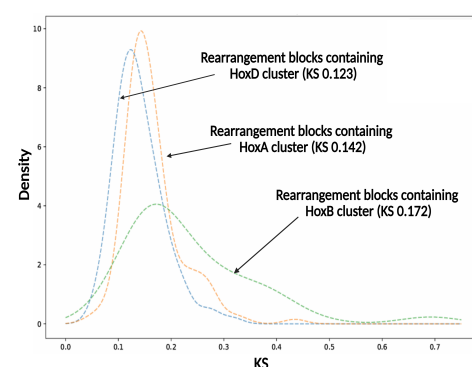


Figure 4. Hox cluster analysis in the Acipenseriformes. (a) (i) Hox genes in the duplicate-resolved paddlefish assembly, with remnants of a second *HoxC* cluster highlighted in green. (ii) *HoxC* cluster and flanking genes in the published paddlefish assembly, the improved assembly, and duplicated *HoxC* clusters in sterlet. The putative *HoxC* remnant in paddlefish, along with its flanking genes, is highlighted in green. (b) Dated phylogenies for (i) *HoxA*, (ii) *HoxB*, and (iii) *HoxD*. The 95% credibility interval (CIR) for each node is shown in red. Corresponding 95% CIRs from prior-only analyses are shown in blue below each node, demonstrating that WGD divergence time priors are sufficiently diffuse. Fossil calibrations (upper and lower bounds) are indicated by black diamonds. (iv) K_s distributions for rearrangement blocks containing duplicated Hox clusters; *HoxA*, *HoxB*, and *HoxD* clusters in Paddlefish.

also follows a post-speciation rediploidisation scenario, but with a clear and striking temporal difference: in paddlefish at ~ 51 Ma (95% CI: 80.27–23.14 Ma), and in sturgeon at ~ 147.92 Ma (95% CI: 182.84–113.00 Ma). *HoxB*, on the other hand, rediploidised in the ancestor of sturgeon and paddlefish, with an inferred rediploidisation time of ~ 171.57 Ma (95% CI: 145.59–212.51 Ma), perhaps coming only very shortly before their lineage divergence (Fig 4). These findings present a clear picture of asynchronous rediploidisation occurring at *Hox* clusters, with stark temporal differences that may reflect the distinct biological and morphological differences observed between the two lineages.

To connect the rearrangement blocks defined in the previous section to the *Hox* cluster dating, we assessed whether these blocks containing the clusters could themselves be “dated” using K_s , and whether the signal was consistent with the more precisely resolved timing of the enclosed *Hox* cluster. In Fig. 4 (b)(iv) and Supplementary Fig. S1 (c), the blocks containing *HoxA* and *HoxD* show K_s peaks close to zero, consistent with more recent rediploidisation and correspondingly fewer subsequent substitutions. The block containing *HoxB*, by contrast, shows a broader peak shifted further from zero, consistent with an older rediploidisation event. This is broadly consistent with the K_s analysis reported in previous studies and supports our conclusion that blocks of genes rediploidise together via structural rearrangements (70).

Discussion

Rearrangement events are often presumed as the mechanism by which rediploidisation occurs following an autopolyploid WGD event, yet only a handful of studies have detected patterns that support this. Here, we have directly taken advantage of the signal coming from asynchronous rediploidisation in slowly-evolving sturgeon and paddlefish genomes, i.e. syntenic clustering of genes with shared rediploidisation history (duplication times), to test this. We identify significant associations between genome rearrangement blocks and gene tree topology duplication time blocks, providing strong evidence that rearrangements are key mechanistic drivers of rediploidisation.

We present an updated and enhanced genome assembly for the American paddlefish, adding 9,859 genes to the genome. Remarkably, despite the WGD event being estimated to have occurred a quarter of a billion years ago, large portions of this genome remain highly similar and were consequently prone to assembly collapse. Here, we present a method for improving the assembly, together with our enhanced annotation. While sequencing techniques have advanced and long-read sequencing is much more commonplace, we highlight the implications for future studies, cautioning that similar assembly errors are likely to affect not only recent polyploids but also more ancient events as exemplified here (79). Epitomizing the value of our improved assembly and annotation we recovered two HoxC cluster regions in paddlefish for the first time after multiple previous studies indicated that one copy had been lost. The newly identified cluster is degenerate and consists of only a HoxC4 gene surrounded by known HoxC flanking genes. Interestingly, our analysis of the Hox cluster rediploidisation patterns supports independent, lineage-specific rediploidisation of the HoxC clusters in sturgeon and paddlefish. This implies that the Hox cluster duplicated independently in each lineage, and consequently the functional trajectories of each cluster pair are expected to be entirely lineage specific. It is therefore important to emphasize that the genetic and functional state of the sturgeon cluster cannot be interpreted as the ancestral condition of the more degenerate paddlefish cluster pair.

Understanding how different ploidy states can persist in a stratified manner along chromosomes over tens of millions of years requires a deeper understanding of meiotic processes following WGD. One potential additional factor is higher-order genome organisation. The association we observe between chromosomal rearrangements and TAD boundaries suggests that the 3D genome architecture may influence where and how rediploidisation occurs, with genes located within the same TAD tending to rediploidise together. If TAD boundaries and rediploidisation timing are indeed linked, this would imply that disruption of TAD structure is selectively disadvantageous. In this framework, TADs may act as both structural and functional constraints on the rediploidisation process

by restricting the range of viable chromosomal rearrangements. Rearrangements that
 disrupt established regulatory domains could be deleterious, whereas those that preserve or
 accommodate existing TAD architecture would be more likely to avoid purifying selection.
 Under such a model, the maintenance of regulatory integrity and its complexity might
 not only be expected to constrain the space and patterns of rediploidisation possible but
 to explain, at least in part, the asynchronous nature of rediploidisation observed here
 and in other lineages (50, 65, 71, 90). Alternative explanations are also possible however.
 Rather than constraining rediploidisation, TAD architecture may itself be shaped by the
 rediploidisation process, or it may even be a complex blend of both, across genomic regions.
 Distinguishing between these possibilities will require further investigation. Nevertheless,
 our findings suggest that 3D genome organisation may represent an important, and largely
 unexplored, component of the evolutionary dynamics of post-WGD genomes.

Our analyses show that early-rediploidising regions are enriched for pathways involved
 meiotic regulation, cell-cycle control, chromosome segregation, DNA repair, genomic
 stability, and stress resistance. By contrast, lineage-specific rediploidising regions show
 little evidence of functional enrichment, suggesting that they are functionally more
 heterogeneous. This finding ameliorates findings in salmonids and *Arabidopsis arenosa*
 in which ohnologs associated with cell-cycle regulation and genome maintenance exhibit
 signatures of adaptive evolution and selection following WGD (27, 34, 55). Although these
 studies demonstrate that such WGD-derived genes can be targets of adaptive evolution,
 our results suggest a more specific pattern: genes associated with cell-cycle regulation and
 genome maintenance are disproportionately represented among early-rediploidising regions.
 This enrichment is consistent with a model in which the earliest phase of rediploidisation
 contributed to the stabilisation of cellular functions required for chromosome pairing,
 segregation fidelity, and genome integrity following WGD. More broadly, our findings
 raise the possibility that particular functional classes of genes are preferentially associated
 with early rediploidisation.

Hox clusters provide clear examples of genomic regions that rediploidised at different

times with cases found both before and after the paddlefish-sturgeon divergence. Our phylogenetic analyses, together with divergence time estimates, strongly support a highly asynchronous pattern of rediploidisation among the clusters. As discussed in (70), divergence estimates that fail to account for asynchronous rediploidisation are likely biased toward more recent dates because they instead capture the timing of average, or most recent ohnolog divergence rather than the WGD event itself. Previous attempts to date the paddlefish WGD event were probably affected by this, such as an inferred date for an independent WGD event in the paddlefish lineage at approximately ~ 42 Ma based on sequences from the *HoxA* and *HoxD* clusters (15). Our results show that this estimate instead closely corresponds to the timing of rediploidisation within the paddlefish *HoxD* cluster. Correctly accounting for these processes is therefore essential for reconstructing the history and timing of WGD (13, 53, 68, 70).

Hox genes are central regulators of body patterning, making their independent rediploidisation particularly noteworthy in the context of lineage-specific morphological evolution (32). We estimate that the *HoxD* cluster in paddlefish rediploidised relatively recently, at approximately ~ 50 Ma as a mean estimate, whereas the corresponding region in sterlet rediploidised much earlier, around ~ 147 Ma. This near 100-year mean difference in duplication time of such an iconic gene cluster provides a stark demonstration of the lineage-specific trajectories rediploidisation can take if it has not completed before speciation (28). Furthermore, assuming rediploidisation was underway at ~ 254 Ma (70), this implies that some regions, including this *HoxD* cluster pair, of the paddlefish genome may have remained effectively tetraploid, with duplicate genes not resolving from tetrasomic singletons, for over 200 million years (as a mean estimate). *HoxA*, *HoxD* and *HoxC* all appear to show rediploidisation post-speciation and thus provide a striking example of extreme asynchronous rediploidisation, with homologous genomic regions resolving to diploid inheritance only after remarkable delay and on vastly different evolutionary timescales in the two lineages. Such differences in gene duplication status may have contributed to the distinct morphologies observed in paddlefish and sturgeons given the

role of *Hox* genes in development (32, 33). For example, paddlefish is characterised by a dramatically elongated rostrum rich in electroreceptors, an adaptation for detecting planktonic prey, while sturgeons have a much shorter snout with sensory barbels and are adapted for benthic suction feeding. Whether delayed rediploidisation of *Hox* clusters contributed to these and other morphological differences between the lineages is of course unclear, but warrants further functional exploration. Interestingly, the estimated timing of rediploidisation for the *HoxB* cluster falls close to the inferred paddlefish–sturgeon divergence time (70), which places the split at approximately ~ 171.6 Ma or ~ 167.5 Ma. This suggests that the *HoxB* cluster likely rediploidised shortly before speciation. Different rediploidisation timings of *Hox* clusters is not unique to Acipenseriformes and has also been documented for *Hox* duplicates in salmonids and other teleosts (56, 71).

Our findings evidence genomic rearrangements as the mechanism of rediploidisation, and further support it as an asynchronous and prolonged evolutionary process that generates ohnologs, sometimes long after WGD. Conflating rediploidisation with the WGD event itself, or using rediploidisation (gene divergence and its timing) as a proxy for WGD, is potentially highly problematic, particularly in autopolyploids (5, 50, 65, 70, 71). Reassessing ancient WGDs with consideration for the mode and tempo of rediploidisation may therefore reshape our interpretation of polyploidisations, including for key events such as those in early vertebrate evolution.

