

Article

Highly Resolved Phylogenetic Relationships within Order Acipenseriformes According to Novel Nuclear Markers

Dehuai Luo^{1,†}, Yanping Li^{1,2,†}, Qingyuan Zhao¹, Lianpeng Zhao¹, Arne Ludwig^{3,*} 
and Zuogang Peng^{1,*} 

¹ Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education), Southwest University School of Life Sciences, Chongqing 400715, China; ldh2015@email.swu.edu.cn (D.L.); liyanping@genomics.cn (Y.L.); zqingyuanbiology@sina.com (Q.Z.); xndxsky_zlp@sina.com (L.Z.)

² Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, BGI, Shenzhen 518083, China

³ Department of Evolutionary Genetics, Leibniz-Institute for Zoo and Wildlife Research, 10315 Berlin, Germany

* Correspondence: pzg@swu.edu.cn (Z.P.); ludwig@izw-berlin.de (A.L.)

† Dehuai Luo and Yanping Li contributed equally to this work.

Received: 26 November 2018; Accepted: 2 January 2019; Published: 10 January 2019



Abstract: Order Acipenseriformes contains 27 extant species distributed across the northern hemisphere, including so-called “living fossil” species of garfish and sturgeons. Previous studies have focused on their mitochondrial genetics and have rarely used nuclear genetic data, leaving questions as to their phylogenetic relationships. This study aimed to utilize a bioinformatics approach to screen for candidate single-copy nuclear genes, using transcriptomic data from sturgeon species and genomic data from the spotted gar, *Lepisosteus oculatus*. We utilized nested polymerase chain reaction (PCR) and degenerate primers to identify nuclear protein-coding (NPC) gene markers to determine phylogenetic relationships among the Acipenseriformes. We identified 193 nuclear single-copy genes, selected from 1850 candidate genes with at least one exon larger than 700 bp. Forty-three of these genes were used for primer design and development of 30 NPC markers, which were sequenced for at least 14 Acipenseriformes species. Twenty-seven NPC markers were found completely in 16 species. Gene trees according to Bayesian inference (BI) and maximum likelihood (ML) were calculated based on the 30 NPC markers (20,946 bp total). Both gene and species trees produced very similar topologies. A molecular clock model estimated the divergence time between sturgeon and paddlefish at 204.1 Mya, approximately 10% later than previous estimates based on cytochrome b data (184.4 Mya). The successful development and application of NPC markers provides a new perspective and insight for the phylogenetic relationships of Acipenseriformes. Furthermore, the newly developed nuclear markers may be useful in further studies on the conservation, evolution, and genomic biology of this group.

Keywords: nuclear markers; phylogeny; Acipenseriformes; molecular clock

1. Introduction

In 2010, the International Union for Conservation of Nature (IUCN) Red List of Threatened Species concluded that “Sturgeon more critically endangered than any other group of species”. Considering IUCN Red List data, twenty-seven species of sturgeon are listed with 63 percent as Critically Endangered—the Red List’s highest category of threat. Four species are now possibly extinct. Although progress was made in recovery of North American species during the last decade,

the situation is dire for most Eurasian species. The latest update of the Red List assessed the status of 18 species of sturgeon from all over Europe and Asia and found that all were threatened (The IUCN Red List of Threatened Species). Within the Eurasian species, some are on the brink of extinction [1].

Mitochondrial genes have been used in molecular phylogenetic analyses and species identification since the 1990s. Their high copy number, easy amplification, and fast evolutionary rate [2,3] make them valuable markers addressing taxonomic issues in a wide range of species. However, there are important drawbacks to their use, such as their being inherited as a unit [3]. Therefore, all mitochondrial genes should be considered collectively as a single marker no matter how many genes are concatenated. Many empirical studies have suggested that the rate of mitochondrial evolution is fast, which limits the value of mitochondrial loci in deep-level phylogeny [4]. Nuclear molecular markers are the preferable choice for the phylogenetic and conservation genetic studies of ancient groups, and have been utilized in many species [5–12]. There are three main classes of molecular markers: nuclear protein coding loci exon-primed intron-crossing markers, and anonymous nuclear markers that have been successfully developed in non-model organisms to address a wide range of questions in phylogenetics and phylogeography [13]. Nuclear protein coding loci are used most frequently for several reasons, including their straightforward handling, low likelihood of paralogs, high number potential loci, and large phylogenetic span [13]. Meanwhile, exons are generally conservative and evolve slowly, making them suitable for the study of deep-level phylogenetics [4,14].

The Acipenseriformes lineages represent some of the most primitive lineages of recent vertebrates [15], with many members of the group under threat [16]. Acipenseriformes represent the “living fossils” of aquatic life on earth and have a history of more than 200 million years [17]. Many species are polyploid [16,18,19], but no reference genome sequence data have been published to date depending on their huge genome sizes. This has led to a focus on mitochondrial genes to establish phylogenetic relationships, with nuclear marker data lacking [15]. However, the results of molecular and morphological studies are controversial, generating uncertainty in the phylogenetic relationships and resulting in discussion about their conservation status. For example, there is no doubt that sturgeon and paddlefish are monophyletic groups [15,18,20–28]. However, most molecular results have indicated that the genus *Huso* is not a monophyletic group: *H. dauricus* and *H. huso* are nested in different positions within the genus *Acipenser* [15,18,22]. Meanwhile, Findeis (1997) and Artyukhin (2010) suggested the opposite based on morphological evidence [28,29]. Moreover, *A. sinensis* and *A. dabryanus* were found by Zhang et al. (2000) to be sister species based the mitochondrial genes NADH dehydrogenase subunit 4L (ND4L) and NADH dehydrogenase subunit 4 (ND4) [20], but cytochrome b (*cytb*), 16S ribosomal DNA (rDNA), and 12S rDNA did not confirm this [25]. Furthermore, *A. oxyrinchus* and *A. sturio* were deemed related based on the similarity of eight mitochondrial sequences (12S ribosomal (rRNA), 16S rDNA, *cytb*, tRNA_{Asp}, tRNA_{Phe}, NADH5, and control region) by Krieger et al. (2010) [15], but this finding did not agree with that of Birstein and DeSalle (1998) [25].

As such it is difficult to address the phylogenetic relationships within the Acipenseriformes using only mitochondrial data [15]. However, many species of Acipenseriformes are polyploid because of widespread genome duplication [22], complicating the search for nuclear molecular markers. With the development of sequencing technologies and analysis methods (especially based on transcriptome and reference genome development), nuclear markers have been reported in other species [6,7,11], providing frameworks for the development of such markers in the Acipenseriformes. To date, there are only a few examples of transcriptome data, such as *A. oxyrinchus* [30], *A. sinensis* [31], *A. baerii* [32], *A. transmontanus* [33], *A. schrenckii* [34], *A. naccarii*, and *A. stellatus* [35]. Fortunately, the available genome of *Lepisosteus oculatus* [36], transcriptomes of sturgeon, and previous studies developing nuclear markers provide new foundations to analyze the phylogenetic relationships within the Acipenseriformes.

