

Whole Genome Sequencing Reveals Autooctoploidy in the Chinese Sturgeon and its Evolutionary Trajectories

Binzhong Wang^{1, #}, Bin Wu^{2, #}, Xueqing Liu^{1, #}, Yacheng Hu^{1, #}, Yao Ming^{2, #}, Mingzhou Bai^{2, #}, Juanjuan Liu¹, Kan Xiao¹, Qingkai Zeng¹, Jing Yang¹, Hongqi Wang², Baifu Guo¹, Chun Tan¹, Zixuan Hu¹, Xun Zhao¹, Yanhong Li², Zhen Yue², Junpu Mei², Wei Jiang¹, Yuanjin Yang¹, Zhiyuan Li¹, Yong Gao¹, Lei Chen^{1, *}, Jianbo Jian^{2, *}, Hejun Du^{1, #, *}

1. Chinese Sturgeon Research Institute, China Three Gorges Corporation, Yichang, Hubei, 443100, China

2. BGI Genomics, BGI-Shenzhen, Shenzhen, 518083, China

Background



Sturgeon caviar

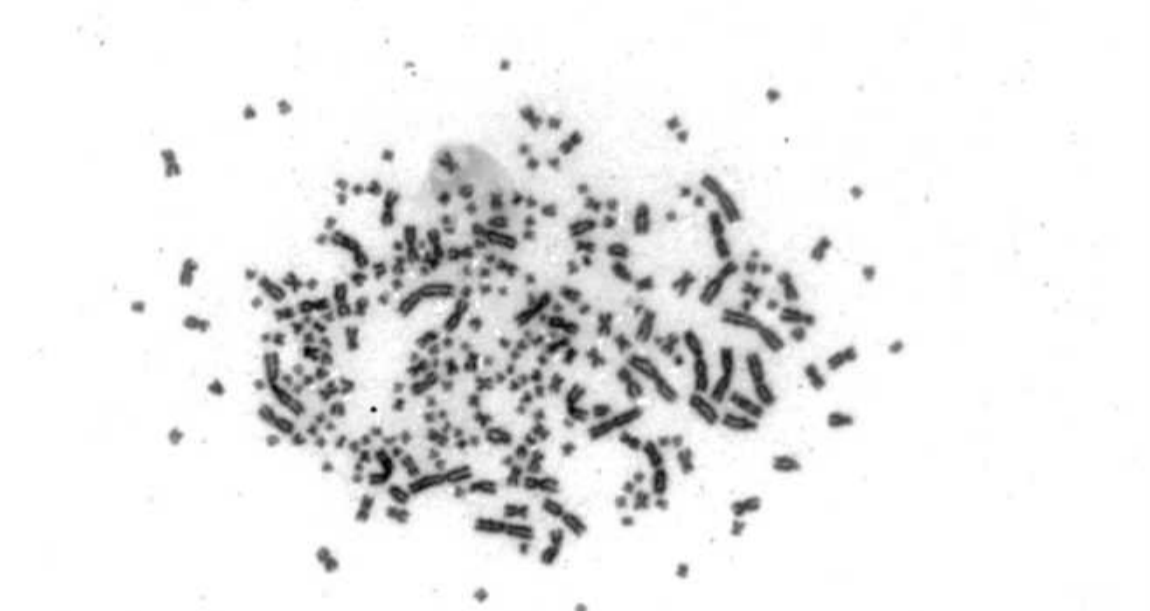


Artificial propagation and artificial releasing

- Acipenseriformes have attracted considerable attention for their valuable caviar and delicious meat, ancient species, wide distribution in the Northern Hemisphere, and threatened or endangered status.
- The whole genome assembly and ploidy evolution analysis is a still challenging task for the Acipenseriformes species because of the large genome contents (3.2-14.2 pg), multiple ploidy compositions (4n, 8n, and 12n), multiple karyotypes (Group A: ~120 chromosomes; Group B: ~240 chromosomes; and Group C: ~360 chromosomes) and nearly half of the microchromosomes.
- Acipenseriformes are thought to have undergone a delayed rediploidization, in which a species radiates an extensive time after a WGD (i.e. a timescale on the order of millions of years), resulting in complex ploidies. However, the ploidy compositions of most Acipenseriformes species have been challenging to clarify. The debates regarding ploidy, for example, whether Groups A and B are diploid and tetraploid or are instead tetraploid and octoploid, or even are paleotetraploidy versus modern/functional diploidy in the case of Group A, have lasted for half a century.
- Chinese sturgeon (*Acipenser sinensis*) is a critically endangered large fish in China. It has a complex ploidy composition of ~264 chromosomes and is a representative of Group B (8n, ~240 chromosomes) containing the most abundant Acipenseriformes species. Currently, the complex polyploid composition of the Chinese sturgeon is unknown, as well as the process of its formation. There is also controversy over the timing of whole genome duplication (WGD).