Materials and Methods

Resolving collapsed duplicates in the paddlefish genome

Cheng et al. (2021)’s American Paddlefish assembly has 60 pairs of chromosomes with a genome size of 1.54GB with 26,017 predicted protein-encoding genes (12). The assembly was sequenced to 30X coverage and short and long reads from this study were downloaded from CNGB under project accession number CNP0000867. Cheng et al. (12) note a smaller than expected genome size following a 17-mer analysis (51). Following a similar

method to Du et al.(17), we identified these collapsed regions and attempted to resolve
them (39, 42, 93).

PacBio long-reads were aligned with bwa (v0.7.17-r1198-dirty) (48) with default
parameters. This was sorted and indexed using SAMtools (v1.16.1) (8). The short reads
were aligned with Bowtie (v2.4.2) (45) using standard parameters and again, indexing and
sorting was done with SAMtools (v1.16.1) (8). To assess the depth of coverage across the
genome, the alignments were split into 10kb regions and mosdepth (v0.3.3) (66) was used
to quantify read depth at each segment with parameter *-by*. 10kb regions with double
the expected read depth (double coverage segments), were separated from the rest of
the genome for the next steps. Using FreeBayes (v1.3.6) (26), a polymorphism VCF was
generated from the short-read alignments. After this, long-reads were used to separate
“haplotypes” in the double coverage regions. HapCUT2 (20), a haplotype assembly tool, was
used to reconstruct individual haplotypes from double-coverage long-read alignments. This
approach enables the identification of genomic regions containing more than one haplotype
and was thus used here to detect potential duplicated loci represented mistakenly as
multiple alternative alleles in the double-coverage long-read BAM file. HapCUT2 requires
as input a BAM file of mapped reads and a corresponding VCF file (20). The software
is designed for diploid genomes and does not currently support phasing of polyploids.
Parts of the paddlefish genome however, exhibited signatures of apparent polyploidy.
To overcome the limitations of HapCUT2 we filtered the VCF to remove polyploid or
erroneous genotype calls (e.g., 4/4, 3/4) prior to analysis. Polyploid genotype fields (e.g.,
4/4) were coerced to diploid genotypes (2/2) using a custom script. While this does
not fully capture the underlying ploidy, it preserves more information than discarding
such variants entirely by retaining the two most common alternative alleles. HapCUT2
results for our double-coverage regions contained multiple assembled haplotypic segments
for a given loci (collapsed duplicates) which were subsequently split into individual files
and annotated with corresponding VCF information using a custom script by Du et al.

(17) (available at <https://github.com/dukecomeback/sterletM.sch>) that was modified for this study. These split files were then processed with fgbio's *HapCutToVcf* script to generate separate VCF's for each assembled haplotype. These VCF files and the reference were used to produce haplotypic contigs in fasta format using *vcf-consensus* from the bcftools package (v1.10.2) (47). The fasta files of the split regions were merged with the original contigs (minus the double coverage contigs) using the Unix *cat* command (See Supplementary Figure 1).

Assembly and scaffolding of the paddlefish genome

Following haplotype splitting, the genome was then reassembled and scaffolded with the available HiC data. Cheng et al. (12) described 60 pairs of chromosomes (n=120), finding 26 macro chromosomes and 34 smaller, micro-chromosomes, a number which aligns with previous karyotype studies (77) and is equivalent to the sterlet genome (17). Assembly and scaffolding were done using Juicer (v1.6) and 3D-DNA (v190716)(18, 19). The Illumina short-reads were aligned to the duplicate resolved contigs with Juicer (v1.6)(19). 3D-DNA (v190716) (18) was then used for assembling the genome with -r=0 flag to ensure no iterative rounds of mis-join correction were carried out. Finally, the scaffold assembly was manually reviewed using Juicebox assembly tools (v1.6)(18).

Paddlefish RNA-seq and genome annotation

To produce high-quality gene predictions we first produced RNA-seq data for fifteen tissues from four paddlefish individuals. Captive-bred American paddlefish were purchased from Osage Catfisheries Inc. (Osage Beach, MO, USA) and euthanized on site by MS-222 overdose, with samples collected from brain, caecum, eye, gill filament, gill raker, gonad, heart, kidney, liver, rostrum, skeletal muscle, skin, spiral valve, spleen, and stomach from four individual fish. Samples were flash frozen using liquid nitrogen and stored in RNeasy lysis buffer at -80°C. Total RNA was prepared by lysing tissue samples in 1ml Tri Reagent (Sigma Aldrich) using a Tissue lyser LT (Qiagen) and 7mm steel bead. Following

incubation for 5 minutes at room temperature, 0.2ml of chloroform was added and the tube
shaken for 15 seconds. After incubation for 2 to 3 minutes at room temperature the tubes
were spun at 12,000g for 15 minutes. The upper, RNA-containing aqueous fraction was
moved to a clean 1.5ml tube, mixed with an equal volume of 70% ethanol, and vortexed.
This solution was added to a spin column from the RNeasy mini kit (Qiagen) and RNA
prepared following the manufacturer's instructions, including the DNaseI treatment step
to remove any contaminating genomic DNA. Purified RNA was quantified using a Qubit
3.0 Fluorometer and then stored at -80°C. Quality control, poly-A enrichment mRNA
library preparation and RNA sequencing of 30 million 150bp paired end Illumina reads
were performed by Novogene America (Sacramento, CA, USA). Summary of paired-end
RNA-seq FASTQ files generated from *Polyodon spathula* (paddlefish) tissues are shown
in Table S4 and have been deposited in the NCBI Sequence Read Archive (SRA) under
BioProject accession PRJNA1512476.

Genome annotation was performed using BRAKER3 (30), integrating de novo gene
prediction with protein homology and RNA-seq evidence. Assembly completeness was
assessed using BUSCO (v5.4.4) (54) with the Actinopterygii odb9 dataset, run with the
-augustus and -long options enabled (54, 75). Protein homology evidence was provided
to BRAKER3 using the Metazoa protein database together with proteomes from the
following selected vertebrate genomes, used also in orthology inference (12): American
paddlefish, elephant shark, zebrafish, medaka, fugu, stickleback, sea lamprey, spotted gar,
human, mouse, and sterlet. Protein sequences were clustered using CD-HIT (24) and
supplied to BRAKER3 using the --prot_seq option. Raw read quality was assessed using
FastQC (v0.11.9) (2) . Reads were trimmed to remove adapters and low-quality bases
using Trimmomatic (v0.39) (7). RNA-seq reads from multiple adult paddlefish tissues were
aligned to the genome using HISAT2 (v2.2.1) (40). RNA-seq reads from multiple adult
paddlefish tissues were aligned to the genome using HISAT2 (v2.2.1) (40). Alignments
were sorted and indexed with SAMtools (v1.16.1) (8) and provided to BRAKER3 via
the --bam option. BRAKER3 was run in combined RNA-seq and protein mode using

GeneMark-ETP and AUGUSTUS for gene prediction, enabling automated training and evidence integration without manual model reconciliation. Functional annotation of predicted gene models was performed using the FANTASIA pipeline (57).

Orthology Assignment

Orthofinder (v2.5.4) (21) was used for orthology inference. We included a diverse set of proteomes that spanned the jawed vertebrate phylogeny including from the Chondrichthyes: ghost shark (*Callorhinchus milii*; GCF_000165045.1) (85); from the Sarcopterygii; human (*Homo sapiens*; GCF_000001405.39), African clawed frog (*Xenopus tropicalis*; GCF_000004195.4), and coelacanth (*Latimeria chalumnae*; GCF_000225785.1); from within Actinopterygii we selected zebrafish (*Danio rerio*; GCF_000002035.6), spotted gar (*Lepisosteus oculatus*; GCF_000242695.1) (9), Grey bichir (*Polypterus senegalus*; GCF_016835505.1) (6). We included a species tree in our Orthofinder (21) run with flag `-s` in line with the accepted relationships in the jawed vertebrate phylogeny to augment orthology inference: `((Callorhinchus milii), ((Latimeria chalumnae, (Xenopus tropicalis, (Homo sapiens)), (Grey birchir, ((Polyodon Spathula, Acipenser ruthenus), ((Danio rerio), (Lepisosteus oculatus)))))))).`

We extracted the longest isoform from each proteome using a custom pipeline, reduced sequence redundancy with CD-HIT (24), and then ran OrthoFinder (v2.5.4) (21) using the `-s` parameter to include a rooted species tree. To improve the resolution of orthogroup relationships, we additionally used the `-y` flag in OrthoFinder when generating Phylogenetic Hierarchical Orthogroups (PHOGs). This option separates paralogous clades below the root of a HOG into distinct hierarchical orthogroups, thereby reducing the lumping of multiple sturgeon–paddlefish ohnolog pairs into single gene families during downstream analyses. Following the protocol described in Redmond et al. (70), PHOGs were further filtered by extracting groups containing two sequences each from sturgeon and paddlefish, together with at least one outgroup sequence for subsequent rooting of

the sturgeon–paddlefish subtree. Sturgeon and paddlefish genes were only selected from
the 60 largest chromosomes.

Ohnolog duplication time inference

The sturgeon-paddlefish ohnolog pairs described above, were subjected to phylogenetic
analysis to estimate the time of rediploidisation relative to the speciation event. MAFFT
(v7.453) was used for multiple sequence alignment of the filtered PHOGs with standard
parameters. Phylogenetic inference by ML was performed with IQ-TREE (v1.6.12)(60)
with the -m JTT+G flag, -bb 1000 flag allowing 1000 ultrafast bootstrap (59) replicates.
These ohnolog gene trees were processed and analysed for duplication time inference (i.e
rediploidisation time). Scripts described and published in (70) were used to infer the
rediploidisation time of each gene tree. The resulting gene trees were summarised into
different groups indicating their duplication time: PreSpec, PostSpec, Other[PreSpec-like,
PostSpec-like].

