



Chromosome-level genome of the Adriatic sturgeon, *Acipenser naccarii*: A resource for polyploid fish genomics

Roberto Biello^{1,*} , Annalisa Scapolatiello² , Sebastiano Fava³ , Victor Hugo Muñoz Mora¹ , Stefano Dalle Palle² , Patrícia Santos¹ , Alessio Iannucci⁴ , Tatiana Tilley⁵ , Nivesh Jain⁵ , Jennifer Balacco⁵ , Brian O'Toole⁵ , Giulio Formenti^{4,5} , Erich D. Jarvis^{5,6} , Claudio Ciofi⁴ , Emiliano Trucchi³ , Leonardo Congiu^{2,7} , Giorgio Bertorelle¹ , Andrea Benazzo^{1,*} 

¹Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy

²Department of Biology, University of Padua, Padua, Italy

³Department of Life and Environmental Sciences, Marche Polytechnic University, Ancona, Italy

⁴Department of Biology, University of Florence, Florence, Italy

⁵The Vertebrate Genome Laboratory, The Rockefeller University, New York, NY, United States

⁶Howard Hughes Medical Institute, Chevy Chase, MD, United States

⁷National Biodiversity Future Center, Palermo, Italy

*Corresponding authors: Roberto Biello, Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy. Email: roberto.biello@unife.it; Andrea Benazzo. Email: andrea.benazzo@unife.it

Abstract

The Adriatic sturgeon, *Acipenser naccarii*, a tetraploid species endemic to the North Adriatic region, has experienced significant population declines, resulting in its classification as “Critically Endangered” by the International Union for Conservation of Nature (IUCN). Historically widespread in the Adriatic Sea’s tributaries, the species is now at high risk of extinction with occasional reproductions occurring in the wild. Using long-read sequencing (PacBio HiFi) and chromatin conformation capture sequencing (Hi-C), we generated a phased reference genome for the tetraploid Adriatic sturgeon. The haploid assembly spans 1.94 Gb across 2,083 scaffolds, with a contig N50 of 1.024 Mb, a scaffold N50 of 39.6 Mb, and a scaffold L90 of 276. Approximately 80% of the genome is contained within the first 60 scaffolds, indicating a high degree of contiguity. The Benchmarking Universal Single-Copy Orthologs (BUSCO) completeness score of the haploid assembly reached 88.9%, whereas combined metrics for all four haploid assemblies increased to 94.1%. This comprehensive genomic resource provides valuable insights into the genetic and evolutionary mechanisms of polyploidy and, more specifically, it will improve our understanding of the genetic diversity of the Adriatic sturgeon, thereby informing targeted conservation strategies for this critically endangered species.

Keywords conservation genomics, endemic species, endemixit, genome assembly, polyploidy

Introduction

The Adriatic sturgeon, *Acipenser naccarii* (Fig. 1a), is a critically endangered tetraploid endemism of the Adriatic region. Once abundant, the species experienced a catastrophic population decline throughout the 20th century, culminating in its near extinction in the wild by the 1990s.

The Adriatic sturgeon’s decline reflects the broader crisis facing sturgeons worldwide. As ancient Acipenseriformes, their late maturity, long lifespans, and size-dependent fecundity once provided evolutionary advantages but now make them highly vulnerable to human impacts (Billard and Lecointre 2000). Many sturgeons are anadromous, requiring access to freshwater spawning grounds like gravel beds for successful reproduction (Sulak and Randall 2002). However, dams and other barriers have disrupted these migrations, cutting off critical habitats

(Williot et al. 2002). Overexploitation for caviar and meat has further accelerated declines by disproportionately removing the largest, most reproductively valuable individuals.

Its persistence today is largely attributable to a single founder stock preserved in a private aquaculture facility (Congiu et al. 2011). Over the past four decades, intensive conservation programs, primarily centered in Italy’s Po River basin, have tried to recover the species population by releasing hundreds of thousands of juvenile sturgeons derived from this limited genetic pool (Arlati and Poliakova 2009). However, the species’ long-life cycle, spanning more than 50 years, means that the success of these efforts will only be measurable over extended timescales. Only recently, in fact, occasional natural reproductions have been recorded in the Ticino and Livenza rivers allowing the reclassification by IUCN from *Critically Endangered* and *Possibly Extinct in the Wild* to *Critically Endangered*.

Received: 20 June 2025. Revised: 18 January 2026. Accepted: 10 March 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of The American Genetic Association.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