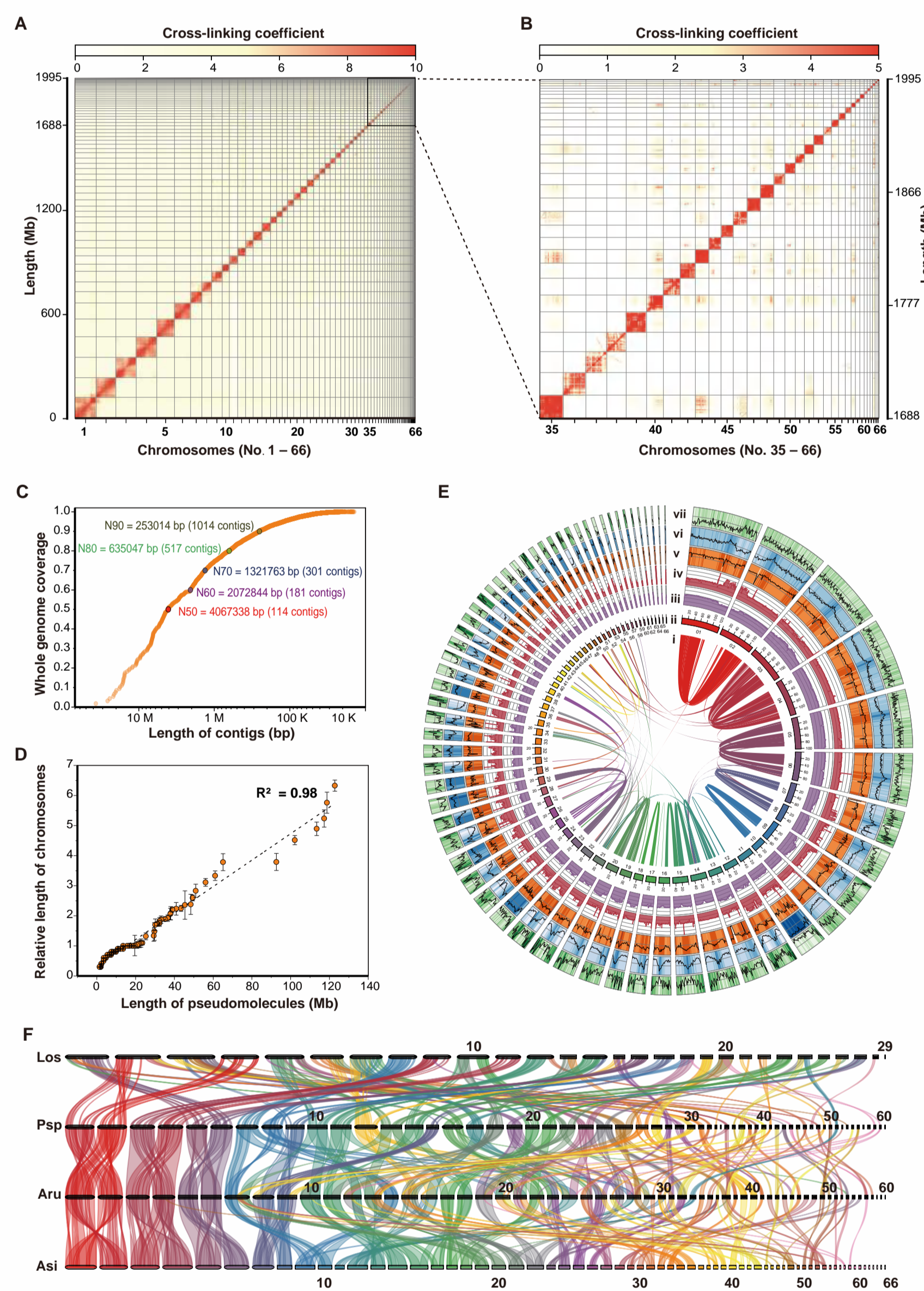
Chinese sturgeon



Karyotype

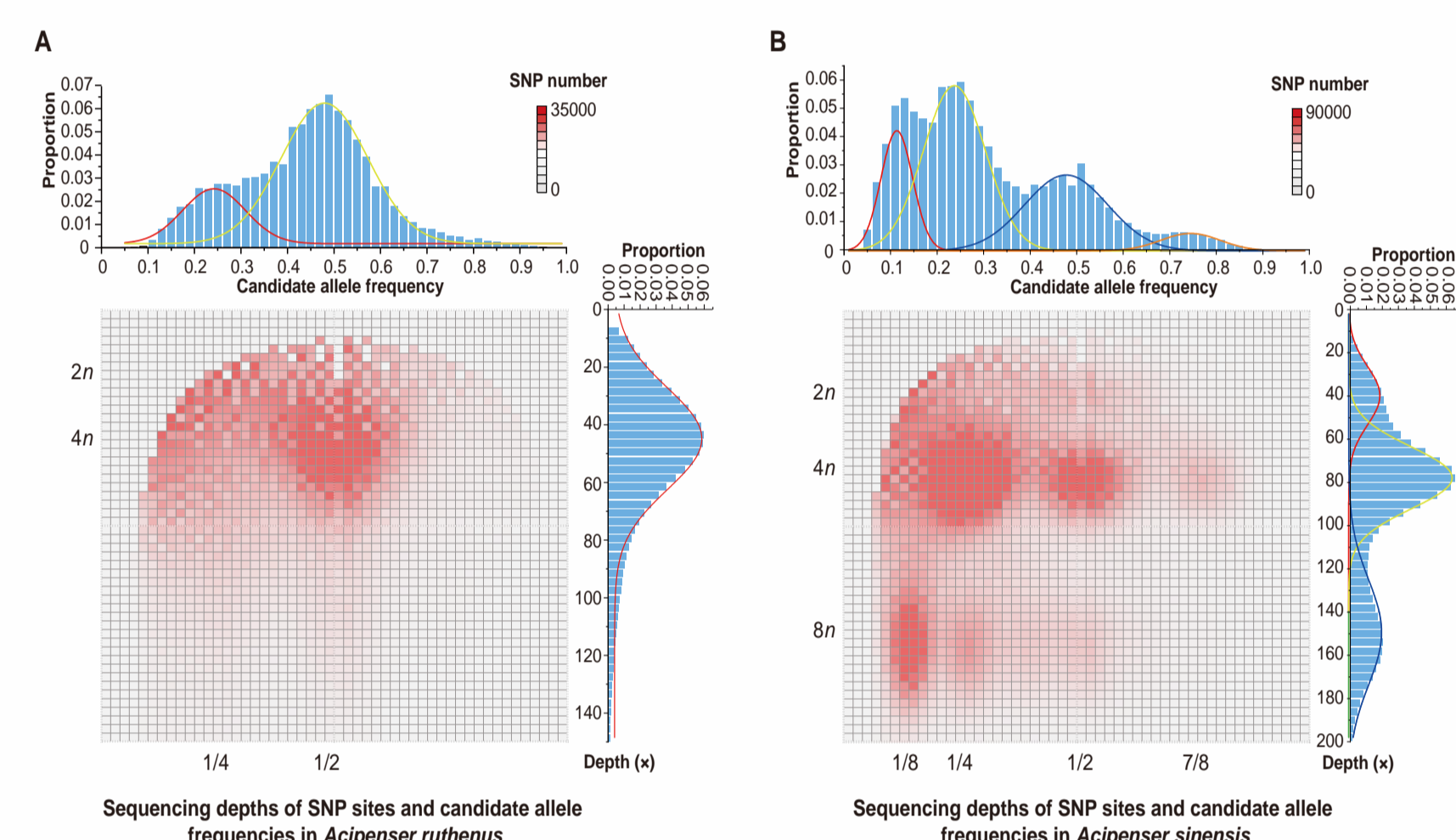
Results

Genome sequencing and assemble

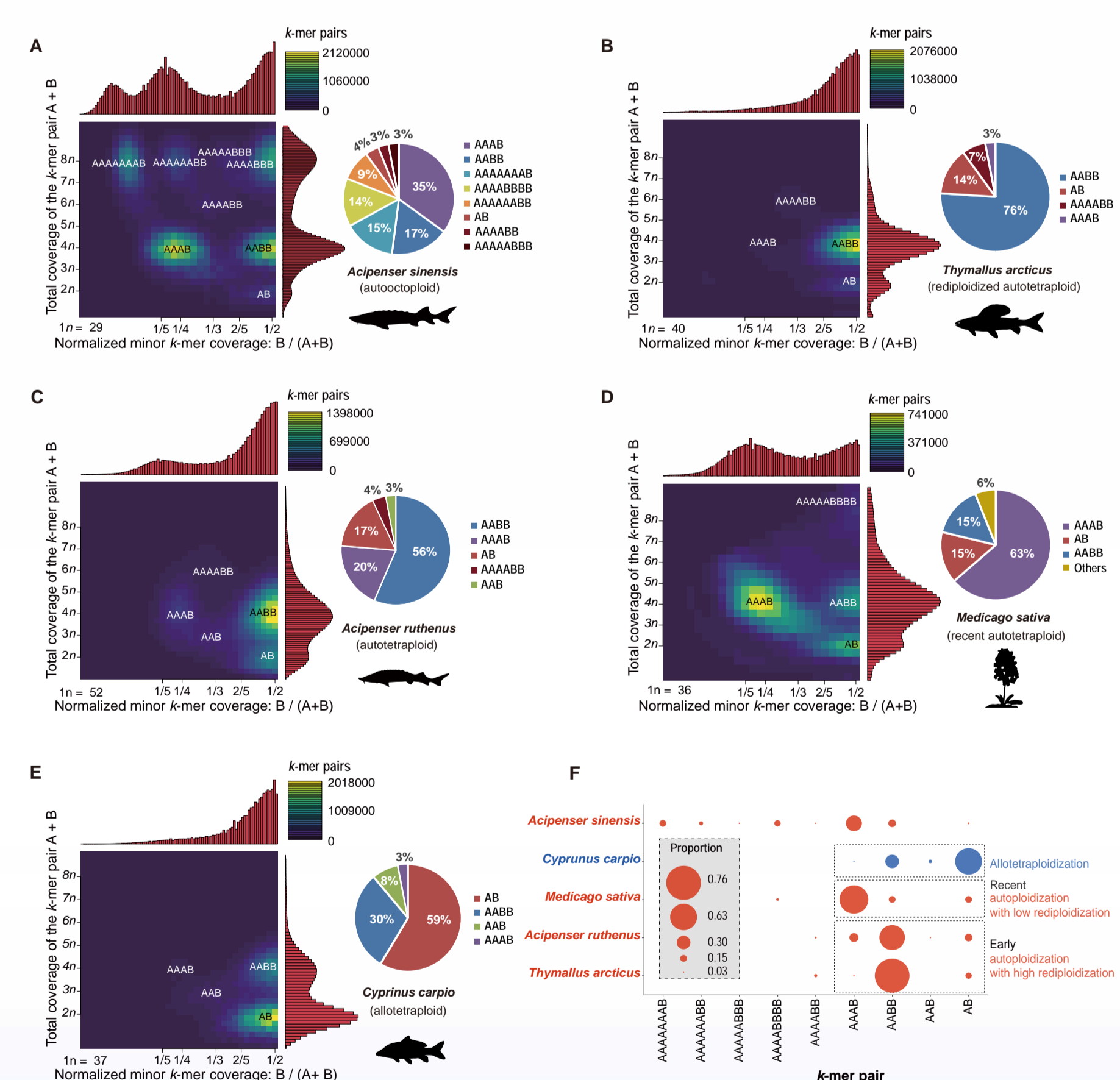


We obtained 421.58 Gb clean Illumina short-read data, 221.96 Gb clean PacBio long-reads data, and 172.87 Gb clean Hi-C data. Illumina and PacBio reads were assembled into the initial contigs of ~1.99 Gb with an N50 size of ~4.07 Mb. Clean Hi-C reads were applied to anchor contigs into 66 scaffolds corresponding to 66 chromosomes of two monoploid genomes. The final genome assembly was 1.99 Gb with a scaffold N50 size of ~48.46 Mb, and 98.3% of assembled sequences were assigned to chromosomes. The completeness of genome assembly was 95.6% based on BUSCO evaluation. We predicted a total of 36,837 protein-encoding genes and annotated 34,950 protein-coding genes.