In the present study, we aimed to develop a similar approach to generate new nuclear protein-coding (NPC) markers for phylogenetic analysis using a transcriptome plus reference genome

strategy. Four new transcriptomes of sturgeon and the *L. oculatus* genome were used as a database. These were analyzed by screening candidate single-copy genes to develop NPC markers and using multi-sequence alignment to find the conserved regions of exons. Two pairs of degenerate primers were designed for each marker for polymerase chain reaction (PCR) amplification. Thus, by first constructing a phylogenetic relationship and then reinvestigating the divergence times of the 16 Acipenseriformes species using the 30 newly developed NPC markers, our study provides new insight into Acipenseriformes phylogeny and improves our understanding of evolution at the nuclear gene level.

2. Materials and Methods

2.1. Datasets

The transcriptome sequencing data of seven species (*Acipenser baerii*, *A. gueldenstaedtii*, *A. naccarii*, *A. schrenckii*, *A. sinensis*, *A. stellatus*, and *A. transmontanus*) were searched and downloaded from the NCBI Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra>). Due to the difference in data size, here we only kept the transcriptome data of four species (*A. baerii*, *A. schrenckii*, *A. sinensis*, and *A. transmontanus*) for subsequent analysis. The accession numbers and other details are presented in Table S1. The SRA files were converted into fastq files using fastq-dump within the SRA Toolkit version 2.7.0 (Bethesda, MD, USA).

2.2. Quality Control and De Novo Transcriptome Assembly

The fastq data were first processed using fastqc version 0.11.3 software [37] (Babraham Institute, Cambridge, UK) to view the summary quality control of the reads. Trimmomatic version 0.33 [38] (RWTH Aachen University, Aachen, Germany) was used to filter the base reads for low quality and artificial errors in the sequence ends. In this step, bases with abnormal content of approximately 8 bp according to fastqc results and those showing low quality ($Q \leq 3$ base) at the start and end of the sequence were removed. The de novo assembly of clean reads was performed using Trinity version 2.0.6 software [39] (Broad Institute, Cambridge, MA, USA), with the min_kmer_cov value set to 2 and all other parameters set to default. Only contigs longer than 200 bp were used for further analysis. Then, the unigenes were extracted from the transcripts and searched against the protein data set from the *L. oculatus* genome using blastx version 2.2.31 (NCBI, Bethesda, MD, USA) with an E-value of 1×10^{-5} . Protein data sets were extracted from the blast results based on the Perl script. The coding sequences (CDS) and protein sequences of the unigenes without blastx hits were predicted using TransDecoder version 3.0.0 (<http://transdecoder.sourceforge.net/>).

2.3. Comparison of Transcripts and Extraction of Protein Sequences

The unigene sequences of four species (*A. baerii*, *A. sinensis*, *A. schrenckii*, and *A. transmontanus*) were compared with the protein sequences from the spotted gar (*L. oculatus*) whole genome using blastx (E-value $< 10^{-5}$). The protein sequences were then extracted according to the alignment results using the Perl script.

2.4. Identification of Orthologous Single-Copy Genes

We identified sturgeon orthologous genes using OrthoMCL version 2.0.9 [40] (<http://www.orthomcl.org>). Briefly, all protein sequences of the four sturgeons and the spotted gar were filtered according to length, with sequences >30 amino acid retained. Sequences between the start codon and the first stop codon which were less than 20% of the complete protein sequence were removed. We then used the all-vs-all BLAST method in the filtered files, and results with E-values $< 10^{-5}$ were retained. Clustering was performed using Markov cluster algorithm (MCL) to obtain gene families. Finally, the orthologous single-copy genes (1:1:1) were extracted using the Perl script.

In addition, the spotted gar genome and the corresponding annotation files were downloaded from Ensembl (<http://www.ensembl.org/index.html>). We used Perl scripts to determine the corresponding genes of the single-copy orthologous sequences of the four-sturgeon species, and single-copy genes containing at least one exon of >700 bp were identified for further analysis.

2.5. Exon Search and Primer Design

Nested PCR and degenerate primer strategies have been widely used in developing NPC markers because they can significantly improve the success rate of PCR [5,12,41]. In order to ensure the universality of primers, we screened 193 nuclear single-copy genes, which were selected from 1850 candidate genes containing at least one exon >700 bp. First, we downloaded specific information in fasta format for the 193 candidate genes of spotted gar from Ensembl (<http://asia.ensembl.org>). Then, we chose 43 nuclear single-copy genes for primer design, many having only one exon. We then integrated the information for each gene in a table and performed multiple alignments of the CDS from the five species. Conserved regions containing no termination codons were selected for designing degenerate primers (Figure S1). The length of each primer was 21–26 bp, with approximately 172 primers designed for the development of nuclear gene markers.

2.6. Taxon Sampling, DNA Extraction, and Experimental Testing

Sixteen Acipenseriformes species were sampled in this study, consisting of four genera (Table 1). Originally, specimens were identified based on morphological characteristics [28]; but species were also identified by sequencing of mitochondrial genes [1,18,22]. All animal experiments were conducted and approved by the Animal Care and Use Committee of Southwest University (No. 2017-7). Genomic DNA was extracted from fin tissues and egg previously stored at -80°C , using the Qiagen DNeasy Kit (Qiagen, Shanghai, China), according to manufacturer's instructions. The 43 NPC candidate loci were amplified using nested PCR. First-round PCR reactions were conducted in 25 μL volumes containing the following: 2.5 μL 10 \times buffer (Mg^{2+} plus), 2.0 μL of 2.5 mM dNTPs, 1 U Taq DNA polymerase (rTaq, TaKaRa; Dalian, China), 1 μL of each first-round primer (10 μM), 1 μL genomic DNA (~ 100 ng/ μL), and double-distilled water to yield a final volume of 25 μL . The following conditions were used for PCR amplification: initial denaturation for 4 min at 94°C ; 35 cycles of 45 s at 94°C , 40 s at 45°C , and 2 min at 72°C ; and a final extension of 10 min at 72°C [12]. Negative controls (i.e., containing no DNA templates) were used in each PCR run to test for contamination and artifacts. Second-round PCR (25 μL) contained the following: 2.5 μL 10 \times buffer (Mg^{2+} plus), 2.0 μL of 2.5 mM dNTPs, 1 U Taq DNA polymerase (rTaq, TaKaRa; Dalian, China), 1 μL of each second-round primer (10 μM), 0.8–1.5 μL first-round PCR product without dilution, and double-distilled water to yield a final volume of 25 μL . The following conditions were used for PCR amplification: an initial denaturation for 4 min at 94°C ; 30 cycles of 45 s at 94°C , 40 s at 50 – 52°C , and 1.5 min at 72°C ; and a final extension of 10 min at 72°C . Samples were then cooled to 4°C . Second-round PCR products were examined via electrophoresis through 1% agarose gels. If multiple bands were present, the correct PCR product according to size was purified by cutting the band from the gel. In order to verify that markers were effective, extraction and amplification of these markers was attempted using the relevant primer pair in three species (*A. dabryanus*, *A. sinensis*, and *Polyodon spathula*). When single-band amplification, sequencing, and BLAST searching against the target gene in these three species were satisfied, the marker was tested in the remaining 13 species. An NPC candidate locus was considered a feasible marker if Sanger sequencing was successful and results were achieved in at least 13 out of the 16 species (Figure S1).