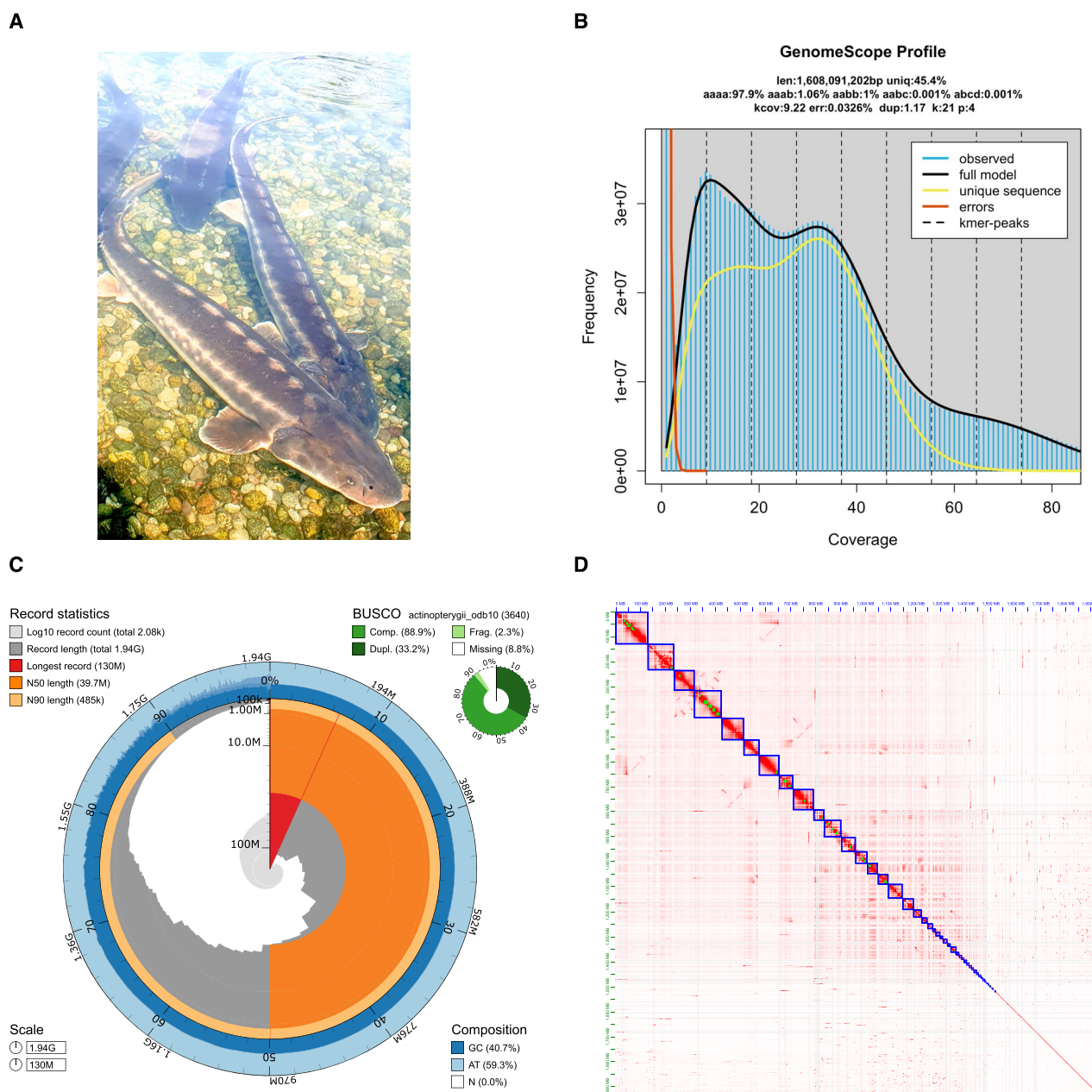


Fig. 1 Visual overview of *Acipenser naccarii* genome assembly metrics. a) Photography of *A. naccarii* individuals (photo credit: Annalisa Scapolatiello). b) K-mer spectrum output and corresponding genome size and heterozygosity estimated with GenomeScope 2.0. c) BlobToolKit snail plot showing a graphical representation of the quality metrics presented in Tables 2 and 3 for the *A. naccarii* assembly (hap1). The longest scaffold (red vertical line), N50 (orange track), N90 (light orange track), GC content (external blue track), and BUSCO scores (outer circular pie chart in green) are shown. d) Hi-C contact map for the scaffolds of the hap1 genome assembly generated with Juicebox.

The ploidy level of sturgeon species has been a subject of scientific debate, primarily due to the complex evolutionary history involving multiple whole-genome duplication events. According to Birstein et al. (1997) and Ludwig et al. (2001), the common ancestor of all sturgeons had a diploid genome with approximately 60 chromosomes. A first major whole-genome duplication likely occurred during the early Jurassic period (Fontana et al. 2008), leading to a tetraploid condition with around 120 chromosomes. Over time, however, this genome underwent a process known as functional diploidization

(Stift et al. 2008), during which, despite the doubled chromosome count, the genome, or at least a relevant part of it, began to function more like a diploid. This has led to debate over whether these sturgeons with 120 chromosomes should be considered tetraploid or functionally diploid. Subsequently, further independent duplication events occurred both within the Pacific and Atlantic sturgeon clades. These events gave rise to species with approximately 240 chromosomes. Although this chromosome count would nominally suggest an octoploid genome, they are mostly functionally tetraploid. To further

complicate the picture, also following this last duplication events, a functional diploidization process likely started to reduce the genomic redundancy (Stift et al. 2008).

The Adriatic sturgeon is one of the 240 chromosome species and is generally considered as a functional tetraploid, as revealed by in situ hybridization analyses of 28S and 5S rDNA (Fontana et al. 1999) and by microsatellite data, with most loci exhibiting up to four alleles per individual. Nevertheless, some microsatellite loci exhibit more than four alleles in numerous individuals (Boscari et al. 2015) either due to local duplication events or as a legacy of the nominal octoploid condition (Ludwig et al. 2001; Havelka et al. 2013). On the other hand, a functional diploidization is taking place as shown by some microsatellite loci with four alleles per individual that segregate according to a double disomic pattern (Dalle Palle et al. 2022). Consequently, in the genome of *A. naccarii*, as well as in all the 240 chromosome species, it is possible that different genomic regions present different levels of ploidy (diploid, tetraploid, octoploid), though functionally tetraploid regions are expected to predominate.

Here, we present a high-quality chromosome-scale reference genome for *A. naccarii* (Fig. 1a), produced as part of ENDEMIXIT (www.endemixit.com) and Vertebrate Genomes Project (<https://www.vertebrategenomesproject.org/>). This high-quality reference genome is a valuable resource for both the conservation of this critically endangered species and the study of genome evolution in sturgeons. As one of the first chromosome-level assemblies available for polyploid sturgeons (Wang et al. 2024), it offers key insights into the genetic and evolutionary mechanisms underlying polyploidy, contributing to a broader understanding of genome duplication in vertebrates. In the face of persistent environmental and anthropogenic challenges, this genomic resource represents a critical step toward to help improving the conservation strategy for this species.

Material and methods

Sampling, genomic DNA extraction, and sequencing

Two juvenile sturgeon specimens of unknown sex, born as a result of the June 2020 breeding program, were collected after natural death and generously donated by Allevamento Storione Ticino, located in Cassolnovo (Italy). The first was used for genome sequencing, whereas the second specimen was utilized for transcriptome analysis.

All the following steps were carried out at the Vertebrate Genomes Laboratory (<https://vertebrategenomelab.org/>) of the Rockefeller University. High molecular weight (HMW) DNA was extracted from muscle with the Circulomics HMW DNA extraction standard TissueRuptor protocol with the Nanobind Tissue Big DNA Kit (PN NB-900-701-01). DNA absorbance was checked for quality and purity control with Nanodrop, and average fragment length was verified with a pulsed field gel electrophoresis. Genomic data from 2 different sequencing technologies were used for the assembly: Pacific Biosciences (PacBio) High Fidelity (HiFi) reads and Hi-C reads from Arima Genomics. PacBio HiFi libraries were prepared using the Pacific Biosciences Express Template Prep Kit 2.0. The library was then size selected (>10 kb) using the Circulomics Short Read Eliminator. The PacBio library was

sequenced on 2 PacBio 8M v3 SMRT Cells on a PacBio Sequel II and 1 PacBio 8M SMRT Cell on a PacBio Sequel IIe using the sequencing kit 2.0 and a 30-h movie. Hi-C libraries were generated by Arima Genomics (<https://arimagenomics.com/>) using muscle in vivo cross-linking with the Arima-HiC kit with 2-enzyme proximity ligation. Proximally ligated DNA was subjected to shearing, size selection (~200 to 600 bp) with Solid-Phase Reversible Immobilization (SPRI) beads, and enrichment with streptavidin beads for the biotin-labeled DNA. KAPA Hyper Prep kit was employed to generate libraries compatible with Illumina technologies. Libraries were amplified through Polymerase Chain Reaction (PCR), purified with SPRI beads, and sequenced on an Illumina HiSeq X (~60× coverage) after a quality check with Bioanalyzer and Quantitative PCR (qPCR).

RNA extraction and sequencing

RNA sequences were obtained from four different tissues: muscle, liver, gonads, and brain. RNA extraction was performed using the Direct-zol RNA Miniprep Kits, following the manufacturer's instructions. The total extracted RNA was subsequently subjected to RNA sequencing on a NextSeq 500 sequencer (Illumina) at the Next-Generation Sequencing (NGS) facility of the University of Padua, Italy.