Hox cluster analysis

We downloaded the complete Hox cluster sequences from the spotted gar (9), sterlet (17),
zebrashark (91) and elephant shark (85). TBLASTN (NCBI BLAST+ suite v2.17.0) was
used to align these genes against our genome assembly. Taking the longest captured
regions we then used Exonerate (v2.4.0)) to align the best-hits (73) and delineate and
verify the open-reading frames of 77 Hox genes across 8 chromosomes. We also took the
protein sequences of elephant shark and human *HoxC* flanking genes *Tspan31*, *Ddx23*,
Rnd1, *Adcy6*, *Smug1*, and *Copz1* to verify the location of duplicate *HoxC* clusters. To
investigate the phylogenetic relationships of the clusters, we used *Hox* orthologs from
Acipenser ruthenus (sterlet), *Polypterus senegalus* (gray bichir), *Amia calva* (bowfin),
Stegostoma tigrinum (zebrashark), *Callorhinchus milii* (elephant shark), *Danio rerio*
(zebrafish), *Lepisosteus oculatus* (spotted gar), *Homo sapiens* (human), and *Latimeria*

chalumnae (coelacanth), obtained from the NCBI database. Sequences were aligned using MASCEv2 (69) and codon alignments were generated using the amino acid alignments as a templates with the same software. These alignments were used to construct ML gene trees for each *Hox* gene in IQTREE (v.2.1.3). using ModelFinder, for model selection and the -bb 1000 flag allowing 1000 ultrafast bootstrap replicates(UFBOOT). For the concatenated alignments the best-fitting model for each concatenated gene family was again determined using IQ-TREE (v.2.1.3). To provide statistical delineate between these classifications, the Approximately Unbiased (AU) test was employed in IQ-TREE (v.2.1.3) to compare the likelihood of the two possible PresSpec against the single PostSpec topology. Gene families were assigned to the best-supported topology where alternative hypotheses could be rejected at $p < 0.05$.

Defining Micro- and Macro-Syntenic blocks and Topology Breakpoints

Micro-syntenic blocks between the sturgeon and paddlefish genomes were identified using OrthoFinder (v2.5.4) (21) and i-ADHoRe (v3.0.01) (67). Proteomes from sterlet, the duplicate-resolved paddlefish assembly, and gray bichir were analysed using OrthoFinder with default parameters. Conserved syntenic blocks were subsequently identified with i-ADHoRe, and genomic coordinates for genes within each block, as well as block boundaries (defined by the first and last gene pairs), were extracted from the `segments.txt` output file. Macro-syntenic blocks, representing larger-scale rearrangements, were constructed following the methodology described in (63) To identify genomic regions associated with transitions in ohnolog gene-tree topology, a Bayesian segmentation approach was applied to paddlefish and sturgeon chromosomes using the algorithm available at <https://github.com/McLysaght-Evolutionary-Genetics/bayesian-segmentation>. In this approach, each chromosome was modeled as an ordered sequence of categorical observations corresponding to posterior probabilities of the gene-tree topology states: PreSpec, PostSpec. A topology breakpoint was defined as the genomic position at which the most probable

topology state transitioned between categories. To assess spatial relationships between chromosomal rearrangements and changes in rediploisation history, genomic distances were calculated between each micro- and macro-synteny block boundary and the nearest topology breakpoints. These observed distances were compared against null distributions generated from randomly sampled genomic coordinates matched by chromosome and block size. All distance calculations and randomisations were performed using custom Python scripts (86);(Supplementary Materials (22)). PreSpec and PostSpec gene trees within each synteny block were quantified using custom scripts (Supplementary Materials (22)). To test whether specific gene categories (PreSpec or PostSpec) were overrepresented within individual synteny blocks relative to genome-wide background frequencies, Fisher's exact tests were performed. For each block and category, a 2×2 contingency table was constructed comparing the number of genes of the focal category within the block to the number outside the block. Statistical analyses were conducted using the *scipy* Python package, with significance evaluated at $p < 0.05$.

Functional investigation with STRING

Protein–protein interaction networks were analysed using STRING v12.0. (78) All predicted protein sequences were queried against the STRING database with the American Paddlefish proteome used as background. STRING's PPI enrichment test was used to evaluate whether observed interactions exceeded random expectation given the size of each protein set. A PPI enrichment p-value $p < 0.05$ was considered indicative of statistically significant interaction structure. Functional enrichment assessments were conducted using STRING's integrated annotation tools. For each gene set, enrichment analyses were performed across: Gene Ontology (Biological Process, Molecular Function, Cellular Component), KEGG pathways, Reactome pathways and STRING Local Network Clusters (LNCs). Significance was evaluated using the Benjamini–Hochberg false-discovery rate (FDR) method, with $FDR < 0.05$ considered significant. STRING's default species-wide proteome was used as the background set for all tests. Semantic similarity was calculated

using REVIGO, and statistical significance was assessed using two-sided permutation tests ($p < 0.01$).

TAD boundaries

Hi-C data were processed using the HiCExplorer v3.7.6 pipeline (89). Raw Hi-C reads were aligned to the reference genome using bamTools (3), and the resulting BAM files were sorted and filtered before matrix construction with hicBuildMatrix with the following parameters: `-binSize 10000 -restrictionSequence GATC (mbol) -danglingSequence GATC -minDistance 150`. The Hi-C matrix was normalised using Knight–Ruiz (KR) balancing with hicCorrectMatrix to correct for coverage biases. TADs were identified using hicFindTADs with FDR correction and the following parameters: `-minDepth 3000 -maxDepth 31500 -minBoundaryDistance 30000 -step 1500 -thresholdComparisons 0.005 -delta 0.01 -correctForMultipleTesting fdr`. These analyses were also attempted for the sterlet sturgeon, but at least with currently available data we were unable to robustly identify TADs and their boundaries (data not shown).

Phylogeneomic divergence dating of hox clusters

Phylogenomic divergence dating was conducted using PhyloBayes v4.1 (46), applying the site-heterogeneous CAT-GTR+G4 substitution model with an autocorrelated log-normal relaxed clock (36), and a birth-death prior with soft fossil calibration bounds (35, 92). Fossil calibration priors (Supplementary Materials (22)) were applied to most nodes in the tree, with the exception of the sturgeon–paddlefish WGD lower-bound node and the node separating Acipenseriformes (sturgeons and paddlefish) from Neopterygii. The calibration scheme largely followed that of (70), which is based on (4). Specifically, a minimum divergence of 121 Ma for sturgeons and paddlefish was used (76), and a lower bound of 381 Ma was set for crown Chondrichthyes (23). A fixed tree topology was specified based on established jawed vertebrate phylogeny (12, 17, 36, 70) and our inference of a shared WGD. The topology used for *HoxA* and *HoxD* was:

((zebra shark,elephant shark),((coelcanth,human),(grey bichir,((zebrafish,(spotted gar, bowfin)),((sterlet,sterlet1),(paddlefish,paddlefish1)))))); was specified based on accepted jawed vertebrate phylogeny (12, 17, 36) Zebra shark and whale shark (Chondrichthyes) were designated as outgroups. To confirm this topology, we conducted a basic concatenated phylogenomic analysis for each of our five datasets using IQ-TREE (60) under the LG+G4 model, with 1000 UFBoot bootstrap replicates. Each PhyloBayes MCMC chain was run for at least 30,000 cycles, discarding the first 5,000 as burn-in before calculating divergence times and 95% credibility intervals (CIRs). To verify the influence of priors, we also performed runs under the prior using the same settings but substituted the site-heterogeneous CAT-GTR+G4 model with a site-homogeneous Poisson model for computational efficiency. Because the prior over divergence times is independent of the substitution model, this approach allowed us to confirm that our WGD divergence priors were sufficiently diffuse and did not unduly constrain the main analysis.

K_S analysis of the *Hox* clusters

Synonymous substitution rates (K_S) values were calculated for *Hox* gene clusters in paddlefish and sterlet. In addition, we included paralogous gene pairs from within rearrangement blocks containing the *Hox* clusters in each species. We assume that these blocks consist of genes that experienced rediploidisation contemporaneously with the *Hox* clusters. K_S values were estimated for each rearrangement block containing a *Hox* cluster pair using the wgd v2 toolkit (11) for each species.

Acknowledgments

This work was supported by funding from the European Research Council, grant agreement 771419 (to A.McL.). A.K.R. was supported by an Irish Research Council Government of Ireland Postdoctoral Fellowship (GOIPD/2021/466), a Royal Society-Research Ireland University Research Fellowship (URF\R1\241884), and a UCD Ad Astra Fellowship.

D.J.M. was supported by the Biotechnology and Biological Sciences Research Council 678
through grants BBS/E/RL/230001B and BB/Z51746X/1 679

Supplementary Figures and Tables

680

Table S1. Categorisation of the possible rooted sturgeon-paddlefish subtrees with duplication nodes coming before (‘PreSpec’) or after (‘PostSpec’) the species diverged, and ‘Other’ trees that only partially match one of these scenarios (either, ‘PreSpec-like’, or ‘PostSpec-like’). Counts from Redmond et al 2023 and this study shown here

	PreSpec Trees	PostSpec Trees	Other Trees
Redmond et al. 2023	1448	2074	1917
This study	1713	2105	2044

Table S2. Counts of the different breakpoint types identified in the study, as well as counts for PreSpec-type and PostSpec-type blocks. Blocks were classified as PreSpec-type or PostSpec-type when they showed a significant enrichment of PreSpec or PostSpec genes relative to the genomic background (one-sided Fisher’s exact test, $p < 0.05$).

	Topology Breakpoints	Micro-Synten Blocks	Macro-Synten Blocks	PreSpec Type	PostSpec Type	TADs
Sterlet	293	1503	200	386	180	N/A
Paddlefish	383	1173	205	282	156	6094

Table S3. Assembly stats

	Paddlefish, Cheng et al. 2021	Paddlefish, Duplicate Resolved Assembly
Genome Size	1.54GB	1.56GB
Scaffold N50	48.9MB	48.9GB
GC content	39%	39.2%
Protein-Coding Genes	26,071	35,930
Chromosome No.	60 pairs (2n+120)	60 pairs (2n+120)
BUSCO	93%[S:51%,D:41%],F:2.5%,M:5.5%,n:3640]	93.14[S:36.5%,D:56.6%],F:1.3%,M:5.6%,n:3640]

Table S4. Summary of paired-end RNA-seq FASTQ files generated from *Polyodon spathula* (paddlefish) tissues deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession PRJNA1512476. The dataset comprises 120 FASTQ files (60 paired-end libraries) from four individuals (A–D) and 15 tissue types per individual, with a total compressed size of approximately 296 GB.