Ploidy analysis

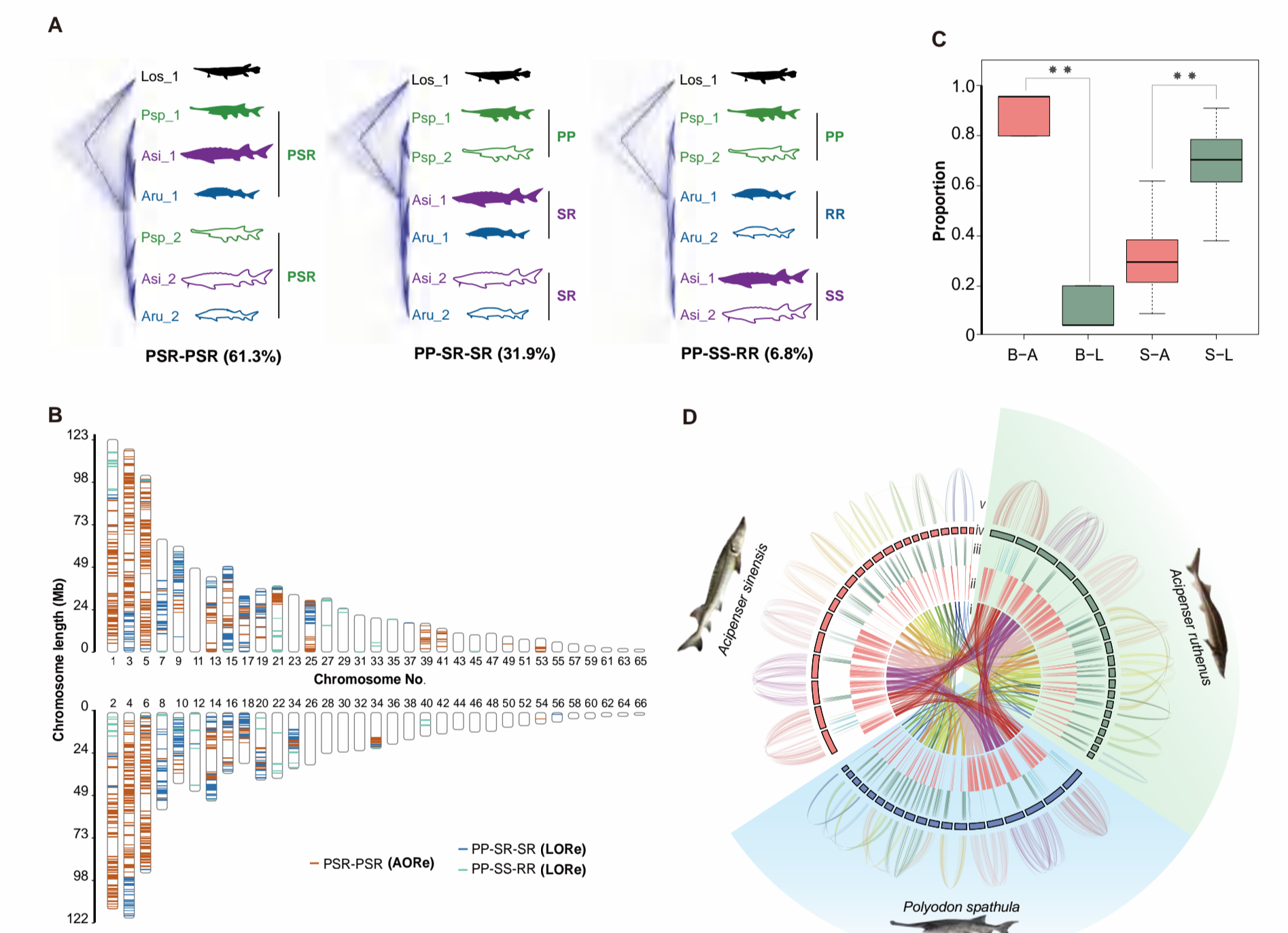


A. sinensis exhibited four peaks in the SNP frequency curve, whereas *A. ruthenus*, the tetraploid sturgeon, only exhibited two peaks. The main peak of ploidy in *A. sinensis* was 4n, pointing to 1/4 and 1/2 candidate allele frequency. We detected the first peak at the position of 1/8 candidate allele frequency in *A. sinensis*. This peak mostly pointed to 8n ploidy, suggesting that the eight monoploids have high similarity and revealed the octoploid features of *A. sinensis*.