Genome size and ploidy estimation

Quality control of HiFi reads was performed using fastqc v0.11.8 (Andrews 2010; <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Residual sequencing adapters were identified and removed using HiFiAdapterFilt v2.0.0 (Sim et al. 2022), setting a minimum of 44 bp and a 97% identity with the known adapter sequence.

To estimate the genome size of the Adriatic sturgeon, the KMC v3.0 tool (Kokot et al. 2017) was used to generate a k-mer spectrum with a k-mer length of 21 and filters set to -ci10 and -cx1200 to exclude low-frequency and excessively high-frequency k-mers, respectively. The k-mer spectrum was then analyzed using GenomeScope v2.0 (Ranallo-Benavidez et al. 2020), which uses k-mer frequency distributions to estimate genome size, repetitive content, and heterozygous/homozygous fraction of the genome.

To assess the ploidy of the genome, Smudgeplot v0.3.0 (Ranallo-Benavidez et al. 2020) and nQuire v1.0 (Weiß et al. 2018) were employed. Smudgeplot provides a reference-free approach to analyze genomes with complex ploidy levels by interpreting patterns in k-mer pair frequencies. Instead, nQuire requires a Binary Alignment Map (BAM) file as input. For this reason, the HiFi reads were aligned to the *Acipenser ruthenus* reference genome (Du et al. 2020). nQuire calculates base frequencies at biallelic Single Nucleotide Polymorphism (SNP) sites. These frequencies are compared with three predefined diploidy, triploidy, and tetraploidy models to infer ploidy levels.

Genome assembly

The *A. naccarii* genome was assembled using the VGP 2.0 assembly pipeline (Rhie et al. 2021), as outlined in Table 1, with an additional bioinformatic step to account for the ploidy of the species during Hi-C scaffolding.

Table 1 Pipeline and software used for the genome assembly.

Assembly step	Software	Version	Reference
Initial assembly			
Reads quality control	FastQC	0.11.8	Andrews (2010)
Adaptors residual removal	HiFiAdapterFilt	2.0.0	Sim et al. (2022)
k-mer spectrum generation	KMC	3.0	Kokot et al. (2017)
Estimation of genome size and heterozygosity	GenomeScope	2.0	Ranallo-Benavidez et al. (2020)
Estimation of genome ploidy	Smudgeplot	0.3.0	Ranallo-Benavidez et al. (2020)
	nQuire	1.0	Weiß et al. (2018)
Genome assembly and scaffolding			
De novo assembly (contigging)	Hifiasm	0.19.3	Cheng et al. (2021a); Cheng et al. (2022)
Hi-C alignment	Chromap	0.2.4	Zhang et al. (2021)
Contact processing	pairtools	1.0.2	Open2C et al. (2024)
Hi-C contact map visualization	Juicebox	2.16	Durand et al. (2016)
Hi-C scaffolding	Yahs	1.2	Zhou et al. (2023)
	RagTag	2.1	Alonge et al. (2022)
Genome annotation			
TE annotation	TE Annotator	1.9.9	Ou et al. (2019)
	DeepTE	1.0	Yan et al. (2020)
	RepeatMasker	4.1.2	Smit et al. (2021)
RNA-seq raw reads trimming	fastp	0.19.5	Chen et al. (2018)
Mapping RNA-seq reads genome	HISAT2	2.2.1	Kim et al. (2019)
SAM/BAM processing	samtools	1.11	Li et al. (2009)
Filtering splice junctions	Portcullis	1.2.4	Mapleson et al. (2018)
Gene prediction	Braker	3.0	Gabriel et al. (2024)
Gene combining	TSEBRA	1.0.3	Gabriel et al. (2021)
Mitochondrial assembly			
De novo genome assembly	MitoHiFi	3.2.1	Uliano-Silva et al. (2023)
Genome quality assessment			
Assembly completeness	BUSCO	5.3.2	Manni et al. (2021)
	Merqury	1.3	Rhie et al. (2020)
Contamination screening			
General contamination screening	BlobTools	1.1.1	Laetsch and Blaxter (2017)
Genome synten			
Synten	GENESPACE	1.2.3	Lovell et al. (2022)

HiFi and Hi-C reads were jointly processed in Hifiasm v0.19.3 (Cheng et al. 2021a; Cheng et al. 2022) in tetraploid mode ($-n$ -hap 4) to generate 4 phased haploid genome assemblies. Each genome assembly was then processed to identify and remove possible artifacts due to the presence of multiple allelic haplotigs from the same genomic regions, similarly to what was performed in Kuhl et al. (2022). Briefly, Hi-C reads were aligned to each assembly independently using Chromap v0.2.4 (Zhang et al. 2021) and pairtools v1.0.2 (Open2C et al. 2024) was then used to retain paired-end reads having a mapping quality >20 . Each scaffold was divided into not-overlapping 50 Kb windows, and the number of Hi-C contacts was counted for each pair of windows based on the alignments. Windows that did not contain any restriction site were excluded from the analysis. The observed frequency distribution of intra-windows contacts was then used to fit a mixture Poisson distribution model and identify the homozygous windows. The noise parameter epsilon, representing the maximum number of spurious contacts between independent regions of the genome,

was estimated by selecting the terminal windows of each scaffold in homozygous state and counting the number of contacts between the ones located in different scaffolds. Epsilon was set to the maximum observed value among all the possible window pairs. Each scaffold pair was compared, retaining the scaffold windows showing a number of contacts higher than epsilon. The window positions along the two scaffolds were used to perform a nonparametric Kendall correlation test. In each comparison, the shorter scaffold was classified as an allelic haplotig and removed from the assembly if it showed a proportion of valid windows greater than 0.6 and a significant correlation test result ($P < 0.01$). The four haploid assemblies were independently purged following the above bioinformatic procedure. A further scaffolding round for each haploid assembly was performed with Yahs v1.2 (Zhou et al. 2023) using Hi-C reads, setting the restriction enzyme motif “GATC, GATC, CTNAG, TTAA” as suggested for Arima Genomics data. Finally, the manual curation of each assembly was performed using Juicebox v2.16 (Durand et al. 2016). To further improve

the quality of the assemblies, potential misassemblies were detected in the most contiguous haploid assembly, among the four assembled, using the RagTag v2.1 (Alonge et al. 2022) *correct* module, setting the chromosome-scale *A. ruthenus* genome as the reference sequence. The correction step was conducted using the long-reads coverage profile to validate each putative misassembly point, hence a breakpoint was inserted only if supported by the reference alignment and a concurrent drop in sequencing depth. The RagTag *scaffold* module was then used to perform a final scaffolding round on the corrected haploid assembly, introducing the proposed joints only if they resulted in an improved Hi-C contact pattern by a visual inspection. The resulting curated haploid assembly was finally used as a guide (reference sequence) to correct and scaffold the other three haploid assemblies using RagTag. Also in this step, the proposed edits from RagTag were introduced in the assemblies if compatible with the observed Hi-C contact matrix using the Juicebox platform.