File name	File size	Read	Sample description
A1_1.fq.gz	2.2G	R1	Paddlefish Fish A brain
A1_2.fq.gz	2.3G	R2	Paddlefish Fish A brain

Continued on next page

Table S4 – continued from previous page

File name	File size	Read	Sample description
A2R_1.fq.gz	2.6G	R1	Paddlefish Fish A kidney
A2R_2.fq.gz	2.7G	R2	Paddlefish Fish A kidney
A3_1.fq.gz	2.5G	R1	Paddlefish Fish A liver
A3_2.fq.gz	2.6G	R2	Paddlefish Fish A liver
A4_1.fq.gz	2.4G	R1	Paddlefish Fish A spleen
A4_2.fq.gz	2.5G	R2	Paddlefish Fish A spleen
A5R_1.fq.gz	2.7G	R1	Paddlefish Fish A skin
A5R_2.fq.gz	2.8G	R2	Paddlefish Fish A skin
A6_1.fq.gz	2.5G	R1	Paddlefish Fish A skeletal muscle
A6_2.fq.gz	2.6G	R2	Paddlefish Fish A skeletal muscle
A7_1.fq.gz	2.3G	R1	Paddlefish Fish A eye
A7_2.fq.gz	2.3G	R2	Paddlefish Fish A eye
A8_1.fq.gz	3.6G	R1	Paddlefish Fish A gill filament
A8_2.fq.gz	3.7G	R2	Paddlefish Fish A gill filament
A9_1.fq.gz	2.6G	R1	Paddlefish Fish A gill raker
A9_2.fq.gz	2.7G	R2	Paddlefish Fish A gill raker
A10_1.fq.gz	2.1G	R1	Paddlefish Fish A rostrum
A10_2.fq.gz	2.2G	R2	Paddlefish Fish A rostrum
A11_1.fq.gz	2.5G	R1	Paddlefish Fish A spiral valve
A11_2.fq.gz	2.6G	R2	Paddlefish Fish A spiral valve
A12_1.fq.gz	2.8G	R1	Paddlefish Fish A stomach
A12_2.fq.gz	2.8G	R2	Paddlefish Fish A stomach
A13_1.fq.gz	2.3G	R1	Paddlefish Fish A heart
A13_2.fq.gz	2.4G	R2	Paddlefish Fish A heart
A14_1.fq.gz	2.4G	R1	Paddlefish Fish A gonad

Continued on next page

Table S4 – continued from previous page

File name	File size	Read	Sample description
A14_2.fq.gz	2.4G	R2	Paddlefish Fish A gonad
A15R_1.fq.gz	2.3G	R1	Paddlefish Fish A pyloric caeca
A15R_2.fq.gz	2.4G	R2	Paddlefish Fish A pyloric caeca
B1_1.fq.gz	2.5G	R1	Paddlefish Fish B, brain
B1_2.fq.gz	2.6G	R2	Paddlefish Fish B, brain
B2_1.fq.gz	2.3G	R1	Paddlefish Fish B, kidney
B2_2.fq.gz	2.3G	R2	Paddlefish Fish B, kidney
B3_1.fq.gz	2.4G	R1	Paddlefish Fish B, liver
B3_2.fq.gz	2.5G	R2	Paddlefish Fish B, liver
B4_1.fq.gz	2.5G	R1	Paddlefish Fish B, spleen
B4_2.fq.gz	2.5G	R2	Paddlefish Fish B, spleen
B5_1.fq.gz	2.2G	R1	Paddlefish Fish B, skin
B5_2.fq.gz	2.3G	R2	Paddlefish Fish B, skin
B6_1.fq.gz	2.3G	R1	Paddlefish Fish B, skeletal muscle
B6_2.fq.gz	2.4G	R2	Paddlefish Fish B, skeletal muscle
B7_1.fq.gz	2.9G	R1	Paddlefish Fish B, eye
B7_2.fq.gz	3.0G	R2	Paddlefish Fish B, eye
B8_1.fq.gz	2.3G	R1	Paddlefish Fish B, gill filament
B8_2.fq.gz	2.4G	R2	Paddlefish Fish B, gill filament
B9_1.fq.gz	2.3G	R1	Paddlefish Fish B, gill raker
B9_2.fq.gz	2.2G	R2	Paddlefish Fish B, gill raker
B10_1.fq.gz	2.5G	R1	Paddlefish Fish B, rostrum
B10_2.fq.gz	2.6G	R2	Paddlefish Fish B, rostrum
B11_1.fq.gz	2.2G	R1	Paddlefish Fish B, spiral valve
B11_2.fq.gz	2.3G	R2	Paddlefish Fish B, spiral valve

Continued on next page

Table S4 – continued from previous page

File name	File size	Read	Sample description
B12_1.fq.gz	2.4G	R1	Paddlefish Fish B, stomach
B12_2.fq.gz	2.5G	R2	Paddlefish Fish B, stomach
B13_1.fq.gz	2.7G	R1	Paddlefish Fish B, heart
B13_2.fq.gz	2.7G	R2	Paddlefish Fish B, heart
B14_1.fq.gz	2.9G	R1	Paddlefish Fish B, gonad
B14_2.fq.gz	3.0G	R2	Paddlefish Fish B, gonad
B15_1.fq.gz	2.8G	R1	Paddlefish Fish B, pyloric caeca
B15_2.fq.gz	2.9G	R2	Paddlefish Fish B, pyloric caeca
C1_1.fq.gz	2.3G	R1	Paddlefish Fish C, brain
C1_2.fq.gz	2.4G	R2	Paddlefish Fish C, brain
C2_1.fq.gz	2.5G	R1	Paddlefish Fish C, kidney
C2_2.fq.gz	2.6G	R2	Paddlefish Fish C, kidney
C3_1.fq.gz	2.4G	R1	Paddlefish Fish C, liver
C3_2.fq.gz	2.6G	R2	Paddlefish Fish C, liver
C4_1.fq.gz	2.1G	R1	Paddlefish Fish C, spleen
C4_2.fq.gz	2.2G	R2	Paddlefish Fish C, spleen
C5_1.fq.gz	2.1G	R1	Paddlefish Fish C, skin
C5_2.fq.gz	2.3G	R2	Paddlefish Fish C, skin
C6_1.fq.gz	2.1G	R1	Paddlefish Fish C, skeletal muscle
C6_2.fq.gz	2.3G	R2	Paddlefish Fish C, skeletal muscle
C7_1.fq.gz	2.5G	R1	Paddlefish Fish C, eye
C7_2.fq.gz	2.7G	R2	Paddlefish Fish C, eye
C8_1.fq.gz	2.3G	R1	Paddlefish Fish C, gill filament
C8_2.fq.gz	2.4G	R2	Paddlefish Fish C, gill filament
C9_1.fq.gz	2.4G	R1	Paddlefish Fish C, gill raker

Continued on next page

Table S4 – continued from previous page

File name	File size	Read	Sample description
C9_2.fq.gz	2.6G	R2	Paddlefish Fish C, gill raker
C10_1.fq.gz	2.2G	R1	Paddlefish Fish C, rostrum
C10_2.fq.gz	2.3G	R2	Paddlefish Fish C, rostrum
C11_1.fq.gz	2.4G	R1	Paddlefish Fish C, spiral valve
C11_2.fq.gz	2.5G	R2	Paddlefish Fish C, spiral valve
C12_1.fq.gz	2.4G	R1	Paddlefish Fish C, stomach
C12_2.fq.gz	2.5G	R2	Paddlefish Fish C, stomach
C13_1.fq.gz	2.3G	R1	Paddlefish Fish C, heart
C13_2.fq.gz	2.5G	R2	Paddlefish Fish C, heart
C14_1.fq.gz	2.3G	R1	Paddlefish Fish C, gonad
C14_2.fq.gz	2.5G	R2	Paddlefish Fish C, gonad
C15_1.fq.gz	2.4G	R1	Paddlefish Fish C, pyloric caeca
C15_2.fq.gz	2.5G	R2	Paddlefish Fish C, pyloric caeca
D1_1.fq.gz	3.1G	R1	Paddlefish Fish D, brain
D1_2.fq.gz	3.3G	R2	Paddlefish Fish D, brain
D2_1.fq.gz	2.2G	R1	Paddlefish Fish D, kidney
D2_2.fq.gz	2.3G	R2	Paddlefish Fish D, kidney
D3_1.fq.gz	2.3G	R1	Paddlefish Fish D, liver
D3_2.fq.gz	2.4G	R2	Paddlefish Fish D, liver
D4_1.fq.gz	3.0G	R1	Paddlefish Fish D, spleen
D4_2.fq.gz	3.1G	R2	Paddlefish Fish D, spleen
D5_1.fq.gz	2.2G	R1	Paddlefish Fish D, skin
D5_2.fq.gz	2.4G	R2	Paddlefish Fish D, skin
D6_1.fq.gz	2.3G	R1	Paddlefish Fish D, skeletal muscle
D6_2.fq.gz	2.5G	R2	Paddlefish Fish D, skeletal muscle

Continued on next page

Table S4 – continued from previous page

File name	File size	Read	Sample description
D7_1.fq.gz	2.4G	R1	Paddlefish Fish D, eye
D7_2.fq.gz	2.5G	R2	Paddlefish Fish D, eye
D8_1.fq.gz	2.3G	R1	Paddlefish Fish D, gill filament
D8_2.fq.gz	2.5G	R2	Paddlefish Fish D, gill filament
D9_1.fq.gz	2.7G	R1	Paddlefish Fish D, gill raker
D9_2.fq.gz	2.9G	R2	Paddlefish Fish D, gill raker
D10_1.fq.gz	2.5G	R1	Paddlefish Fish D, rostrum
D10_2.fq.gz	2.6G	R2	Paddlefish Fish D, rostrum
D11_1.fq.gz	2.2G	R1	Paddlefish Fish D, spiral valve
D11_2.fq.gz	2.4G	R2	Paddlefish Fish D, spiral valve
D12_1.fq.gz	2.3G	R1	Paddlefish Fish D, stomach
D12_2.fq.gz	2.5G	R2	Paddlefish Fish D, stomach
D13_1.fq.gz	2.4G	R1	Paddlefish Fish D, heart
D13_2.fq.gz	2.5G	R2	Paddlefish Fish D, heart
D14_1.fq.gz	3.0G	R1	Paddlefish Fish D, gonad
D14_2.fq.gz	3.2G	R2	Paddlefish Fish D, gonad
D15_1.fq.gz	2.6G	R1	Paddlefish Fish D, pyloric caeca
D15_2.fq.gz	2.8G	R2	Paddlefish Fish D, pyloric caeca