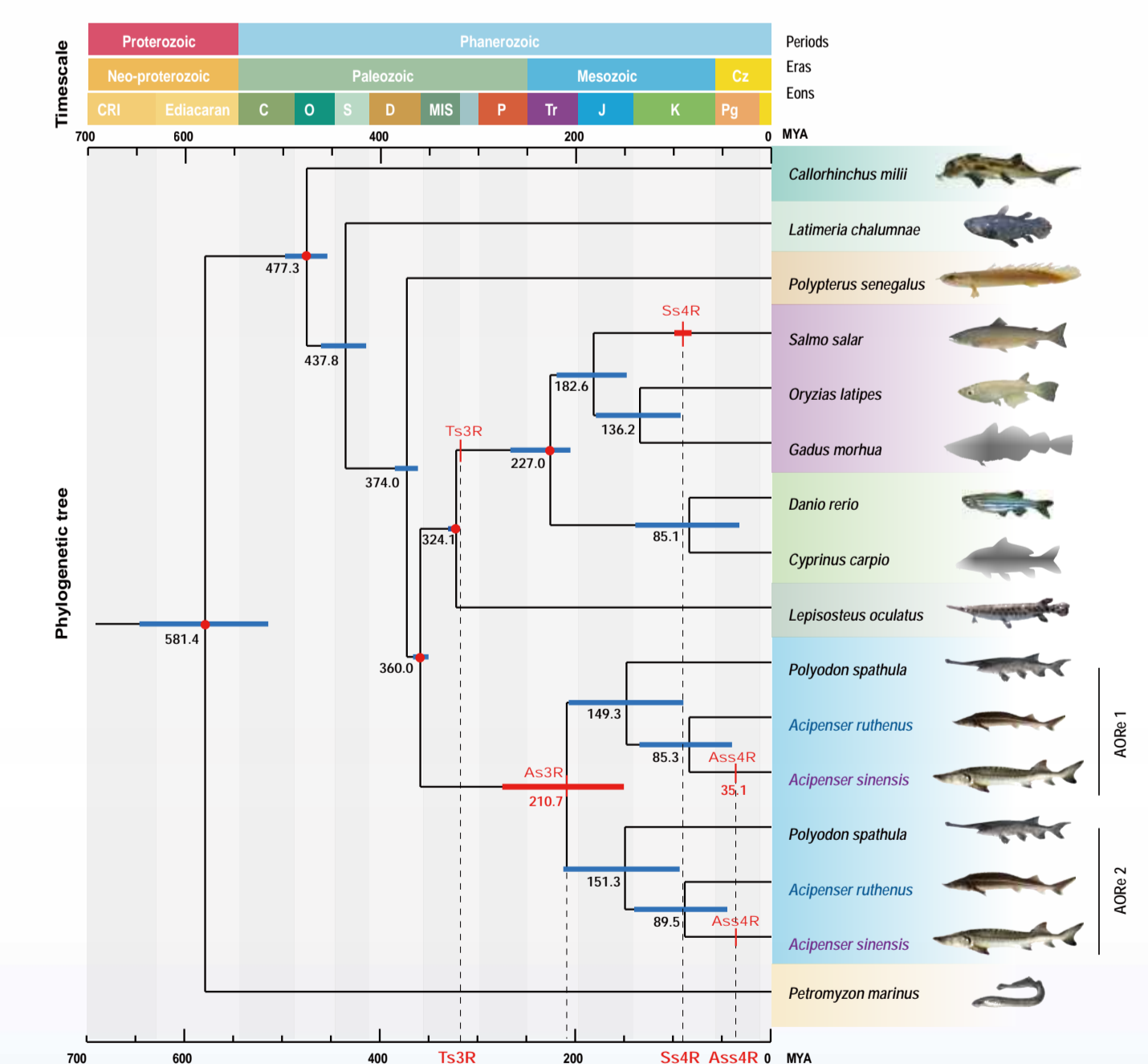


The proportion of autopolyploidization characteristics, including AAAAAAAB, AAAAAABB, AAAAABBB, and AAAB, is up to 62% in *A. sinensis*, suggesting that *A. sinensis* is an autooctoploid.

Whole genome duplication and divergence



We screened 1438 gene families with collinearity among the genomes of *A. sinensis* (S), *A. ruthenus* (R), *P. spatula* (P), and *Lepisosteus osseus* with the gene copy number of 2:2:2:1. We constructed 1,438 topologies using the screened gene families, and three types of representative topologies model were collected. The PSR-PSR type was the dominant topology and accounted for 61.3% of 736 screened gene families, indicating that the three species underwent a common duplication event. Lineage-specific Ohnolog Resolution (LORE) cannot accurately reflect WGD due to delayed differentiation of LORE, whereas, Ancestral Ohnolog Resolution (AORE) can better reflect the evolutionary trajectory of Acipenseriformes. AORE and LORE were screened.