We assessed the quality of our genome assemblies using three methods. First, we used the BUSCO quality control tool to check for genome completeness using a set of conserved single-copy orthologous genes. We ran BUSCO v5.3.2 (Manni et al. 2021) in the genome mode with default parameters on the Actinopterygii dataset (Actinopterygii_odb10) that contains 3,640 orthologous genes. Second, we used Merquy v1.3 (Rhie et al. 2020) to estimate base-level accuracy, expressed as the phred-scaled Quality Value (QV) and assembly completeness by comparing the k-mers in the assembly with those observed in the HiFi reads. Third, we used the blobtools pipeline v1.1.1 (Laetsch and Blaxter 2017) by generating taxon-annotated Guanine-Cytosine (GC) content-coverage plots (known as “BlobPlots”). Each scaffold was annotated with taxonomy information based on blastn (Basic Local Alignment Search Tool) v2.12.0 (Camacho et al. 2009) searches against the National Center for Biotechnology Information nucleotide database with the options “-outfmt ‘6 qseqid staxids bitscore std sscinames sskindoms stitle’ -culling_limit 5 -evaluate 1e-25.” To calculate average coverage per scaffold, we aligned the PacBio HiFi raw reads to the assembly using bwa-mem version 0.7.17 (Burrow-Wheeler Aligner) (Li et al. 2009) with default parameters. The resulting BAM file was sorted with samtools v1.11 (Li et al. 2009) and passed to blobtools along with the table of blastn results.

Mitochondrial genome assembly

The mitochondrial genome was assembled and annotated with MitoHiFi v3.2.1 (Uliano-Silva et al. 2023), using a genome from *A. naccarii* (MK078265.1; Cheng et al. 2021b) as a reference.

Genome annotation

To identify and annotate repetitive elements, we first generated a de novo repeat library using the Extensive de novo TE Annotator v1.9.9 (Ou et al. 2019). Subsequently, we refined the library using DeepTE v1.0 (Yan et al. 2020), which employs convolutional neural networks to classify unknown elements at the order and superfamily levels. Then, we used RepeatMasker v4.1.2 (Smit et al. 2021; <http://www.repeatmasker.org>) with the final library to mask the genome and parsed the RepeatMasker output file with RM_TRIPS script (https://github.com/clbutler/RM_TRIPS). Transposable element landscapes were generated using the RepeatMasker script calcDivergenceFromAlign.pl.

Quality control and trimming for adapters and low-quality bases (quality score <20) of the RNA-seq raw reads were performed using fastqc v0.11.8 (Andrews 2010) and fastp v0.19.5 (Chen et al. 2018), respectively. After trimming, reads shorter than 75 nucleotides were discarded. High-quality reads were then mapped to the soft-masked assembly with HISAT2 v2.2.1 (Kim et al. 2019) and sorted with samtools v1.11 (Li et al. 2009). All the BAM files were filtered to remove invalid splice junctions with Portcullis v1.2.4 (Mapleson et al. 2018). Filtered RNA-seq alignments were passed to Braker v3.0 (Gabriel et al. 2024), together with protein sequences of three closely related species, *Acipenser oxyrinchus* (GCA_030684275.1), *A. ruthenus* (GCF_902713425.1), and *Polyodon spathula* (GCF_017654505.1). The Braker gene prediction pipeline was run with the options “-softmasking.” This pipeline uses StringTie (Pertea et al. 2015) to assemble the RNA-seq reads followed by rounds of GeneMark and AUGUSTUS training and gene prediction (Stanke and Waack 2003). Gene sets were combined with TSEBRA v1.0.3 (Gabriel et al. 2021). The completeness of the final gene set was checked with BUSCO v5.3.2 (Manni et al. 2021) against the Actinopterygii dataset (Actinopterygii_odb10), using the longest transcript of each gene as the representative transcript.

Synteny analysis

Syntenic blocks of genes in *A. naccarii* were identified across the four haploid assemblies and compared with *A. ruthenus* (GCF_902713425.1), *A. oxyrinchus* (GCA_030684275.1), and *Huso huso* (GCA_036884735.1), using GENESPACE v1.2.3 (Lovell et al. 2022). As outgroup, the genome of *P. spathula* (GCF_017654505.1) was used. Riparian plots were obtained using the synteny calculation blocks obtained from GENESPACE.

Results and discussion

Genome size and ploidy estimation

GenomeScope reported a haploid genome size of ~1.6 Gbp for *A. naccarii*, which is 11 % to 20% smaller than those reported for other sturgeon species, such as *Acipenser gueldenstaedtii* (1.9 Gbp), *A. ruthenus* (1.8 Gbp), *Acipenser baerii* (2.0 Gbp), and *Acipenser sinensis* (1.9 Gbp) (Zhou et al. 2011; Du et al. 2020; Wang et al. 2024) (Fig. 1b). The reported homozygosity percentage was 97.9% (aaaa), and the most frequent types of heterozygous alleles were aaab at 1.06% and aabb at 1% (Fig. 1c). Typically, autotetraploid genomes have aaab > aabb, whereas allotetraploid genomes have the opposite (aaab < aabb). In the case of the Adriatic sturgeon, there was not a substantial difference between aaab and aabb, which does not allow to assess if the two events of whole-genome duplication in this species (Vasil'eva 1999; Fontana et al. 2008) occurred through auto- or allopolyploidy duplications. The Smudgeplot analysis identified aabb (35%) and aaab (12%) genotype proportions, suggesting a predominantly tetraploid genome structure in the Adriatic sturgeon. The remaining 53% of the genome was fractionated in higher ploidy levels that could not be clearly resolved (Supplementary Fig. S1). This pattern may reflect technical limitations of the tool due to the high repetitiveness of the genome (54.6%), a level well above the condition under which Smudgeplot reliably infers ploidy in autotetraploid datasets

Table 2 Genome assembly statistics for each assembled haploid copy of *Acipenser naccarii* and *Acipenser ruthenus* (GCF_902713425.1).

	hap1	hap2	hap3	hap4	<i>Acipenser ruthenus</i>
Total length (Gb)	1.94	1.74	1.65	1.84	1.89
Number of scaffolds	2,083	2,370	1,881	3,438	1,730
Scaffold N50 (Mb)	39.6	37.89	43.81	42.13	42.7
Scaffold L50	14	13	11	12	12
Scaffold N90 (Mb)	0.485	0.405	0.684	0.313	9.613
Scaffold L90	276	271	101	280	47
Number of contigs	4,788	5,872	5,744	8,291	2,163
Contigs N50 (Mb)	1.024	0.806	0.756	0.700	8.8
Contigs L50	512	555	581	653	65

Table 3 BUSCO scores for each assembled haploid copy and all joint assemblies of *Acipenser naccarii* and *Acipenser ruthenus* (GCF_902713425.1).