Table 1. Sample collection list of all species used in this study.

| Species | Common Name | Collection Locality | Distribution ¹ |
|-----------------------------|--------------------|--|---------------------------|
| <i>Acipenser</i> | | | |
| <i>A. baerii</i> | Siberian sturgeon | Germany | SAO |
| <i>A. dabryanus</i> | Yangtze sturgeon | Yibin, Sichuang, China | CH |
| <i>A. fulvescens</i> | Lake sturgeon | Wolfgangsee, Wisconsin, America | GL |
| <i>A. gueldenstaedtii</i> | Danube sturgeon | Germany | PC |
| <i>A. oxyrinchus</i> | Atlantic sturgeon | Canada | MGM/NWA |
| <i>A. persicus</i> | Persian sturgeon | Iran | PC |
| <i>A. ruthenus</i> | Sterlet sturgeon | The Danube, Germany | NEA/SAO/PC |
| <i>A. schrenckii</i> | Amur sturgeon | Wanzhou, Chongqing, China | ASJ |
| <i>A. sinensis</i> | Chinese sturgeon | Yangtze river fisheries research institute, Hubei, China | CH |
| <i>A. stellatus</i> | Starry sturgeon | The Danube, Romania | PC |
| <i>A. sturio</i> | Sturgeon | Gironde River, France | NEA/PC |
| <i>A. transmontanus</i> | White sturgeon | Amur river, China | NEP |
| <i>Huso</i> | | | |
| <i>H. dauricus</i> | Kaluga | Amur river, China | ASJ |
| <i>H. huso</i> | Beluga | The Danube, Romania | PC |
| <i>Pseudoscaphirhynchus</i> | | | |
| <i>P. kaufmanni</i> | Amu Darya sturgeon | Amu Darya, Turkmenistan | PC |
| <i>Polyodon</i> | | | |
| <i>P. spathula</i> | Paddlefish | Wanzhou, Chongqing, China | MGM |

¹ NEP: North Eastern Pacific; GL: Great Lakes, Hudson Bay & St. Lawrence R., NWA: North Western Atlantic; MGM: Mississippi R. & Gulf of Mexico; NEA: Northeastern Atlantic, including White, Baltic & North seas; PC: Ponto-Caspian Region, including Mediterranean, Aegean, Black, Caspian & Aral seas; SAO: Siberia & Arctic Ocean; ASJ: Amur R., Sea of Okhotsk & Sea of Japan; CH: China (Bemis and Kynard, 1997).

2.7. Sequence Assembly and Phylogenetic Analyses

Second-round PCR products were sequenced using Sanger sequencing. The forward and reverse sequences of each NPC maker were manually assembled using ContigExpress version 3.0.0 (Invitrogen; Carlsbad, CA, USA) in the Vector NIT software suite [42] and visually examined (Figure S1). The sequences were aligned using the clustalW algorithm in MEGA version 7 [43]. Ambiguously aligned regions were removed using Gblocks version 0.91b [44], with all gaps allowed ($-b5 = a$), utilizing the ‘codon’ model setting and setting all other values to default. We used Sequence Matrix version 1.8 [45] to concatenate successful NPC markers. The variable sites (variability) and parsimony information (PI) sites of each alignment and concatenated data were calculated by MEGA 7. The supermatrix data were divided into four parts for saturation testing by DAMBE version 5 [46]. Then, we manually defined five partitioning strategies: two partitions (1- and 2-codon positions together as a partition and one partition for the 3-codon position), three partitions (one partition for each codon position), 30 partitions (each gene as a partition), 60 partitions (one partition for 1- and 2-codon positions together and the 3-codon position as a partition across 30 genes), and 90 partitions (codon position partitioning across 30 genes). The comparisons of the five partitioning strategies and selections of corresponding nucleotide substitution models were conducted under the corrected Akaike information criterion implemented in PartitionFinder version 2 [47].

The supermatrix dataset was separately analyzed with both maximum likelihood (ML) and Bayesian inference (BI) methods with 60 schemes. RAxML version 8.0 [48] was used to analyze the partition ML, with the GTR + Γ + I model assigned for each partition. Branch support for the resulting phylogeny was evaluated with 500 rapid bootstrapping replicates (with $-f$ as an option) implemented in RAxML. MrBayes version 3.2 [49] was used to analyze partition BI. All model parameters were unlinked, with one cold chain and three heated chains for the Markov chain Monte Carlo (MCMC) process, which began with random starting trees. The analysis was run for 30 million generations and sampled every 1000 generations, with the first 25% representing burn-in. Posterior probabilities were obtained from the 50% majority rule consensus tree of the remaining topologies using TreeAnnotator, and the above process was run independently twice. Chain stationarity was visualized by plotting

likelihoods against the generation number using the program Tracer version 1.7 [50]. Effective sample sizes greater than 200 ensured the convergence of the results.

The construction of the species tree was completed by ASTRAL version 5.5.9 [51] based on gene trees estimated from the 30 NPC markers. This Java program estimates a species tree given a set of unrooted gene trees and support is calculated using local posterior probabilities [52]. We used 30 maximum likelihood gene trees as input trees, derived using the GTR + Γ + I model in RAxML with five hundred bootstrap replicates.

2.8. Hypothesis Testing

To test the statistical significance of alternative hypotheses for the Acipenseridae family within the Acipenseriformes, we compared several previous studies [15,24,28,29] with respect to the relationships among three genera within the family (i.e., *Pseudoscaphirhynchus*, *Huso*, and *Acipenser*). Topology tests were conducted using the datasets of all 16 species. The site-log-likelihood values were calculated under the GTR + Γ + I model for nucleotides using TREE-PUZZLE version 5.3 [53] (command-line option: -wsl). The obtained values were used as input for the software, CONSEL [54]. Hypotheses were statistically tested by AU [55], KH [56], and SH [57].

2.9. Estimating Divergence Time

In order to increase the speed of calculation, molecular dating was estimated using the relaxed-clock based program MCMCTREE version 4.9e (University College London, London, UK) [58]. In this analysis, the best tree from the hypothesis testing was used as the reference tree. All 30 NPC markers were concatenated as a single “supermatrix” (20,946 bp), and the alignment was divided into three partitions corresponding to the 1st, 2nd, and 3rd codon sites. The approximate likelihood method (usedata = 2) [58] was used for divergence time estimation. For each locus, the root age was set at “<2.0” and the model at “HKY85 + G”. The molecular clock model used independent rates (clock = 2). At present, there is only one report on the complete estimation of divergence times between species within the Acipenseriformes order [18]. Therefore, two nodes (C1 and C2) were considered as time-calibrated points with lognormal distributions and soft constraint bands (allowing violation probability of 0.025) in the present study. The C1 calibration point was estimated as the split time between sturgeon and paddlefish, based on a fossil of a Hauterivian-age polyodontid, dating from Early Cretaceous (~132–200 Mya) [17,59]. The C2 calibration point was used to estimate the node that *A. oxyrinchus* and *A. sturio* splitting with other sturgeons based on the acipenserid fossil from the Campanian age, dating from Late Cretaceous (~74 Mya) [60] (Table S2). We used 500,000 generations for the Markov chain Monte Carlo (MCMC) analysis, with the first 20% of all samples discarded as burn-in and samples collected every 200 generations thereafter up to 20,000. An independent rates model (clock = 2), which follows a lognormal distribution, was used for the MCMC search.