The phylogenetic tree based on protein sequences of AORE in Acipenseriformes showed that the Acipenseriforme-specific common WGD (As3R) occurred 210.7 MYA. Divergence of paddlefish and sturgeon occurred ~150 MYA. *A. sinensis* and *A. ruthenus* diverged 89.5–85.3 MYA. *A. sinensis*-specific WGD (Ass4R) occurred ~35.12 MYA.

Conclusion



- Provided the first Chinese sturgeon reference genome with highly complex polyploidy. We sequenced the first octoploid sturgeon genome, and this is also the first octoploid animal genome to be sequenced to date. This will provide data support for the conservation of the Chinese sturgeon.
- Clarified the ploidy composition, allo- or auto-polyploidization of the Chinese sturgeon genome. We conclude that *A. sinensis* is an autooctoploid.
- Obtained a more reasonable WGD and divergence time and demonstrated the common evolution of Acipenseriformes. Acipenseriformes shared a common WGD event (As3R) dating back to 210.7 MYA before the divergence ~150 MYA of Acipenseridae and Polyodontidae, and *A. sinensis* underwent an additional lineage-specific WGD (Ass4R) around 35.12 MYA.

Acknowledgments

This work was supported by the Three Gorges Environmental Funds of China Three Gorges Corporation (XN270).