	hap1	hap2	hap3	hap4	Joint assemblies	<i>Acipenser ruthenus</i>
Complete BUSCO (%)	88.9	88.7	87.4	88.1	94.1	98
Complete and single-copy BUSCOs (%)	55.7	58.3	57.2	51.5	2.3	27.4
Complete and duplicated BUSCOs (%)	33.2	30.4	30.2	36.6	91.8	70.6
Fragmented BUSCOs (%)	2.3	2.3	2.2	2.6	1.9	0.2
Missing assembly BUSCOs (%)	8.8	9	10.4	9.3	4	1.8
k-mer completeness (%)	64.8	63.6	63.2	65.5	95.2	/
QV	57.9	57.5	56.9	56.6	57.2	49.83

(<38% repetitiveness; [Ranallo-Benavidez et al. 2020](#)). Under such complexity, k-mer based approaches may overestimate ploidy because the signal from repetitive k-mers exceeds that from heterozygous k-mers. However, considering the evolutionary history of the Adriatic sturgeon, which likely underwent two whole-genome duplications ([Fontana et al. 2001](#); [Biltueva et al. 2020](#)), it could be considered octoploid on an evolutionary scale. Therefore, the smudges above tetraploidy observed in [Supplementary Fig. S1](#) should be interpreted with caution as they might reflect remnants of the historical octoploid state. To further investigate the polyploid structure of the *A. naccarii* genome and possibly overcome the Smudgeplot's extremely high-coverage requirements, we employed the read mapping approach implemented in nQuire. nQuire was applied to each of the 60 chromosomes individually. Thirty-five chromosomes fit the tetraploid model, whereas five matched the diploid model ([Supplementary Fig. S2](#)). The remaining 20 chromosomes did not fit any model, likely due to a low presence of biallelic sites or potential octoploidy, as suggested by Smudgeplot. However, nQuire does not support an octoploid model. Despite differences in the estimated proportion of tetraploid regions (74.7% in nQuire vs. 47% in Smudgeplot), both methods indicated that the Adriatic sturgeon genome is predominantly tetraploid. These results provided a first insight into the complex ploidy levels present in the genome of this species.

Genome assembly

We generated 5.96 million HiFi reads and 357 million of Hi-C read pairs, corresponding to an approximate sequencing coverage of 59X and 67X, respectively. The assembly process resulted in

four haploid assemblies having a genome size between 1.65 and 1.94 Gbp, a N50 from 37.8 to 56.8 Mb, and L50 from 11 to 33 scaffolds ([Fig. 1c](#); [Table 2](#)). Some differences among haplotypes are observed, which likely reflect both the complexity of the species' ploidy and the presence of additional haplotigs in the larger assemblies. The first assembled haploid copy (hereafter hap1) was the most contiguous assembly characterized by 2,083 scaffolds, a total length of 1.94 Gbp, a contig N50 of 1.024 Mb, a scaffold N50 of 39.6Mbp, and a scaffold L90 of 276. Assembly statistics at each step of RagTag-based correction and scaffolding of hap1 are summarized in [Supplementary Table S1](#). The assembly size was in agreement with the genome size estimated from HiFi reads and it was close to the 1.8 Gbp reported for *A. ruthenus* ([Du et al. 2020](#)). The 80% of the assembly was contained in the first 60 scaffolds, suggesting a close proximity to the chromosome scale for several scaffolds ([Fig. 1d](#)). Compared with the diploid chromosome-scale *A. ruthenus* assembly, our tetraploid assembly showed similar contiguity metrics at the scaffold level (Scaffold N50 39.6 vs 42.7 Mb) but starting from a more fragmented contig set (Contig N50 1.02 vs 8.8 Mb) as expected due to challenges posed by the sorting of polyploid genomes ([Table 2](#)).

The hap1 assembly was moderately complete, reaching a BUSCO score of 88.9% according to the ctinopterygii_odb10 gene set and a k-mer completeness of 64.8% ([Table 3](#)). Both completeness statistics were significantly higher when computed jointly over the four assemblies, reaching a BUSCO score of 94.1% and a k-mer completeness of 95.2% ([Table 3](#)). The phred-scaled nucleotide error rate of the four assemblies was similar among them varying between 56.6 and 57.9, corresponding to one erroneous base every 450 and 620 Kbp, respectively. The estimated gene completeness was slightly lower than *A. ruthenus*