3. Results

3.1. Quality Control and Assembly

After removing artifact erroneous bases at the ends of the sequences and low-quality reads, the high-quality reads were used for further analysis. The software, Trinity [39], was used to guarantee transcripts. In the same species, all reads from different libraries or tissues were pooled. The longest isoform of a gene served as the unigene. Finally, we obtained 13,666, 125,179, 211,209, and 234,179 unigenes with an N50 of 787, 653, 654, and 704 bp from *A. sinensis*, *A. transmontanus*, *A. schrenckii*, and *A. baerii*, respectively.

3.2. Identification of Putative Proteins

We analyzed the possible protein sequences to identify orthologous genes for comparison of the species divergence of sequences. The unigenes of each species were used to predict CDS and

protein sequences. We recovered 37,751, 36,648, 32,470, and 36,490 protein sequences (N50: 224, 216, 319, and 323 bp, respectively) from *A. sinensis*, *A. transmontanus*, *A. schrenckii*, and *A. baerii* based on the search results from blastx. The remaining unigenes were predicted using TransDecoder pipeline, obtaining 9530, 7485, 12,226, and 14,040 amino acid sequences with N50 lengths of 195, 188, 177, and 191 bp, for the four species respectively (Table S3). These two groups of protein sequences were combined and used for the identification of homologous genes. Finally, these amino acid sequences from *A. sinensis*, *A. transmontanus*, *A. schrenckii*, and *A. baerii* were used for further analysis. A total of 21,820 homologous gene families were obtained in four sturgeons and spotted gar protein sequences clustered by using OrthoMCL. Among them, 1850 single-copy orthologous genes were identified, including spotted gar genes. In addition, 193 nuclear single-copy genes that contained at least one exon of >700 bp were selected from the spotted gar genome.

3.3. Characteristics of NPC Markers

According to the abovementioned screening criteria, we successfully found 30 new NPC markers for Acipenseriformes, which were amplified in at least 14 species. Fragments ranged from 467 bp to 993 bp, with an average length of 698 bp. In a total of 480 PCR reactions (30 loci \times 16 species), 476 (99.1%) were successfully amplified, with 469 (97.7%) producing high-quality bands used for Sanger sequencing (Figure S2). The PCR primer information of the 30 NPC markers, including their genome locations, is listed in Table S4 and their GenBank accession numbers are shown in Table S5. Their variability ranged from 4% to 16%, with an average of 9%, and their parsimony-informative sites ranged from 2% to 16%, with an average of 4% (Table S6). Twenty-seven NPC markers were amplified in all 16 species. The NPC marker 'fam43a' could not be amplified from *H. dauricus* or *Pseudoscaphirhynchus kaufmanni*, and 'ENSLOCG00000018097' and 'cebpa' could not be amplified from *P. kaufmanni* (Figure S2).

3.4. Reconstructing Phylogenetic Relationships in the Acipenseriformes

The supermatrix combined 30 NPC markers based on 20,946 bp. The results of the saturation tests showed that the values of the saturation index (*Iss*) for the first codon, second codon, third codon, first and second codons, and all three sites of protein-coding regions were significantly smaller than the critical values, *Iss.cSym* or *Iss.cAsym* (Table S7).

The results of PartitionFinder 2 supported the 60 schemes as the best-fitting partitioning schemes using both ML and BI analyses. Both the ML tree and BI tree produced highly similar topologies. Ten out of thirteen internal nodes produced bootstrap support of >75% in the ML tree and all internal nodes had posterior probabilities >0.95 in the BI tree (Figure 1). The species tree produced similar topologies with ML and BI trees in most lineages, and most internal nodes had high branch support (Figure 2). The gene and species trees can divide the Acipenseridae into three clades: clade 1, containing *A. sturio* and *A. oxyrinchus* as sister species; clade 2, containing the Pacific species *A. dabryanus*, *A. sinensis*, *A. schrenckii*, *A. transmontanus*, and *H. dauricus*; and clade 3, containing the Atlantic sturgeons *A. baerii*, *A. fulvescens*, *A. gueldenstaedtii*, *A. persicus*, *A. ruthenus*, *A. stellatus*, *H. huso*, and *P. kaufmanni*.

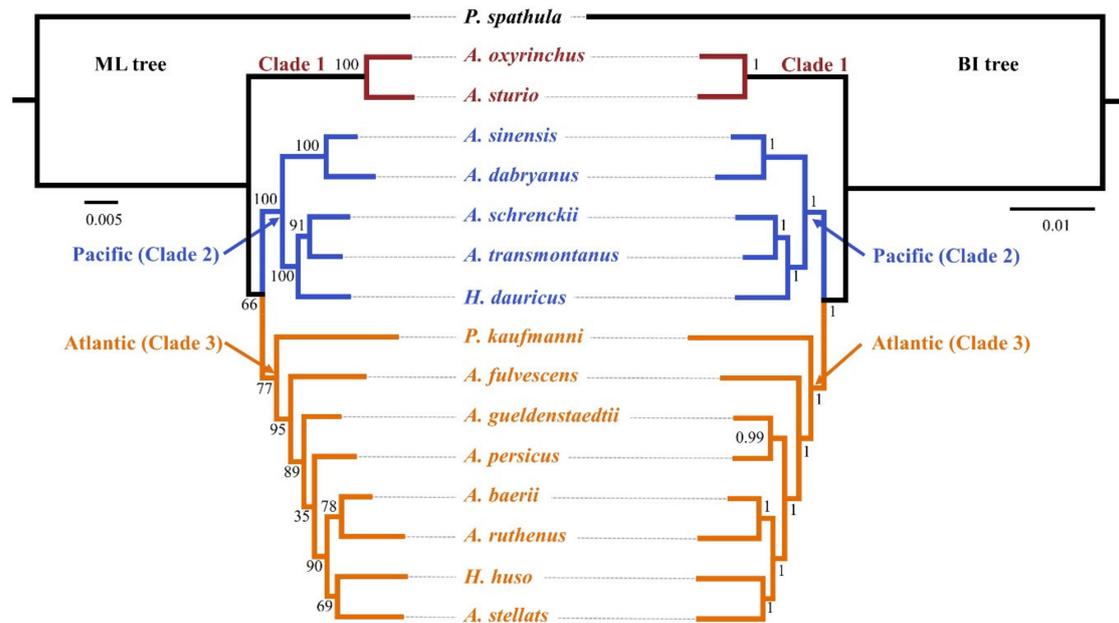


Figure 1. Phylogenetic relationships of Acipenseriformes inferred from 30 nuclear protein-coding (NPC) markers. Left: maximum likelihood (ML) tree; right: Bayesian inference (BI) tree; bootstrap support and Bayesian posterior probability are shown on the nodes.