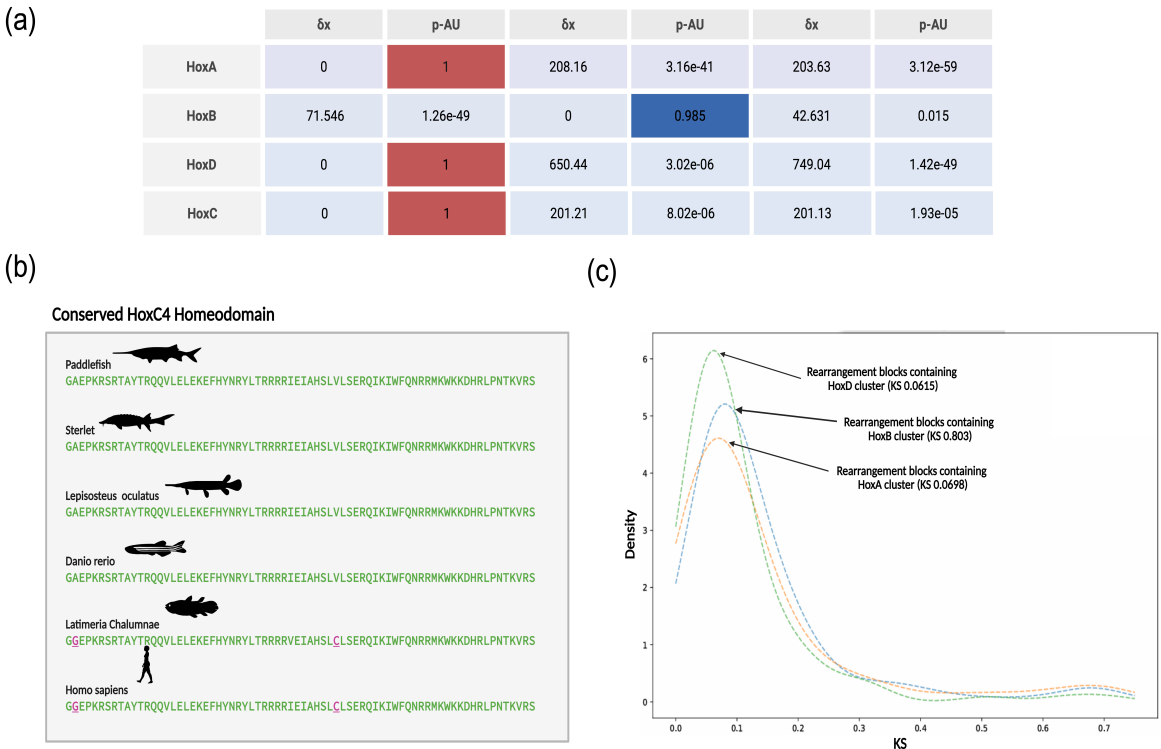


Figure S1. AU-test results for the concatenate Hox clusters in the sterlet and paddlefish (A) AU-test results for the concatenated gene trees (B) Alignment of itHoxC4 genes accross vertebrates showing clear conseravtion in the Paddlefish sequence (C) K_s distributions for rearrangement blocks containing duplicated Hox clusters; *HoxA*, *HoxB*, and *HoxD* clusters in Sterlet.

Metric	Value
Sequenced reads	298 M
Valid Hi-C contacts	161 M
Interchromosomal	68.8 M
Intrachromosomal ≥ 20 kb	19.7 M
Duplicates	31 k
Self-ligation/dangling ends	Low

Table S5. Hi-C QC Summary of Hi-C sequencing, mapping, and library quality statistics for the paddlefish Hi-C dataset used for genome-wide TAD identification.

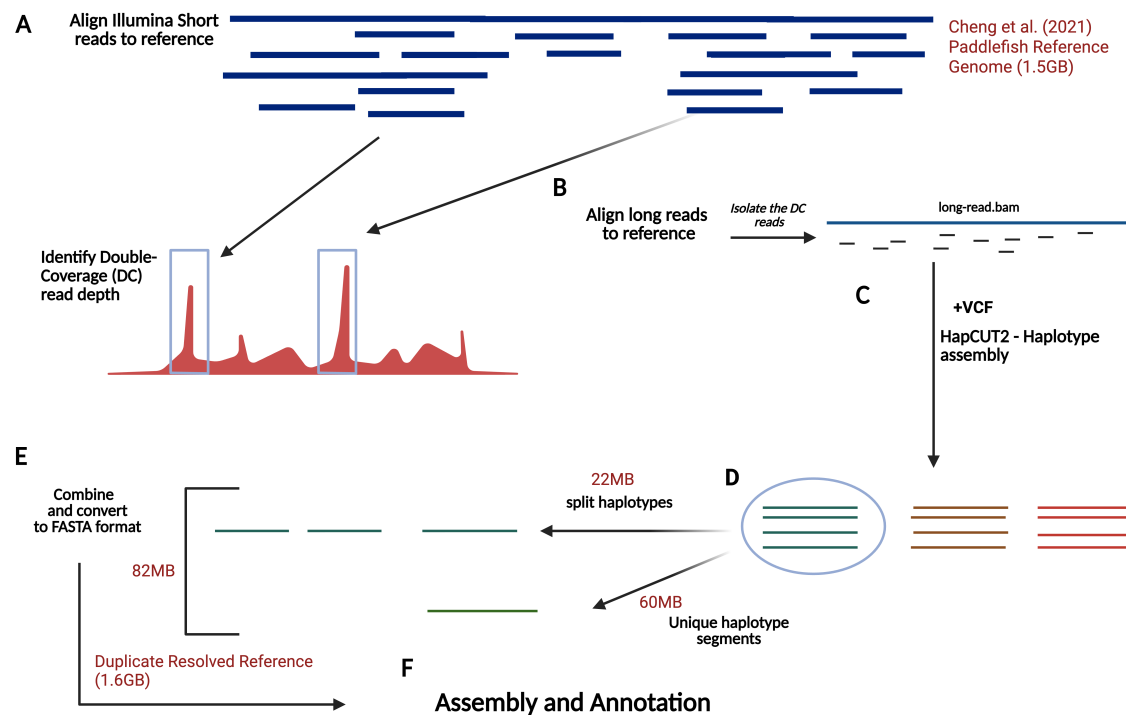


Figure S2. Schematic of the methodology used to resolve collapsed reads in the Cheng et al (2021), paddleFish Bassembly (a) Illumina short reads are aligned to the reference, and regions with double coverage are identified using Mosdepth (v0.3.3) (66). (b) Double coverage (DC) regions are isolated from the mapped long reads (c) In conjunction with the DC long-read BAM file, a VCF generated with freebayes (26) is utilized as input for HapCUT2 ((20)) to reconstruct individual haplotypes within the double-coverage mapped long reads (d) For each double-coverage region, a haplotyped VCF file is generated. Some of these files contain more than one haplotypic segment. These are our putative collapsed duplicates and so are split based on the haplotypic information from the VCF file (e) The split and unique regions are then integrated with the original genome fasta file (f) Subsequently, the duplicate resolved reference is scaffolded and annotated

References

- [1] Laurent Abi-Rached, André Gilles, Takashi Shiina, Pierre Pontarotti, and Hidetoshi Inoko. Evidence of en bloc duplication in vertebrate genomes. *Nature Genetics*, 31(1):100 – 105, 05 2002. doi: 10.1038/ng855.
- [2] Simon Andrews. Fastqc: a quality control tool for high throughput sequence data <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. *Babraham Bioinformatics*, 2010.
- [3] Derek W. Barnett, Erik K. Garrison, Aaron R. Quinlan, Michael P. Strömberg, and Gabor T. Marth. Bamtools: a c++ api and toolkit for analyzing and managing bam files. *Bioinformatics*, 27(12):1691–1692, 04 2011. ISSN 1367-4803. doi: 10.1093/bioinformatics/btr174. URL <https://doi.org/10.1093/bioinformatics/btr174>.
- [4] Michael J Benton, Philip CJ Donoghue, Robert J Asher, Matt Friedman, Thomas J Near, Jakob Vinther, et al. Constraints on the timescale of animal evolutionary history. *Oxford University Press*, 2015.
- [5] Camille Berthelot, Frédéric Brunet, Domitille Chalopin, Amélie Juanchich, Maria Bernard, Benjamin Noël, Pascal Bento, Corinne Da Silva, Karine Labadie, Adriana Alberti, Jean-Marc Aury, Alexandra Louis, Patrice Dehais, Philippe Bardou, Jérôme Montfort, Christophe Klopp, Cédric Cabau, Christine Gaspin, Gary H. Thorgaard, Mekki Boussaha, Edwige Quillet, René Guyomard, Delphine Galiana, Julien Bobe, Jean-Nicolas Volff, Carine Genêt, Patrick Wincker, Olivier Jaillon, Hugues Roest Crollius, and Yann Guiguen. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications*, 5(1):3657, 2014. doi: 10.1038/ncomms4657.
- [6] Xupeng Bi, Kun Wang, Liandong Yang, Hailin Pan, Haifeng Jiang, Qiwei Wei, Miaoquan Fang, Hao Yu, Chenglong Zhu, Yiran Cai, Yuming He, Xiaoni Gan, Honghui Zeng, Daqi Yu, Youan Zhu, Huifeng Jiang, Qiang Qiu, Huanming Yang,

- Yong E. Zhang, Wen Wang, Min Zhu, Shunping He, and Guojie Zhang. Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes. *Cell*, 184(5): 1377–1391.e14, 2021. ISSN 0092-8674. doi: 10.1016/j.cell.2021.01.046.
- [7] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 08 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu170. URL <https://doi.org/10.1093/bioinformatics/btu170>.
- [8] James K Bonfield, John Marshall, Petr Danecek, Heng Li, Valeriu Ohan, Andrew Whitwham, Thomas Keane, and Robert M Davies. HTSlib: C library for reading/writing high-throughput sequencing data. *GigaScience*, 10(2):giab007–, 2021. ISSN 2047-217X. doi: 10.1093/gigascience/giab007.
- [9] Ingo Braasch, Andrew R Gehrke, Jeramiah J Smith, Kazuhiko Kawasaki, Tereza Manousaki, Jeremy Pasquier, Angel Amores, Thomas Desvignes, Peter Batzel, Julian Catchen, Aaron M Berlin, Michael S Campbell, Daniel Barrell, Kyle J Martin, John F Mulley, Vydivanathan Ravi, Alison P Lee, Tetsuya Nakamura, Domitille Chalopin, Shaohua Fan, Dustin Weisel, Cristian Cañestro, Jason Sydes, Felix E G Beaudry, Yi Sun, Jana Hertel, Michael J Beam, Mario Fasold, Mikio Ishiyama, Jeremy Johnson, Steffi Kehr, Marcia Lara, John H Letaw, Gary W Litman, Ronda T Litman, Masato Mikami, Tatsuya Ota, Nil Ratan Saha, Louise Williams, Peter F Stadler, Han Wang, John S Taylor, Quenton Fontenot, Allyse Ferrara, Stephen M J Searle, Bronwen Aken, Mark Yandell, Igor Schneider, Jeffrey A Yoder, Jean-Nicolas Volff, Axel Meyer, Chris T Amemiya, Byrappa Venkatesh, Peter W H Holland, Yann Guiguen, Julien Bobe, Neil H Shubin, Federica Di Palma, Jessica Alföldi, Kerstin Lindblad-Toh, and John H Postlethwait. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nature Genetics*, 48(4):427–437, 2016. ISSN 1061-4036. doi: 10.1038/ng.3526.
- [10] Chase D. Brownstein, Daemin Kim, Oliver D. Orr, Gabriela M. Hogue, Bryn H.