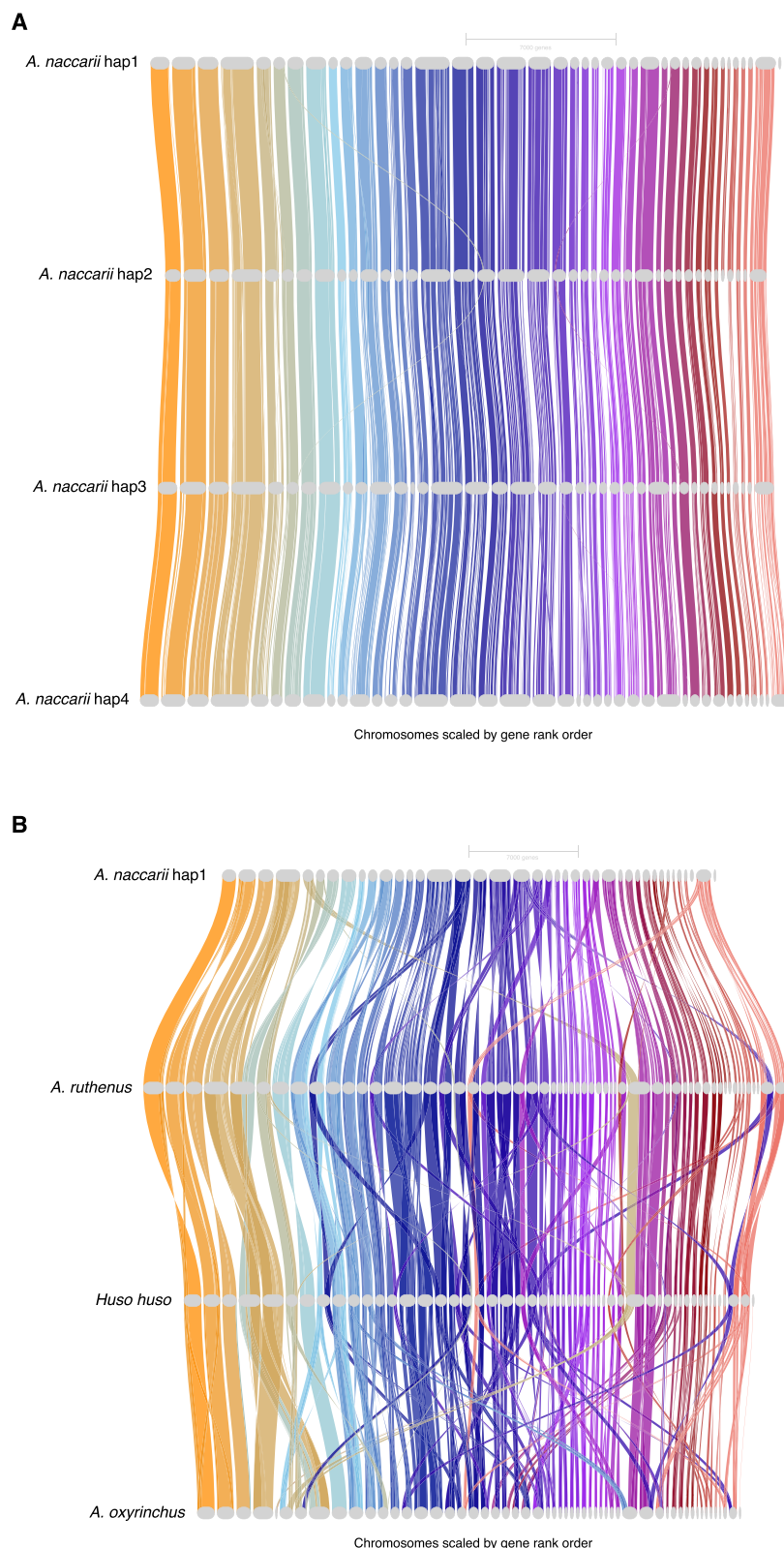


Fig. 2 Synteny among the four *Acipenser naccarii* haploid assemblies and compared with *Acipenser ruthenus*. GENESPACE generated synteny map among (bottom to top) the four genome assemblies of *A. naccarii* (a) and among the *A. naccarii* hap1, *A. ruthenus*, *Huso huso*, and *Acipenser oxyrinchus* (b). Chromosomes are shown as segments, and syntenic blocks between each pair of genomes are shown as connecting links. Only scaffolds with strong and unambiguous syntenic signal were included.

genome (BUSCO 94.1% vs 98%) but still in line with the high-quality genome assembly standards (Rhie et al. 2021) (Table 3). To further assess assembly quality, we performed contamination screening using BlobTools for each haplotype, which did not reveal any evidence of obvious contamination (Supplementary Figs S3–S6). In addition, we evaluated k-mer completeness using Merqury, both by considering all haplotypes jointly (Supplementary Fig. S7) and each haplotype separately (Supplementary Fig. S8). These analyses indicate a high level of k-mer completeness with little missing single-copy genome content, whereas the observed duplicated k-mers are consistent with the presence of haplotigs resulting from haplotype-resolved assembly, particularly in haplotype 4, which shows a higher number of contigs and an increased proportion of duplicated BUSCO genes.

We applied nQuire to the first 60 scaffolds of the *A. naccarii* hap1 assembly individually. Thirty-four scaffolds fit the tetraploid model, whereas 13 matched the diploid model and two the triploid model (Supplementary Fig. S9). The remaining scaffolds did not conform to any tested model, potentially reflecting genomic regions with higher ploidy or issues related to model inference, as suggested by Smudgeplot analyses.

The mitochondrial genome was 17,033 bp with successful annotations of all 37 genes found in the reference mitochondrial genome (Supplementary Fig. S10).

Genome annotation

We generated 16 Gb of strand-specific RNA-seq data to support genome annotation. A range of 27,050 to 29,814 protein-coding genes (36,458 to 40,061 transcripts) were predicted across the four haploid *A. naccarii* genome assemblies (Supplementary Table S2). The four annotations captured between 90.7% and 92.2% of the BUSCO Actinopterygii gene set as complete copies (Supplementary Table S2). The identification of repetitive elements in the four haplotypes revealed a repeat content ranging from 40.67% to 45.96% (Supplementary Tables S3–S6). The major class of repetitive elements was constituted by DNA transposons, with the DNA_hAT family representing the most abundant component, accounting for 11.56% to 14.58% of the genome across the haplotypes (Supplementary Fig. S11). Given their abundance, DNA_hAT elements may have contributed to ancestral genome duplications and structural evolution in sturgeons, representing an interesting focus for future study.

Synteny

The comparison of the four haploid *A. naccarii* genome assemblies revealed a high degree of synteny (Fig. 2a), confirming the consistency of all copies within our assembly (for details of the order across haplotypes, see Supplementary Table S7).

Additionally, a comparative analysis between the genomes of *A. naccarii* (hap1) and *A. ruthenus* demonstrated a high level of collinearity and synteny, with the majority of chromosomes showing no signs of major structural rearrangements (Fig. 2b). Although a few rearrangements were detected, their presence is expected given that these two species have been evolving separately for approximately 11.2 million years (Kumar et al. 2017). Such variations likely reflect natural evolutionary divergence rather than errors in genome assembly. It should be noted that the apparent

difference in size between the assemblies in the synteny plots reflects that only orthologous genes shared across genomes are shown.

When comparing *A. naccarii* to *H. huso* and *A. oxyrinchus*, the extent of chromosomal rearrangements increased, particularly in *A. oxyrinchus*, which diverged from the other species approximately 171 million years ago (Peng et al. 2007) (Fig. 2b). This greater divergence is consistent with the accumulation of structural variations over the evolutionary timescale and highlights the deep phylogenetic split between *A. oxyrinchus* and the other sturgeon species. It is worth noting that the classification of *H. huso* in a genus different from *Acipenser* is not supported by molecular data and that this species is actually grouped together with *A. naccarii* within the cluster of Atlantic-origin species, whereas *A. oxyrinchus*, along with *Acipenser sturio*, forms the basal cluster of all sturgeon species. (Krieger et al. 2008).

Conclusions

We present the first high-quality genome assembly for the Adriatic sturgeon, *A. naccarii*, a critically endangered, tetraploid species. Despite the challenges of assembling polyploid genomes, our phased assembly achieved high levels of completeness and contiguity, comparable to those of other recently assembled sturgeon species. We found a high degree of synteny among the haplotypes and with the closely related *A. ruthenus*. In contrast, deeper genomic divergence was observed in *A. oxyrinchus*, reflecting long-term structural rearrangements. Our analysis revealed a predominantly tetraploid genome with complex patterns of ploidy variation, offering new insights into genome evolution in sturgeons following whole-genome duplication events. The integration of genome assembly and annotation with transcriptomic data presents a valuable opportunity to investigate the process of functional diploidization and its impact on gene regulation, expression patterns, and evolutionary trajectories. In particular, the identification of heterozygous regions and allele-specific expression patterns will enable detailed studies of trait segregation and inheritance. This approach not only aids in understanding the evolutionary dynamics of tetraploidy but also validates candidate genes associated with diploidization, providing critical insights into the biology and adaptive potential of sturgeons. This genomic resource also has significant practical implications. It facilitates the construction of specific genetic markers, such as MIPs (Boscari et al. 2024), in order to support the development of robust conservation strategies, including population genetic studies, and the identification of breeding lines with optimal genetic diversity. In conclusion, this genome enhances our understanding of polyploid evolution in vertebrates and provides a critical resource for the conservation of this ancient and endangered species.

Acknowledgments

We deeply thank Camil Jaron for his help with GenomeScope and Smudgeplot analyses.