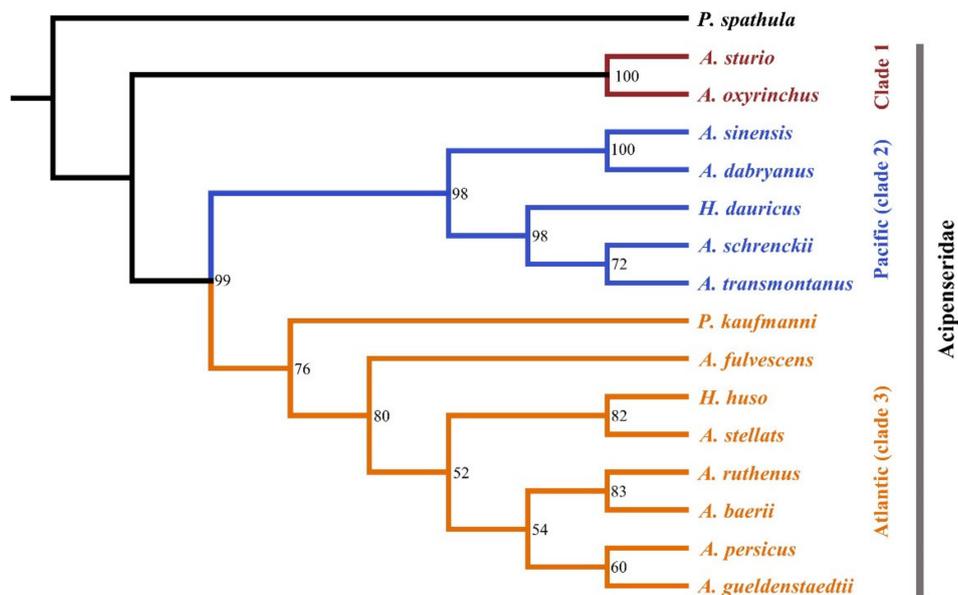


Figure 2. Species tree estimation of Acipenseriformes based on 30 NPC markers using the method of accurate species tree algorithm (ASTRAL). Bootstrap support is shown on the nodes.

3.5. Hypothesis Testing

The topology test strongly supported our gene tree (Table 2). The genus *Huso* was not found to be a monophyletic group, where *H. dauricus* and *H. huso* were dispersed and nested within the genus *Acipenser*. *P. kaufmanni* was also nested in the genus *Acipenser*. The topologies that rejected the hypothesis were (*Pseudoscaphirhynchus* + (*Huso* + *Acipenser*)) [28] and (*Huso* + (*Pseudoscaphirhynchus* + *Acipenser*)) [29]; both were based on morphological traits.

Table 2. Tree topology testing.

| Rank | Topology Tested | log L | AU | KH | SH |
|------|---|------------|---------------------|-------|-------|
| 1 | Best tree (ML tree) | −47,716.74 | 1.000 | 1.000 | 1.000 |
| 2 | ((((ASTE,PK),AR)),HH),(((AT,ASC),(ASI,AD)),HD) ¹ | −47,944.33 | 6×10^{-8} | 0 | 0.011 |
| 3 | ((ASTE,PK),((AR,HH),HD),AF) ² | −48,422.00 | 2×10^{-25} | 0 | 0 |
| 4 | ((Acipenser, PK),(HH,HD)) ³ | −48,869.58 | 0.008 | 0 | 0 |
| 5 | ((((HH,HD),(AR,ASC)),ASTE),(ASTU,AO)),PK) ⁴ | −48,915.92 | 2×10^{-62} | 0 | 0 |

¹ Krieger et al., 2010 [15]; ² Birstein et al., 2002 [24]; ³ Findeis, 1997 [29]; ⁴ Artyukhin, 2010 [28]. AD = *Acipenser dabryanus*, AO = *Acipenser oxyrinchus*, AR = *Acipenser ruthenus*, ASC = *Acipenser schrenckii*, ASTE = *Acipenser stellatus*, ASI = *Acipenser sinensis*, ASTU = *Acipenser sturio*, AT = *Acipenser transmontanus*, HD = *Huso dauricus*, HH = *Huso huso*, PK = *Pseudoscaphirhynchus kaufmanni*.

3.6. Divergence Times

The estimated divergence times of the 16 species of Acipenseriformes are shown in Figure 3. Our results suggest that the estimated divergence of sturgeon and paddlefish occurred at 204.1 Mya, with a 95% credibility interval of 180.3–233.6 Mya. Our times are not only earlier than the 141.4 Mya and 184.4 Mya [18] calculated based on partial mitochondrial genome sequences and *cytb* data, respectively, but are also earlier than the results of 114.1 Mya for data set 1 and 145.2 Mya for data set 2 obtained by Inoue et al. (2005) [61]. The divergence time of *A. sturio*–*A. oxyrinchus* clade and the other sturgeon species was estimated about 144.9 Mya, with a 95% credibility interval of 129.9–175.3 Mya. The Pacific-Atlantic split in this species occurred at 117.6 Mya (95% credibility interval of 98.9–144.8 Mya). *P. kaufmanni* separated at 103.3 Mya (95% credibility interval of 85.3–128.0 Mya). In addition, the *A. sinensis* and *A. dabryanus* divergence time dated back to 44.7 Mya, with a 95% credibility interval of 31.4–61.8 Mya.