- Tracy, M. Worth Pugh, Randal Singer, Chelsea Myles-McBurney, Jon Michael Mollish, 734
Jeffrey W. Simmons, Solomon R. David, Gregory Watkins-Colwell, Eva A. Hoffman, 735
and Thomas J. Near. Hidden species diversity in an iconic living fossil vertebrate. 736
Biology Letters, 18(11):20220395, 11 2022. ISSN 1744-9561. doi: 10.1098/rsbl.2022. 737
0395. URL <https://doi.org/10.1098/rsbl.2022.0395>. 738
- [11] Hengchi Chen, Arthur Zwaenepoel, and Yves Van de Peer. wgd v2: a suite of tools to 739
uncover and date ancient polyploidy and whole-genome duplication. *Bioinformatics*, 740
40(5):btae272, 04 2024. ISSN 1367-4811. doi: 10.1093/bioinformatics/btae272. URL 741
<https://doi.org/10.1093/bioinformatics/btae272>. 742
- [12] Peilin Cheng, Yu Huang, Yunyun Lv, Hao Du, Zhiqiang Ruan, Chuangju Li, Huan 743
Ye, Hui Zhang, Jinming Wu, Chengyou Wang, Rui Ruan, Yanping Li, Chao Bian, 744
Xinxin You, Chengcheng Shi, Kai Han, Junming Xu, Qiong Shi, and Qiwei Wei. The 745
American Paddlefish Genome Provides Novel Insights into Chromosomal Evolution 746
and Bone Mineralization in Early Vertebrates. *Molecular Biology and Evolution*, 38 747
(4):msaa326, 2020. ISSN 0737-4038. doi: 10.1093/molbev/msaa326. 748
- [13] James W Clark and Philip CJ Donoghue. Constraining the timing of whole genome 749
duplication in plant evolutionary history. *Proceedings of the Royal Society B: Biological 750
Sciences*, 284(1858):20170912, 2017. 751
- [14] Gavin C. Conant and Kenneth H. Wolfe. Turning a hobby into a job: How duplicated 752
genes find new functions. *Nature Reviews Genetics*, 9(12):938–950, 2008. ISSN 753
1471-0056. doi: 10.1038/nrg2482. 754
- [15] Karen D. Crow, Christopher D. Smith, Jan-Fang Cheng, Günter P. Wagner, and 755
Chris T. Amemiya. An Independent Genome Duplication Inferred from Hox Paralogs 756
in the American Paddlefish—A Representative Basal Ray-Finned Fish and Important 757
Comparative Reference. *Genome Biology and Evolution*, 4(9):937–953, 2012. ISSN 758
1759-6653. doi: 10.1093/gbe/evs067. 759

- [16] Paramvir Dehal and Jeffrey L Boore. Two Rounds of Whole Genome Duplication
in the Ancestral Vertebrate. *PLoS Biology*, 3(10):e314, 2005. ISSN 1544-9173. doi:
10.1371/journal.pbio.0030314.
- [17] Kang Du, Matthias Stöck, Susanne Kneitz, Christophe Klopp, Joost M. Woltering,
Mateus Contar Adolphi, Romain Feron, Dmitry Prokopov, Alexey Makunin, Ilya
Kichigin, Cornelia Schmidt, Petra Fischer, Heiner Kuhl, Sven Wuertz, Jörn Gessner,
Werner Kloas, Cédric Cabau, Carole Iampietro, Hugues Parrinello, Chad Tomlinson,
Laurent Journot, John H. Postlethwait, Ingo Braasch, Vladimir Trifonov, Wesley C.
Warren, Axel Meyer, Yann Guiguen, and Manfred Scharl. The sterlet sturgeon
genome sequence and the mechanisms of segmental rediploidization. *Nature Ecology
& Evolution*, pages 1–12, 2020. doi: 10.1038/s41559-020-1166-x.
- [18] Olga Dudchenko, Sanjit S. Batra, Arina D. Omer, Sarah K. Nyquist, Marie Hoeger,
Neva C. Durand, Muhammad S. Shamim, Ido Machol, Eric S. Lander, Aviva Presser
Aiden, and Erez Lieberman Aiden. De novo assembly of the *Aedes aegypti* genome
using Hi-C yields chromosome-length scaffolds. *Science*, 356(6333):92–95, 2017. ISSN
0036-8075. doi: 10.1126/science.aal3327.
- [19] Neva C Durand, Muhammad S Shamim, Ido Machol, Suhas S P Rao, Miriam H
Huntley, Eric S Lander, and Erez Lieberman Aiden. Juicer Provides a One-Click
System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*, 3(1):95–98,
2016. ISSN 2405-4712. doi: 10.1016/j.cels.2016.07.002.
- [20] Peter Edge, Vineet Bafna, and Vikas Bansal. HapCUT2: robust and accurate
haplotype assembly for diverse sequencing technologies. *Genome Research*, 27(5):
801–812, 2017. ISSN 1088-9051. doi: 10.1101/gr.213462.116.
- [21] David M. Emms and Steven Kelly. OrthoFinder: phylogenetic orthology infer-
ence for comparative genomics. *Genome Biology*, 20(1):238, 2019. doi: 10.1186/
s13059-019-1832-y.

- [22] Figshare. <https://figshare.com/s/26c4deaa37f28e13ff98> - a duplicate-resolved pad-
dlefish genome provides insights into the mechanisms of rediploidisation and hox
cluster evolution. *n/a*, 01 2026. ISSN 0737-4038. doi: 10.1093/molbev/msaa015.
URL <https://figshare.com/s/26c4deaa37f28e13ff98>.
- [23] Linda Frey, Michael Coates, Michał Ginter, Vachik Hairapetian, Martin Rücklin,
Iwan Jerjen, and Christian Klug. The early elasmobranch phoebodus: phylogenetic
relationships, ecomorphology and a new time-scale for shark evolution. *Proceedings*
of the Royal Society B, 286(1912):20191336, 2019.
- [24] Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. CD-HIT:
accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28
(23):3150–3152, 2012. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts565.
- [25] Rebecca F. Furlong and Peter W. H. Holland. Were vertebrates octoploid? *Philo-*
sophical Transactions of the Royal Society of London. Series B: Biological Sciences,
357(1420):531–544, 2002. ISSN 0962-8436. doi: 10.1098/rstb.2001.1035.
- [26] Erik Garrison and Gabor Marth. Haplotype-based variant detection from short-read
sequencing. *arxiv*, 2012.
- [27] Gareth B Gillard, Lars Grønvold, Line L Røsæg, Matilde Mengkrog Holen, Øystein
Monsen, Ben F Koop, Eric B Rondeau, Manu Kumar Gundappa, John Mendoza,
Daniel J Macqueen, et al. Comparative regulomics supports pervasive selection on
gene dosage following whole genome duplication. *Genome Biology*, 22(1):103, 2021.
- [28] Manu Kumar Gundappa, Thu-Hien To, Lars Grønvold, Samuel A.M. Martin, Sigbjørn
Lien, Juergen Geist, David Hazlerigg, Simen R. Sandve, and Daniel J. Macqueen.
Genome-wide reconstruction of rediploidization following autopolyploidization across
one hundred million years of salmonid evolution. *bioRxiv*, page 2021.06.05.447185,
2021. doi: 10.1101/2021.06.05.447185.

- [29] Eric J Hilton and Lance Grande. Review of the fossil record of sturgeons, family
acipenseridae (actinopterygii: Acipenseriformes), from north america. *Journal of*
Paleontology, 80(4):672–683, 2006.
- [30] Katharina J Hoff, Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke. Whole-
genome annotation with braker. In *Gene prediction: methods and protocols*, pages
65–95. Springer, 2019.
- [31] Karsten Hokamp, Aoife McLysaght, and Kenneth H. Wolfe. The 2r hypothesis and
the human genome sequence. *Journal of Structural and Functional Genomics*, 3(1-4):
95–110, 2003. ISSN 1345-711X. doi: 10.1023/a:1022661917301.
- [32] Peter WH Holland and Jordi Garcia-Fernàndez. Hoxgenes and chordate evolution.
Developmental biology, 173(2):382–395, 1996.
- [33] Peter WH Holland, Jordi Garcia-Fernàndez, Nic A Williams, and Arend Sidow.
Gene duplications and the origins of vertebrate development. *Development*, 1994
(Supplement):125–133, 1994.
- [34] Jesse D Hollister, Brian J Arnold, Elisabeth Svedin, Katherine S Xue, Brian P
Dilkes, and Kirsten Bomblies. Genetic adaptation associated with genome-doubling
in autotetraploid arabidopsis arenosa. *PLoS genetics*, 8(12):e1003093, 2012.
- [35] Jun Inoue, Philip CJ Donoghue, and Ziheng Yang. The impact of the representation
of fossil calibrations on bayesian estimation of species divergence times. *Systematic*
biology, 59(1):74–89, 2010.
- [36] Iker Irisarri, Denis Baurain, Henner Brinkmann, Frédéric Delsuc, Jean-Yves Sire,
Alexander Kupfer, Jörn Petersen, Michael Jarek, Axel Meyer, Miguel Vences, et al.
Phylotranscriptomic consolidation of the jawed vertebrate timetree. *Nature ecology*
& evolution, 1(9):1370–1378, 2017.
- [37] Olivier Jaillon, Jean-Marc Aury, Frédéric Brunet, Jean-Louis Petit, Nicole Stange-
Thomann, Evan Mauceli, Laurence Bouneau, Cécile Fischer, Catherine Ozouf-Costaz,