Author contributions

Roberto Biello (Conceptualization, Investigation, Writing—original draft, Writing—review & editing), Annalisa Scapolatiello (Investigation, Writing—original draft, Writing—review & editing),

Sebastiano Fava (Investigation, Writing—original draft, Writing—review & editing), Victor Hugo Muñoz Mora (Investigation, Writing—original draft, Writing—review & editing), Stefano Dalle Palle (Methodology, Writing—review & editing), Patrícia Santos (Investigation, Writing—review & editing), Alessio Iannucci (Methodology, Writing—review & editing), Tatiana Tilley (Methodology, Writing—review & editing), Nivesh Jain (Methodology, Writing—review & editing), Jennifer Balacco (Methodology, Writing—review & editing), Brian O'Toole (Methodology, Writing—review & editing), Giulio Formenti (Writing—review & editing), Erich Jarvis (Writing—review & editing), Claudio Ciofi (Writing—review & editing), Emiliano Trucchi (Conceptualization, Investigation, Writing—review & editing), Leonardo Congiu (Conceptualization, Resources, Writing—original draft, Writing—review & editing), Giorgio Bertorelle (Conceptualization, Funding acquisition, Writing—review & editing), and Andrea Benazzo (Conceptualization, Investigation, Writing—original draft, Writing—review & editing)

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Funding

This study was supported by the University of Ferrara (Italy) and funded by the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) 2017 grant 201794ZXTL. LC was partially funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4—Call for tender No. 3138 of 16 December 2021, rectified by Decree n. 3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union—NextGenerationEU; Award Number: Project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP C93C22002810006, Project title “National Biodiversity Future Center - NBFC.” Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission, nor can they be held responsible for them.

Conflict of interest

None declared.

Data availability

Raw sequencing data and genome assemblies are available under NCBI BioProjects PRJNA1259054, PRJNA1267827, PRJNA1267829, PRJNA1267830, PRJNA1267831. Genome assemblies, genome annotations, and RepeatMasker outputs are available on Zenodo ([10.5281/zenodo.15705491](https://doi.org/10.5281/zenodo.15705491)). The mitochondrial genome sequence is available in GenBank under accession number [PX733385.1](https://doi.org/10.5281/zenodo.15705491) Scripts for removing multiple allelic haplotigs from genome assemblies are available at https://github.com/POPGG-UniFe/Sturgeon_genome.

References

- Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 2022;23:258. <https://doi.org/10.1186/s13059-022-02823-7>
- Andrews S. FastQC: A quality control tool for high throughput sequence data. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Arlati G, Poliakova L. Restoration of Adriatic sturgeon (*Acipenser naccarii*) in Italy: situation and perspectives. In: Carmona R, García-Gallego M, Ruiz-Rejón M, eds. *Biology, conservation and sustainable development of sturgeons*. Vol. 29. Springer; 2009:237–238. https://doi.org/10.1007/978-1-4020-8437-9_14.
- Billard R, Lecointre G. Biology and conservation of sturgeon and paddlefish. *Rev Fish Biol Fish.* 2000;10:355–392. <https://doi.org/10.1023/A:1012231526151>
- Biltueva LS, Prokopov DY, Romanenko SA, Interesova EA, Schartl M, Trifonov VA. Chromosome distribution of highly conserved tandemly arranged repetitive DNAs in the Siberian sturgeon (*Acipenser baerii*). *Genes.* 2020;11:1375. <https://doi.org/10.3390/genes11111375>
- Birstein VJ, Hanner R, DeSalle R. Phylogeny of the Acipenseriformes: cytogenetic and molecular approaches. *Environmental Biology of Fishes.* 1997;48:127–55. <https://doi.org/10.1023/a:1007366100353>
- Boscari E, Palle SD, Vitulo N, Scapolatiello A, Schiavon L, Cariani A, Papetti C, Zane L, Marino IA, Congiu L. MIPs: multi-locus intron polymorphisms in species identification and population genomics. *Scientific Reports.* 2024;14:17870. <https://doi.org/10.1038/s41598-024-68065-8>
- Boscari E, Vidotto M, Martini D, Papetti C, Ogden R, Congiu L. Microsatellites from the genome and the transcriptome of the tetraploid Adriatic sturgeon, *Acipenser naccarii* (Bonaparte, 1836), and cross-species applicability to the diploid beluga sturgeon, *Huso huso* (Linnaeus, 1758). *J Appl Ichthyol.* 2015;31:977–983. <https://doi.org/10.1111/jai.12906>
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC bioinformatics.* 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421>
- Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* 2021a;18:170–175. <https://doi.org/10.1038/s41592-020-01056-5>
- Cheng P-L, Yu D, Liu H-Z, Du H, Wei Q-W. Molecular phylogeny of Acipenseriformes based on complete mitochondrial genome sequence. *Acta Hydrobiol Sin.* 2021b;45:487–494. <https://doi.org/10.7541/2021.2020.050>
- Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, Li H. Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol.* 2022;40:1332–1335. <https://doi.org/10.1038/s41587-022-01261-x>
- Congiu L, Pujolar JM, Forlani A, Cenadelli S, Dupanloup I, Barbisan F, Galli A, Fontana F. Managing polyploidy in ex situ con-