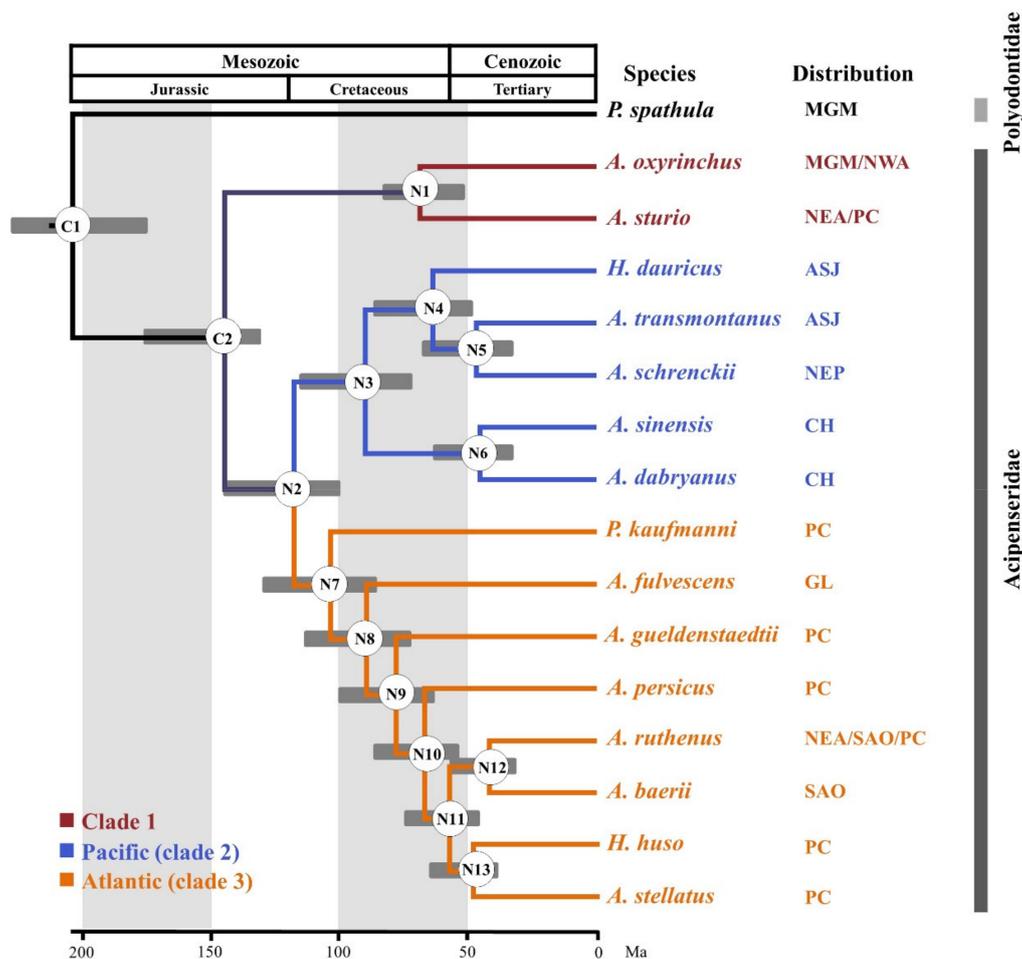


Figure 3. Timetree of 16 Acipenseriformes species. The node number represents the corresponding number. C1 and C2 are calibration nodes. N1–N13 indicate the nodes of interest for which estimated dates are presented in Table S8. CH: China; ASJ: Amur River, Sea of Okhotsk, Sea of Japan; NEP: Northeast Pacific.

4. Discussion

Polyploidization is a common phenomenon in animals and plants [62], which unfortunately causes problems for the use of nuclear markers in phylogenetic analyses. These problems make conventional methods of nuclear marker development very inefficient. On the contrary, searching conservative sequences based on multi-sequence alignment provides a new method to develop large scale nuclear markers [5,7,9,63]. Many Acipenseriformes species are polyploid and lack reference genome sequence data; as a result, there has been strong dependence on mitochondrial data of the phylogenetic relationships among species, with nuclear data rarely used. Nowadays, most Acipenseriformes are rare and often close to extinction, making sampling of wild specimens very difficult.

Twenty to fifty loci are often sufficient to answer phylogenetic questions [64,65]. Therefore, we used a strategy of multi-sequence alignment of transcriptomes combined with spotted gar genomic data to develop single-copy NPC markers to construct the phylogenetics of Acipenseriformes. The success rate of PCR amplification is an important parameter [41]. Our success rate was 99.1%, representing 30 NPC markers (100% for 27 markers) in 16 species of Acipenseriformes (Figure S2); the combination of low degenerate primers and nested PCR was important for this high rate of success. Phylogenetic analysis showed that the gene and species trees have highly similar topologies and most internal nodes are highly supported, showing that our NPC markers are reliable. This strategy for developing NPC markers can be used for the development of molecular markers in other polyploid groups that also

lack reference genomes but have sufficient transcriptome data. Notably, the set of 30 NPC markers can be used for the remaining 11 sturgeon species.

The results from the nuclear gene trees and species tree strongly support the classification of Acipenseriformes into two families, Acipenseridae and Polyodontidae, which is consistent with previous morphological and molecular studies [15,18,21–25,27,28,51,66]. The monophyly of sturgeon and paddlefish is not disputed. Nonetheless, the phylogenetic relationships of the four genera in the family Acipenseridae have been controversial for a long time. Our results suggest the existence of three clades, where clade 1 is represented by *A. sturio* and *A. oxyrinchus* as sister species with a unified basal lineage. All other species of *Acipenser*, *Huso*, and *Pseudoscaphirhynchus* can be divided into Pacific species (clade 2) and Atlantic species (clade 3) [15,18]. Our data support the Pacific clade being a monophyletic group, in agreement with previous molecular studies [15,18,22,26]. However, the internal phylogenetic relationships are different from those of previous studies. For example, *A. dabryanus* and *A. sinensis* have been found to be a sister group with *A. schrenckii* and *A. transmontanus* based on many mitochondrial DNA data, but nuclear markers suggest that *A. schrenckii* and *A. transmontanus* should be a sister group with *H. dauricus*. Presently, *A. dabryanus* and *A. sinensis* are found in China; *H. dauricus* and *A. schrenckii* are found in the Amur River, Sea of Okhotsk, and Sea of Japan; and *A. transmontanus* is found in the North Eastern Pacific (Table 1, Figure 3). Therefore, our results are consistent with geographical distribution of species. Currently, the species composition of the Atlantic clade is complex and contains at least three genera, including *Acipenser*, *Huso*, and *Pseudoscaphirhynchus*. The genus *Huso* is not a monophyletic group, with *H. dauricus* and *H. huso* dispersed and nested within the genus *Acipenser*, consistent with previous molecular studies [15,18,22]. The genus *Pseudoscaphirhynchus* is also nested in the genus *Acipenser*. Previous studies have indicated that *P. kaufmanni* is closely related with *A. stellatus* within the Atlantic clade [15,24]. However, our data do strongly support that *P. kaufmanni* is located basal in the Atlantic clade, which has not been noted previously.

Prior to this study, there was only one complete report on the divergence time of 24 species of Acipenseriformes; therefore, we choose *cytb* results for comparison [18]. Considering the slower saturation of nuclear exon data, time estimates based on nuclear exons would seem to be more reliable [67]. Notably, our results showed that the nuclear gene data set was narrower and more accurate in its credibility intervals. We estimated the divergence time for sturgeon and paddlefish at 204.1 Mya, similar to the oldest Acipenseriformes fossil record (ca. 200 Mya) [17,59]. The Ponto-Caspian region has been very unstable over the last 150 Mya [68], which may have led to the diversification of the Acipenseridae [26]. Based on our dating results, the divergence time of Acipenseridae was estimated to be 144.9 Mya (Table S8), more recently than the 171.6 Mya estimated previously [18], but close to 150 Mya, supporting the hypothesis of Bemis and Kynard (1997) [26]. Interestingly, we determined that the Pacific-Atlantic split occurred at around 117.6 Mya, around the time that the Tethys Sea separated into two oceans (120 Mya) [68]. We suggest that the continued shrinkage of the Tethys Sea greatly changed the habitat, resulting in a significant diversification of the sturgeon species. Additionally, the divergence time of the genus *Pseudoscaphirhynchus* was estimated to be 103.3 Mya, less recently than estimated by all previous studies. However, due to the incomplete sampling of species, the actual divergence time may be a little earlier than our results indicate.

5. Conclusions

The transcriptome plus reference genome data mining strategy utilized here to identify suitable nuclear protein-coding loci for novel marker development has been proven successful in Acipenseriformes. The newly developed 30 NPC markers had a high rate of experimental success (~99.1%), with excellent performance of the concatenated dataset describing phylogenetic relationships. This dataset and markers were useful in reconstructing the relationships within the Acipenseriformes order, yielding the novel findings that *P. kaufmanni* is identified located in the base position of the Atlantic clade of species. As more sequences become available in the future, further studies with more comprehensive taxon sampling are essential to fully resolve the evolutionary history of

Acipenseriformes. Additionally, our marker set offers a great opportunity for species identification including artificial produced hybrids. Trade control is essential for protection of last remaining wild stocks.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/10/1/38/s1>, Figure S1: Flowchart for bioinformatics screening and experimental design, Figure S2: PCR amplification rendering of the 30 NPC markers in 16 Acipenseriformes species, Table S1: Transcriptome information of four sturgeons, Table S2: Maximum (U) and minimum (L) constrains (Mya) and calibration information for nodes in Figure 1 in the molecular dating analysis, Table S3: Number of predicted protein sequences (N) according to detected unigenes, Table S4: Summary information of 30 NPC markers, Table S5: List of GenBank accession numbers for the markers obtained in the present study, Table S6: Summary information of 30 NPC markers in 16 species of Acipenseriformes, Table S7: Saturation test performed by using DAMBE, Table S8: Detailed results of Bayesian molecular dating using MCMCTREE, and a comparison among shared nodes in previous study.