- Alain Bernot, Sophie Nicaud, David Jaffe, Sheila Fisher, Georges Lutfalla, Car- 837
 ole Dossat, Béatrice Segurens, Corinne Dasilva, Marcel Salanoubat, Michael Levy, 838
 Nathalie Boudet, Sergi Castellano, Véronique Anthouard, Claire Jubin, Vanina 839
 Castelli, Michael Katinka, Benoît Vacherie, Christian Biémont, Zineb Skalli, Laurence 840
 Cattolico, Julie Poulain, Véronique de Berardinis, Corinne Cruaud, Simone Duprat, 841
 Philippe Brottier, Jean-Pierre Coutanceau, Jérôme Gouzy, Genis Parra, Guillaume 842
 Lardier, Charles Chapple, Kevin J. McKernan, Paul McEwan, Stephanie Bosak, 843
 Manolis Kellis, Jean-Nicolas Volff, Roderic Guigó, Michael C. Zody, Jill Mesirov, 844
 Kerstin Lindblad-Toh, Bruce Birren, Chad Nusbaum, Daniel Kahn, Marc Robinson- 845
 Rechavi, Vincent Laudet, Vincent Schachter, Francis Quétier, William Saurin, Claude 846
 Scarpelli, Patrick Wincker, Eric S. Lander, Jean Weissenbach, and Hugues Roest 847
 Crollius. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the 848
 early vertebrate proto-karyotype. *Nature*, 431(7011):946–957, 2004. ISSN 0028-0836. 849
 doi: 10.1038/nature03025. 850
- [38] Caelinn James, Marco Trevisan-Herraz, David Juan, and Daniel Rico. Evolutionary 851
 analysis of gene ages across tads associates chromatin topology with whole-genome 852
 duplications. *Cell Reports*, 43(4), 2024. 853
- [39] David R Kelley and Steven L Salzberg. Detection and correction of false segmental 854
 duplications caused by genome mis-assembly. *Genome Biology*, 11(3):R28, 2010. 855
 ISSN 1465-6906. doi: 10.1186/gb-2010-11-3-r28. 856
- [40] Daehwan Kim, Ben Langmead, and Steven L Salzberg. HISAT: a fast spliced aligner 857
 with low memory requirements. *Nature Methods*, 12(4):357–360, 2015. ISSN 1548-7091. 858
 doi: 10.1038/nmeth.3317. 859
- [41] Benjamin L. King, J. Andrew Gillis, Heather R. Carlisle, and Randall D. Dahn. A 860
 natural deletion of the *ijhoxc/ij* cluster in elasmobranch fishes. *Science*, 334(6062): 861
 1517–1517, 2011. doi: 10.1126/science.1210912. URL <https://www.science.org/> 862
<https://doi.org/10.1126/science.1210912>. 863

- [42] Byung June Ko, Chul Lee, Juwan Kim, Arang Rhie, Dong Ahn Yoo, Kerstin Howe, Jonathan Wood, Seoae Cho, Samara Brown, Giulio Formenti, Erich D. Jarvis, and Heebal Kim. Widespread false gene gains caused by duplication errors in genome assemblies. *Genome Biology*, 23(1):205, 2022. ISSN 1474-7596. doi: 10.1186/s13059-022-02764-1.
- [43] Jenő Káldy, Attila Mozsár, Gyöngyvér Fazekas, Móni Farkas, Dorottya Lilla Fazekas, Georgina Lea Fazekas, Katalin Goda, Zsuzsanna Gyöngy, Balázs Kovács, Kenneth Semmens, Miklós Bercsényi, Mariann Molnár, and Eszter Patakiné Várkonyi. Hybridization of russian sturgeon (*acipenser gueldenstaedtii*, brandt and ratzeberg, 1833) and american paddlefish (*polyodon spathula*, walbaum 1792) and evaluation of their progeny. *Genes*, 11(7), 2020. ISSN 2073-4425. doi: 10.3390/genes11070753. URL <https://www.mdpi.com/2073-4425/11/7/753>.
- [44] Bruce T. Lahn and David C. Page. Four Evolutionary Strata on the Human X Chromosome. *Science*, 286(5441):964–967, 1999. ISSN 0036-8075. doi: 10.1126/science.286.5441.964.
- [45] Ben Langmead, Cole Trapnell, Mihai Pop, and Steven L Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3):R25, 2009. ISSN 1465-6906. doi: 10.1186/gb-2009-10-3-r25.
- [46] Nicolas Lartillot, Thomas Lepage, and Samuel Blanquart. Phylobayes 3: a bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*, 25(17):2286–2288, 06 2009. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp368. URL <https://doi.org/10.1093/bioinformatics/btp368>.
- [47] Heng Li. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21):2987–2993, 2011. ISSN 1367-4803. doi: 10.1093/bioinformatics/btr509.
- [48] Heng Li and Richard Durbin. Fast and accurate short read alignment with Bur-

- rows–Wheeler transform. *Bioinformatics*, 25(14):1754–1760, 2009. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp324.
- [49] Jiong-Tang Li, Guang-Yuan Hou, Xiang-Fei Kong, Chun-Yan Li, Jian-Ming Zeng, Heng-De Li, Gui-Bao Xiao, Xiao-Min Li, and Xiao-Wen Sun. The fate of recent duplicated genes following a fourth-round whole genome duplication in a tetraploid fish, common carp (*Cyprinus carpio*). *Scientific Reports*, 5(1):8199, 2015. doi: 10.1038/srep08199.
- [50] Sigbjørn Lien, Ben F Koop, Simen R Sandve, Jason R Miller, Matthew P Kent, Torfinn Nome, Torgeir R Hvidsten, Jong S Leong, David R Minkley, Aleksey Zimin, Fabian Grammes, Harald Grove, Arne Gjuvsland, Brian Walenz, Russell A Hermansen, Kris von Schalburg, Eric B Rondeau, Alex Di Genova, Jeevan K A Samy, Jon Olav Vik, Magnus D Vigeland, Lis Caler, Unni Grimholt, Sissel Jentoft, Dag Inge Våge, Pieter de Jong, Thomas Moen, Matthew Baranski, Yniv Palti, Douglas R Smith, James A Yorke, Alexander J Nederbragt, Ave Tooming-Klunderud, Kjetill S Jakobsen, Xuanning Jiang, Dingding Fan, Yan Hu, David A Liberles, Rodrigo Vidal, Patricia Iturra, Steven J M Jones, Inge Jonassen, Alejandro Maass, Stig W Omholt, and William S Davidson. The Atlantic salmon genome provides insights into rediploidization. *Nature*, 533(7602):200–205, 2016. ISSN 0028-0836. doi: 10.1038/nature17164.
- [51] Binghang Liu, Yujian Shi, Jianying Yuan, Xuesong Hu, Hao Zhang, Nan Li, Zhenyu Li, Yanxiang Chen, Desheng Mu, and Wei Fan. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects, 2020.
- [52] Helen S Long, Simon Greenaway, George Powell, Ann-Marie Mallon, Cecilia M Lindgren, and Michelle M Simon. Making sense of the linear genome, gene function and tads. *Epigenetics & Chromatin*, 15(1):4, 2022.
- [53] Daniel J. Macqueen and Ian A. Johnston. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species

- diversification. *Proceedings of the Royal Society B: Biological Sciences*, 281(1778): 916
20132881, 2014. ISSN 0962-8452. doi: 10.1098/rspb.2013.2881. 917
- [54] Mosè Manni, Matthew R. Berkeley, Mathieu Seppey, and Evgeny M. Zdobnov. 918
Busco: Assessing genomic data quality and beyond. *Current Protocols*, 1(12):e323, 919
2021. doi: <https://doi.org/10.1002/cpz1.323>. URL <https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpz1.323>. 920
921
- [55] Sarah Marburger, Patrick Monnahan, Paul J Seear, Simon H Martin, Jordan Koch, 922
Pirita Paaajanen, Magdalena Bohutínská, James D Higgins, Roswitha Schmickl, 923
and Levi Yant. Interspecific introgression mediates adaptation to whole genome 924
duplication. *Nature Communications*, 10(1):5218, 2019. 925
- [56] Kyle J Martin and Peter WH Holland. Enigmatic orthology relationships between hox 926
clusters of the african butterfly fish and other teleosts following ancient whole-genome 927
duplication. *Molecular Biology and Evolution*, 31(10):2592–2611, 2014. 928
- [57] Gemma I. Martínez-Redondo, Israel Barrios-Núñez, Marçal Vázquez-Valls, Ana M. 929
Rojas, and Rosa Fernández. Illuminating the functional landscape of the dark 930
proteome across the Animal Tree of Life through natural language processing models. 931
bioRxiv, page 2024.02.28.582465, 2024. doi: 10.1101/2024.02.28.582465. 932
- [58] Uirá Souto Melo, Robert Schöpflin, Rocio Acuna-Hidalgo, Martin Atta Mensah, 933
Björn Fischer-Zirnsak, Manuel Holtgrewe, Marius-Konstantin Klever, Seval Türkmen, 934
Verena Heinrich, Ilina Datkhaeva Pluym, et al. Hi-c identifies complex genomic 935
rearrangements and tad-shuffling in developmental diseases. *The American Journal* 936
of Human Genetics, 106(6):872–884, 2020. 937
- [59] Bui Quang Minh, Minh Anh Thi Nguyen, and Arndt Von Haeseler. Ultrafast 938
approximation for phylogenetic bootstrap. *Molecular biology and evolution*, 30(5): 939
1188–1195, 2013. 940

- [60] Bui Quang Minh, Heiko A Schmidt, Olga Chernomor, Dominik Schrempf, Michael D 941
Woodhams, Arndt von Haeseler, and Robert Lanfear. IQ-TREE 2: New Models and 942
Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology* 943
and Evolution, 37(5):1530–1534, 02 2020. ISSN 0737-4038. doi: 10.1093/molbev/ 944
msaa015. URL <https://doi.org/10.1093/molbev/msaa015>. 945
- [61] Yoichiro Nakatani and Aoife McLysaght. Macrosynteny analysis shows the absence of 946
ancient whole-genome duplication in lepidopteran insects. *Proceedings of the National* 947
Academy of Sciences, 116(6):1816–1818, 2019. doi: 10.1073/pnas.1817937116. URL 948
<https://www.pnas.org/doi/abs/10.1073/pnas.1817937116>. 949
- [62] Yoichiro Nakatani, Prashant Shingate, Vydianathan Ravi, Nisha E Pillai, Ar- 950
avind Prasad, Aoife McLysaght, and Byrappa Venkatesh. Reconstruction of proto- 951
vertebrate, proto-cyclostome and proto-gnathostome genomes provides new insights 952
into early vertebrate evolution. *Nature communications*, 12(1):4489, 2021. 953
- [63] Lukasz Niezabitowski, Anthony K Redmond, and Aoife McLysaght. An automated 954
pipeline for reconstructing whole genome duplications. *bioRxiv*, 2026. doi: 10.64898/ 955
2026.08.10.743943. URL [https://www.biorxiv.org/content/early/2026/08/11/](https://www.biorxiv.org/content/early/2026/08/11/2026.08.10.743943) 956
2026.08.10.743943. 957
- [64] Mariam Okhovat, Jake VanCampen, Kimberly A Nevonen, Lana Harshman, Weiyu 958
Li, Cora E Layman, Samantha Ward, Jarod Herrera, Jackson Wells, Rory R Sheng, 959
et al. Tad evolutionary and functional characterization reveals diversity in mammalian 960
tad boundary properties and function. *Nature communications*, 14(1):8111, 2023. 961
- [65] Elise Parey, Alexandra Louis, Jérôme Montfort, Yann Guiguen, Hugues Roest Crol- 962
lius, and Camille Berthelot. An atlas of fish genome evolution reveals delayed 963
rediploidization following the teleost whole-genome duplication. *Genome Research*, 964
32(9):1685–1697, 2022. ISSN 1088-9051. doi: 10.1101/gr.276953.122. 965
- [66] Brent S Pedersen and Aaron R Quinlan. Mosdepth: quick coverage calculation for 966