- servation genetics: the case of the critically endangered Adriatic sturgeon (*Acipenser naccarii*). *PLoS One*. 2011;6:e18249. <https://doi.org/10.1371/journal.pone.0018249>
- Dalle Palle S, Boscari E, Bordignon SG, Muñoz-Mora VH, Bertorelle G, Congiu L. Different chromosome segregation patterns coexist in the tetraploid Adriatic sturgeon *Acipenser naccarii*. *Diversity*. 2022;14:14–745. <https://doi.org/10.3390/d14090745>
- Du K, Stöck M, Kneitz S, Klopp C, Woltering JM, Adolphi MC, Feron R, Prokopov D, Makunin A, Kichigin I, et al. The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat Ecol Evol*. 2020;4:841–852. <https://doi.org/10.1038/s41559-020-1166-x>
- Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for hi-C contact maps with unlimited zoom. *Cell Syst*. 2016;3:99–101. <https://doi.org/10.1016/j.cels.2015.07.012>
- Fontana F, Lanfredi M, Chicca M, Congiu L, Tagliavini J, Rossi R. Fluorescent in situ hybridization with rDNA probes on chromosomes of *Acipenser ruthenus* and *Acipenser naccarii* (Osteichthyes, Acipenseriformes). *Genome*. 1999;42:1008–1012. <https://doi.org/10.1139/g99-030>
- Fontana F, Tagliavini J, Congiu L. Sturgeon genetics and cytogenetics: recent advancements and perspectives. *Genetica*. 2001;111:359–373. <https://doi.org/10.1023/a:1013711919443>
- Fontana F, Congiu L, Mudrak VA, Quattro JM, Smith TI, Ware K, Doroshov SI. Evidence of hexaploid karyotype in shortnose sturgeon. *Genome*. 2008;51:113–119. <https://doi.org/10.1139/G07-112>
- Gabriel L, Hoff KJ, Bruna T, Borodovsky M, Stanke M. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics*. 2021;22:566. <https://doi.org/10.1186/s12859-021-04482-0>
- Gabriel L, Bruna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res*. 2024;34:769–777. <https://doi.org/10.1101/gr.278090.123>
- Havelka M, Hulák M, Bailie DA, Prodöhl PA, Flajšhans M. Extensive genome duplications in sturgeons: new evidence from microsatellite data. *J Appl Ichthyol*. 2013;29:704–708. <https://doi.org/10.1111/jai.12224>
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37:907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kokot M, Długosz M, Deorowicz S. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics*. 2017;33:2759–2761. <https://doi.org/10.1093/bioinformatics/btx304>
- Krieger J, Hett AK, Fuerst PA, Artyukhinand E, Ludwig A. The molecular phylogeny of the order Acipenseriformes revisited. *J Appl Ichthyol*. 2008;24:36–45. <https://doi.org/10.1111/j.1439-0426.2008.01088.x>
- Kuhl H, Du K, Scharlt M, Kalous L, Stöck M, Lamatsch DK. Equilibrated evolution of the mixed auto-/allopolyploid haplotype-resolved genome of the invasive hexaploid Prussian carp. *Nat Commun*. 2022;13:4092. <https://doi.org/10.1038/s41467-022-31515-w>
- Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 2017;34:1812–1819. <https://doi.org/10.1093/molbev/msx116>
- Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies. *F1000Research*. 2017;6:1287. <https://doi.org/10.12688/f1000research.12232.1>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*. 2022;11:e78526. <https://doi.org/10.7554/eLife.78526>
- Ludwig A, Belfiore NM, Pitra C, Svirsky V, Jenneckens I. Genome duplication events and functional reduction of ploidy levels in sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). *Genetics*. 2001;158:1203–1215. <https://doi.org/10.1093/genetics/158.3.1203>
- Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. *Curr Protoc*. 2021;1:e323. <https://doi.org/10.1002/cpz1.323>
- Mapleson D, Venturini L, Kaithakottil G, Swarbreck D. Efficient and accurate detection of splice junctions from RNA-seq with portcullis. *GigaScience*. 2018;7:giy131. <https://doi.org/10.1093/gigascience/giy131>
- Open2C, Abdennur N, Fudenberg G, Flyamer IM, Galitsyna AA, Goloborodko A, Imakaev M, Venev SV. Pairtools: from sequencing data to chromosome contacts. *PLoS Comput Biol*. 2024;20:e1012164. <https://doi.org/10.1371/journal.pcbi.1012164>
- Ou S, Su W, Liao Y, Chougule K, Agda JR, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol*. 2019;20:275–218. <https://doi.org/10.1186/s13059-019-1905-y>
- Peng Z, Ludwig A, Wang D, Diogo R, Wei Q, He S. Age and biogeography of major clades in sturgeons and paddlefishes (Pisces: Acipenseriformes). *Mol Phylogenet Evol*. 2007; 42:854–862. <https://doi.org/10.1016/j.ympev.2006.09.008>
- Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33:290–295. <https://doi.org/10.1038/nbt.3122>
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11:1432. <https://doi.org/10.1038/s41467-020-14998-3>
- Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;21:245. <https://doi.org/10.1186/s13059-020-02134-9>
- Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021;592:737–746. <https://doi.org/10.1038/s41586-021-03451-0>
- Sim SB, Corpuz RL, Simmonds TJ, Geib SM. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics*. 2022;23:157. <https://doi.org/10.1186/s12864-022-08375-1>

- Smit AFA, Hubley R, Green P. RepeatMasker Version 4.1.2.2021. <http://www.repeatmasker.org>
- Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*. 2003;19:ii215–ii225. <https://doi.org/10.1093/bioinformatics/btg1080>
- Stift M, Berenos C, Kuperus P, van Tienderen PH. Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to Rorippa (yellow cress) microsatellite data. *Genetics*. 2008;179:2113–2123. <https://doi.org/10.1534/genetics.107.085027>
- Sulak KJ, Randall M. Understanding sturgeon life history: enigmas, myths, and insights from scientific studies. *J Appl Ichthyol*. 2002;18:519–528. <https://doi.org/10.1046/j.1439-0426.2002.00413.x>
- Uliano-Silva M, Ferreira JG, Krashenninnikova K, Formenti G, Abueg L, Torrance J, Myers EW, Durbin R, Blaxter M, McCarthy SA. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC bioinformatics*. 2023;24:288. <https://doi.org/10.1186/s12859-023-05385-y>
- Vasil'eva ED. Some morphological characteristics of acipenserid fishes: considerations of their variability and utility in taxonomy. *J Appl Ichthyol*. 1999;15:32–34. <https://doi.org/10.1111/j.1439-0426.1999.tb00201.x>
- Wang B, Wu B, Liu X, Hu Y, Ming Y, Bai M, Li Y, Zhang Y, Zhang H, Shi Q, et al. Whole-genome sequencing reveals autooctoploidy in Chinese sturgeon and its evolutionary trajectories. *Genomics, Proteomics & Bioinformatics*. 2024;22:qzad002. <https://doi.org/10.1093/gpbjnl/qzad002>
- Weiß CL, Pais M, Cano LM, Kamoun S, Burbano HA. nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics*. 2018;19:1–8. <https://doi.org/10.1186/s12859-018-2128-z>
- Williot P, Arlati G, Chebanov M, Gulyas T, Kasimov R, Kirschbaum F, Patriche N, Pavlovskaya LP, Poliakova L, Pourkazemi M, et al. Status and management of Eurasian sturgeon: an overview. *Int Rev Hydrobiol*. 2002;87:483–506. [https://doi.org/10.1002/1522-2632\(200211\)87:5/6<483::AID-IROH483>3.0.CO;2-K](https://doi.org/10.1002/1522-2632(200211)87:5/6<483::AID-IROH483>3.0.CO;2-K)
- Yan H, Bombarely A, Li S. DeepTE: a computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics*. 2020;36:4269–4275. <https://doi.org/10.1093/bioinformatics/btaa519>
- Zhang H, Song L, Wang X, Cheng H, Wang C, Meyer CA, Liu T, Tang M, Aluru S, Yue F, et al. Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat Commun*. 2021;12:6566. <https://doi.org/10.1038/s41467-021-26865-w>
- Zhou H, Fujimoto T, Adachi S, Yamaha E, Arai K. Genome size variation estimated by flow cytometry in *Acipenser mikadoi*, *Huso dauricus* in relation to other species of Acipenseriformes. *J Appl Ichthyol*. 2011;27:484–491. <https://doi.org/10.1111/j.1439-0426.2010.01648.x>
- Zhou C, McCarthy SA, Durbin R. YaHS: yet another hi-C scaffolding tool. *Bioinformatics*. 2023;39:btac808. <https://doi.org/10.1093/bioinformatics/btac808>