Author Contributions: Conceptualization, Z.P. and A.L.; methodology, D.L. and Y.L.; software, D.L., Q.Z., and L.Z.; validation, D.L. and Y.L.; formal analysis, D.L.; investigation, D.L. and Y.L.; resources, Z.P. and A.L.; data curation, D.L.; writing—original draft preparation, D.L.; writing—review and editing, Z.P., A.L., and Y.L.; visualization, D.L.; supervision, Z.P.; project administration, Z.P.; funding acquisition, Z.P.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 31272283 and 31572254, and the Fundamental Research Funds for the Central Universities, grant number XDJK2015A011.

Acknowledgments: We thank Sovan Lek, Bo Ma, Dengqiang Wang, and Hua Ye for providing fish tissue samples. We also thank Xingxing Shen for helping with data analysis and Yang Mao and Yabing Niu for their help with laboratory work. We are grateful to Christopher BurrIDGE for comments on an early version of the manuscript.

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

References

- Ludwig, A. Identification of Acipenseriformes species in trade. *J. Appl. Ichthyol.* **2008**, *24*, 2–19. [[CrossRef](#)]
- Simon, C.; Frati, B.F.; Stewart, J.B.; Beckenbach, A.T. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.* **2006**, *37*, 545–579. [[CrossRef](#)]
- Harrison, R.G. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends Ecol. Evol.* **1989**, *4*, 6–11. [[CrossRef](#)]
- Townsend, T.; Alegre, R.; Kelley, S.; Wiens, J.; Reeder, T.W. Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: An example from squamate reptiles. *Mol. Phylogenet. Evol.* **2008**, *47*, 129–142. [[CrossRef](#)]
- Li, Z.; De La Torre, A.R.; Sterck, L.; Cánovas, F.M.; Avila, C.; Merino, I.; Cabezas, J.A.; Cervera, M.T.; Ingvarsson, P.K.; Van de Peer, Y. Single-copy genes as molecular markers for phylogenomic studies in seed plants. *Genome Biol. Evol.* **2017**, *9*, 1130–1147. [[CrossRef](#)] [[PubMed](#)]
- Teasdale, L.C.; Köhler, F.; Murray, K.D.; O'Hara, T.; Moussalli, A. Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture. *Mol. Ecol. Resour.* **2016**, *16*, 1107–1123. [[CrossRef](#)] [[PubMed](#)]
- Rothfels, C.J.; Larsson, A.; Li, F.W.; Sigel, E.M.; Huiet, L.; Burge, D.O.; Ruhsam, M.; Graham, S.W.; Stevenson, D.W.; Wong, G.K.; et al. Transcriptome-mining for single-copy nuclear markers in ferns. *PLoS ONE* **2013**, *8*, e76957. [[CrossRef](#)]
- Peng, Z.; Elango, N.; Wildman, D.E.; Yi, S.V. Primate phylogenomics: Developing numerous nuclear non-coding, non-repetitive markers for ecological and phylogenetic applications and analysis of evolutionary rate variation. *BMC Genom.* **2009**, *10*, 247. [[CrossRef](#)]
- Li, C.; Ortí, G.; Gong, Z.; Lu, G. A practical approach to phylogenomics: The phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evol. Biol.* **2007**, *7*, 44. [[CrossRef](#)]
- Duarte, J.M.; Wall, P.K.; Edger, P.P.; Landherr, L.L.; Hong, M.; Pires, P.K.; Leebens-Mack, J.; Depamphilis, C.W. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol. Biol.* **2010**, *10*, 61. [[CrossRef](#)]
- Deng, H.; Zhang, G.Q.; Lin, M.; Wang, Y.; Liu, Z.J. Mining from transcriptomes: 315 single-copy orthologous genes concatenated for the phylogenetic analyses of Orchidaceae. *Ecol. Evol.* **2015**, *5*, 3800–3807. [[CrossRef](#)] [[PubMed](#)]