- genomes and exomes. *Bioinformatics*, 34(5):867–868, 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/btx699.
- [67] Sebastian Proost, Jan Fostier, Dieter De Witte, Bart Dhoedt, Piet Demeester, Yves Van de Peer, and Klaas Vandepoele. i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Research*, 40(2): e11–e11, 2012. ISSN 0305-1048. doi: 10.1093/nar/gkr955.
- [68] Minbo Qi, James Clark, Edmund R R Moody, Davide Pisani, and Philip C J Donoghue. Molecular dating of the teleost whole genome duplication (3r) is compatible with the expectations of delayed rediploidization. *Genome Biology and Evolution*, 16(7):evae128, 06 2024. ISSN 1759-6653. doi: 10.1093/gbe/evae128. URL <https://doi.org/10.1093/gbe/evae128>.
- [69] Vincent Ranwez, Emmanuel J P Douzery, Cédric Cambon, Nathalie Chantret, and Frédéric Delsuc. Macse v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Molecular Biology and Evolution*, 35(10):2582–2584, 08 2018. ISSN 0737-4038. doi: 10.1093/molbev/msy159. URL <https://doi.org/10.1093/molbev/msy159>.
- [70] Anthony K. Redmond, Dearbhaile Casey, Manu Kumar Gundappa, Daniel J. Macqueen, and Aoife McLysaght. Independent rediploidization masks shared whole genome duplication in the sturgeon-paddlefish ancestor. *Nature Communications*, 14(1):2879, 2023. doi: 10.1038/s41467-023-38714-z.
- [71] Fiona M. Robertson, Manu Kumar Gundappa, Fabian Grammes, Torgeir R. Hvidsten, Anthony K. Redmond, Sigbjørn Lien, Samuel A. M. Martin, Peter W. H. Holland, Simen R. Sandve, and Daniel J. Macqueen. Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biology*, 18(1):111, 2017. ISSN 1474-7596. doi: 10.1186/s13059-017-1241-z.

- [72] Hidetoshi Shimodaira. An approximately unbiased test of phylogenetic tree selection. *Systematic biology*, 51(3):492–508, 2002.
- [73] Guy St C Slater and Ewan Birney. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6(1):31, 2005. ISSN 1471-2105. doi: 10.1186/1471-2105-6-31.
- [74] Douglas E. Soltis, Victor A. Albert, Jim Leebens-Mack, Charles D. Bell, Andrew H. Paterson, Chunfang Zheng, David Sankoff, Claude W. dePamphilis, P. Kerr Wall, and Pamela S. Soltis. Polyploidy and angiosperm diversification. *American Journal of Botany*, 96(1):336–348, 2009. ISSN 00029122, 15372197. URL <http://www.jstor.org/stable/27793091>.
- [75] Mario Stanke and Burkhard Morgenstern. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, 33(suppl_2):W465–W467, 2005. ISSN 0305-1048. doi: 10.1093/nar/gki458.
- [76] CC Swisher, Xiaolin Wang, Zhonghe Zhou, Yuanqing Wang, Fan Jin, Jiangyong Zhang, Xing Xu, Fucheng Zhang, and Yuan Wang. Further support for a cretaceous age for the feathered-dinosaur beds of liaoning, china: New 40 ar ÷ 39 ar dating of the yixian and tuchengzi formations. *Chinese Science Bulletin*, 47:136–139, 2002.
- [77] Radka Symonová, Miloš Havelka, Chris T. Amemiya, William Mike Howell, Tereza Kořínková, Martin Flajšhans, David Gela, and Petr Ráb. Molecular cytogenetic differentiation of paralogs of Hox paralogs in duplicated and re-diploidized genome of the North American paddlefish (*Polyodon spathula*). *BMC Genetics*, 18(1):19, 2017. doi: 10.1186/s12863-017-0484-8.
- [78] Damian Szklarczyk, Rebecca Kirsch, Mikaela Koutrouli, Katerina Nastou, Farrokh Mehryary, Radja Hachilif, Annika L Gable, Tao Fang, Nadezhda T Doncheva, Sampo Pyysalo, Peer Bork, Lars J Jensen, and Christian von Mering. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for

- any sequenced genome of interest. *Nucleic Acids Research*, 51(D1):D638–D646, 2022. 1019
ISSN 0305-1048. doi: 10.1093/nar/gkac1000. 1020
- [79] Tiina Sävilammi, Craig R Primmer, Srinidhi Varadharajan, René Guyomard, Yann 1021
Guiguen, Simen R Sandve, L Asbjørn Vøllestad, Spiros Papakostas, and Sigbjørn Lien. 1022
The chromosome-level genome assembly of european grayling reveals aspects of a 1023
unique genome evolution process within salmonids. *G3 Genes—Genomes—Genetics*, 1024
9(5):1283–1294, 05 2019. ISSN 2160-1836. doi: 10.1534/g3.118.200919. URL 1025
<https://doi.org/10.1534/g3.118.200919>. 1026
- [80] Morgane Thomas-Chollier, Valérie Ledent, Luc Leyns, and Michel Vervoort. A 1027
non-tree-based comprehensive study of metazoan hox and parahox genes prompts 1028
new insights into their origin and evolution. *BMC evolutionary biology*, 10(1):73, 1029
2010. 1030
- [81] Y Uno, C Nishida, C Takagi, N Ueno, and Y Matsuda. Homoeologous chromosomes 1031
of xenopus laevis are highly conserved after whole-genome duplication. *Heredity*, 111 1032
(5):430–436, 2013. 1033
- [82] Anne-Laure Valton and Job Dekker. Tad disruption as oncogenic driver. *Current* 1034
opinion in genetics & development, 36:34–40, 2016. 1035
- [83] Yves Van de Peer, Eshchar Mizrachi, and Kathleen Marchal. The evolutionary 1036
significance of polyploidy. *Nature Reviews Genetics*, 18(7):411–424, 2017. 1037
- [84] Carlos Vargas-Chávez, Lisandra Benítez-Álvarez, Gemma I Martínez-Redondo, Lucía 1038
Álvarez-González, Judit Salces-Ortiz, Klara Eleftheriadi, Nuria Escudero, Nadège 1039
Guiglielmoni, Jean-François Flot, Marta Novo, et al. An episodic burst of massive 1040
genomic rearrangements and the origin of non-marine annelids. *Nature ecology &* 1041
evolution, 9(7):1263–1279, 2025. 1042
- [85] Byrappa Venkatesh, Alison P. Lee, Vydianathan Ravi, Ashish K. Maurya, Michelle M. 1043
Lian, Jeremy B. Swann, Yuko Ohta, Martin F. Flajnik, Yoichi Sutoh, Masanori 1044

- Kasahara, Shawn Hoon, Vamshidhar Gangu, Scott W. Roy, Manuel Irimia, Vladimir 1045
- Korzh, Igor Kondrychyn, Zhi Wei Lim, Boon-Hui Tay, Sumanty Tohari, Kiat Whye 1046
- Kong, Shufen Ho, Belen Lorente-Galdos, Javier Quilez, Tomas Marques-Bonet, 1047
- Brian J. Raney, Philip W. Ingham, Alice Tay, LaDeana W. Hillier, Patrick Minx, 1048
- Thomas Boehm, Richard K. Wilson, Sydney Brenner, and Wesley C. Warren. Elephant 1049
- shark genome provides unique insights into gnathostome evolution. *Nature*, 505(7482): 1050
- 174–179, 2014. ISSN 0028-0836. doi: 10.1038/nature12826. 1051
- [86] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, 1052
- David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan 1053
- Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, 1054
- Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, 1055
- C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, 1056
- Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, 1057
- Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and 1058
- SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing 1059
- in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2. 1060
- [87] Jonathan F Wendel. Genome evolution in polyploids. *Plant molecular evolution*, 1061
- pages 225–249, 2000. 1062
- [88] Kenneth H Wolfe and Denis C Shields. Molecular evidence for an ancient duplication 1063
- of the entire yeast genome. *Nature*, 387(6634):708–713, 1997. ISSN 0028-0836. doi: 1064
- 10.1038/42711. 1065
- [89] Joachim Wolff, Vivek Bhardwaj, Stephan Nothjunge, Gautier Richard, Gina Ren- 1066
- schler, Ralf Gilsbach, Thomas Manke, Rolf Backofen, Fidel Ramírez, and Björn A 1067
- Grüning. Galaxy hicexplorer: a web server for reproducible hi-c data analysis, quality 1068
- control and visualization. *Nucleic Acids Research*, 46(W1):W11–W16, 06 2018. ISSN 1069
- 0305-1048. doi: 10.1093/nar/gky504. URL <https://doi.org/10.1093/nar/gky504>. 1070
- [90] Chuanshuai Xie, Zitu Ma, Chaowei Zhou, Kexin Ma, Haoyu Wang, Jiahong Wu, 1071

- Yan Zhou, Yongrui Lu, Da Ji, Xuedie Gu, et al. Chromosomal fusions trigger rediploidization of autopolyploid genomes. *Nature*, pages 1–8, 2026.
- [91] Kazuaki Yamaguchi, Yoshinobu Uno, Mitsutaka Kadota, Osamu Nishimura, Ryo Nozu, Kiyomi Murakumo, Rui Matsumoto, Keiichi Sato, and Shigehiro Kuraku. Elasmobranch genome sequencing reveals evolutionary trends of vertebrate karyotype organization. *Genome Research*, 33(9):1527–1540, 2023.
- [92] Ziheng Yang and Bruce Rannala. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, 23(1):212–226, 09 2005. ISSN 0737-4038. doi: 10.1093/molbev/msj024. URL <https://doi.org/10.1093/molbev/msj024>.
- [93] Xingtang Zhang, Ruoxi Wu, Yibin Wang, Jiaxin Yu, and Haibao Tang. Unzipping haplotypes in diploid and polyploid genomes. *Computational and Structural Biotechnology Journal*, 18:66–72, 2019. ISSN 2001-0370. doi: 10.1016/j.csbj.2019.11.011.