12. Che, L.H.; Zhang, S.Q.; Li, Y.; Liang, D.; Pang, H.; Ślipiński, A.; Zhang, P. Genome-wide survey of nuclear protein-coding markers for beetle phylogenetics and their application in resolving both deep and shallow-level divergences. *Mol. Ecol. Resour.* **2017**, *17*, 1342–1358. [[CrossRef](#)] [[PubMed](#)]
13. Thomson, R.C.; Wang, I.J.; Johnson, J.R. Genome-enabled development of DNA markers for ecology, evolution and conservation. *Mol. Ecol.* **2010**, *19*, 2184–2195. [[CrossRef](#)] [[PubMed](#)]
14. Morgan, C.C.; Creevey, C.J.; O'Connell, M.J. Mitochondrial data are not suitable for resolving placental mammal phylogeny. *Mamm. Genome* **2014**, *25*, 636–647. [[CrossRef](#)] [[PubMed](#)]
15. Krieger, J.; Hett, A.K.; Fuerst, P.A.; Artyukhin, E.A.; Ludwig, A. The molecular phylogeny of the order Acipenseriformes revised. *J. Appl. Ichthyol.* **2010**, *24*, 36–45. [[CrossRef](#)]
16. Rajkov, J.; Shao, Z.; Berrebi, P. Evolution of polyploidy and functional diploidization in sturgeons: Microsatellite analysis in 10 sturgeon species. *J. Hered.* **2014**, *105*, 521–531. [[CrossRef](#)]
17. Grande, L.; Bemis, W.E. Osteology and phylogenetic relationships of fossil and recent paddlefishes (Polyodontidae) with comments on the interrelationships of Acipenseriformes. *J. Vertebr. Paleontol.* **1991**, *11*, 1–121. [[CrossRef](#)]
18. Peng, Z.; Ludwig, A.; Wang, D.; Rui, D.; Wei, Q.; He, S. Age and biogeography of major clades in sturgeons and paddlefishes (pisces: Acipenseriformes). *Mol. Phylogenet. Evol.* **2007**, *42*, 854–862. [[CrossRef](#)]
19. Fontana, F.; Tagliavini, J.; Congiu, L. Sturgeon genetics and cytogenetics: Recent advancements and perspectives. *Genetica* **2001**, *111*, 359–373. [[CrossRef](#)]
20. Zhang, S.; Zhang, Y.; Zheng, X.; Chen, Y.; Deng, H.; Wang, D.; Wei, Q.; Zhang, Y.; Long, N.; Wu, Q. Molecular phylogenetic systematics of twelve species of acipenseriformes based on mtDNA *ND4L-ND4* gene sequence analysis. *Sci. China Life Sci.* **2000**, *43*, 129–137. [[CrossRef](#)]
21. Ludwig, A.; May, B.; Debus, L.; Jenneckens, I. Heteroplasmy in the mtdna control region of sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). *Genetics* **2000**, *156*, 1933–1947. [[PubMed](#)]
22. Ludwig, A.; Belfiore, N.M.; 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. [[PubMed](#)]
23. Krieger, J.; Fuerst, P.A. Evidence for a slowed rate of molecular evolution in the order Acipenseriformes. *Mol. Biol. Evol.* **2002**, *19*, 891–897. [[CrossRef](#)] [[PubMed](#)]
24. Birstein, V.J.; Doukakis, P.; Desalle, R. Molecular phylogeny of Acipenseridae: Nonmonophyly of scaphirhynchinae. *Copeia* **2002**, *2002*, 287–301. [[CrossRef](#)]
25. Birstein, V.J.; Desalle, R. Molecular phylogeny of Acipenserinae. *Mol. Phylogenet. Evol.* **1998**, *9*, 141–155. [[CrossRef](#)] [[PubMed](#)]
26. Bemis, W.E.; Kynard, B. Sturgeon rivers: An introduction to Acipenseriform biogeography and life history. *Environ. Biol. Fishes* **1997**, *48*, 167–183. [[CrossRef](#)]
27. Bemis, W.E.; Findeis, E.K.; Grande, L. An overview of Acipenseriformes. *Environ. Biol. Fishes* **1997**, *48*, 25–71. [[CrossRef](#)]
28. Artyukhin, E.N. Morphological phylogeny of the order Acipenseriformes. *J. Appl. Ichthyol.* **2010**, *22*, 66–69. [[CrossRef](#)]
29. Findeis, E.K. Osteology and phylogenetic interrelationships of sturgeons (Acipenseridae). *Environ. Biol. Fishes* **1997**, *48*, 73–126. [[CrossRef](#)]
30. Kaitetidou, E.; Ludwig, A.; Gessner, J.; Sarropoulou, E. Expression patterns of atlantic sturgeon (*Acipenser oxyrinchus*) during embryonic development. *G3-Genes Genoms Genet.* **2017**, *7*, 533–542. [[CrossRef](#)]
31. Yue, H.; Li, C.; Du, H.; Zhang, S.; Wei, Q. Sequencing and de novo assembly of the gonadal transcriptome of the endangered Chinese sturgeon (*Acipenser sinensis*). *PLoS ONE* **2015**, *10*, e0127332. [[CrossRef](#)]
32. Song, W.; Jiang, K.; Zhang, F.; Lin, Y.; Ma, L. Transcriptome sequencing, de novo assembly and differential gene expression analysis of the early development of *Acipenser baeri*. *PLoS ONE* **2015**, *10*, e0137450. [[CrossRef](#)] [[PubMed](#)]
33. Chen, M.Y.; Liang, D.; Zhang, P. Selecting question-specific genes to reduce incongruence in phylogenomics: A case study of jawed vertebrate backbone phylogeny. *Syst. Biol.* **2015**, *64*, 1104–1120. [[CrossRef](#)] [[PubMed](#)]
34. Yuan, L.; Zhang, X.; Li, L.; Jiang, H.; Chen, J. High-throughput sequencing of microrna transcriptome and expression assay in the sturgeon, *Acipenser schrenckii*. *PLoS ONE* **2014**, *9*, e115251. [[CrossRef](#)]

35. Vidotto, M.; Grapputo, A.; Boscari, E.; Barbisan, F.; Coppe, A.; Grandi, G.; Kumar, A.; Congiu, L. Transcriptome sequencing and de novo annotation of the critically endangered adriatic sturgeon. *BMC Genom.* **2013**, *14*, 407. [[CrossRef](#)] [[PubMed](#)]
36. Braasch, I.; Gehrke, A.R.; Smith, J.J.; Kawasaki, K.; Manousaki, T.; Pasquier, J.; Amores, A.; Desvignes, T.; Batzel, P.; Catchen, J.; et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* **2016**, *48*, 427–437. [[CrossRef](#)] [[PubMed](#)]
37. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*; Babraham Institute: Cambridge, UK, 2010.
38. Bolger, A.M.; Marc, L.; Bjoern, U. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
39. Haas, B.J.; Alexie, P.; Moran, Y.; Manfred, G.; Blood, P.D.; Joshua, B.; Matthew Brian, C.; David, E.; Bo, L.; Matthias, L. De novo transcript sequence reconstruction from rna-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)]
40. Li, L.; Stoeckert, C.J., Jr.; Roos, D.S. Orthomcl: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **2003**, *13*, 2178–2189. [[CrossRef](#)]
41. Shen, X.-X.; Liang, D.; Wen, J.-Z.; Zhang, P. Multiple genome alignments facilitate development of NPCL markers: A case study of tetrapod phylogeny focusing on the position of turtles. *Mol. Biol. Evol.* **2011**, *28*, 3237–3252. [[CrossRef](#)]
42. Lu, G.; Moriyama, E.N. Vector NTI, a balanced all-in-one sequence analysis suite. *Brief. Bioinform.* **2004**, *5*, 378–388. [[CrossRef](#)] [[PubMed](#)]
43. Kumar, S.; Stecher, G.; Tamura, K. Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)] [[PubMed](#)]
44. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [[CrossRef](#)] [[PubMed](#)]
45. Vaidya, G.; Lohman, D.J.; Meier, R. Sequencematrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* **2011**, *27*, 171–180. [[CrossRef](#)]
46. Xia, X. Dambe 5: A comprehensive software package for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* **2013**, *30*, 1720–1728. [[CrossRef](#)]
47. Lanfear, R.; Frandsen, P.B.; Wright, A.M.; Senfeld, T.; Calcott, B. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **2017**, *34*, 772–773. [[CrossRef](#)] [[PubMed](#)]
48. Stamatakis, A. Raxml version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]
49. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)]
50. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior summarisation in bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [[CrossRef](#)]
51. Zhang, C.; Sayyari, E.; Mirarab, S. Astral-III: Increased scalability and impacts of contracting low support branches. In Proceedings of the RECOMB International Workshop on Comparative Genomics, Barcelona, Spain, 4–6 October 2017; pp. 53–75.
52. Sayyari, E.; Mirarab, S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* **2016**, *33*, 1654–1668. [[CrossRef](#)]
53. Schmidt, H.A.; Strimmer, K.; Vingron, M.; von Haeseler, A. Tree-puzzle: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **2002**, *18*, 502–504. [[CrossRef](#)] [[PubMed](#)]
54. Shimodaira, H.; Hasegawa, M. Consel: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* **2001**, *17*, 1246–1247. [[CrossRef](#)] [[PubMed](#)]
55. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **2002**, *51*, 492–508. [[CrossRef](#)] [[PubMed](#)]
56. Kishino, H.; Hasegawa, M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J. Mol. Evol.* **1989**, *29*, 170–179. [[CrossRef](#)] [[PubMed](#)]

57. Shimodaira, H.; Hasegawa, M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **1999**, *16*, 1114–1116. [[CrossRef](#)]
58. Dos Reis, M.; Yang, Z. Approximate likelihood calculation on a phylogeny for bayesian estimation of divergence times. *Mol. Biol. Evol.* **2011**, *28*, 2161–2172. [[CrossRef](#)] [[PubMed](#)]
59. Jin, F. Late mesozoic Acipenseriformes (osteichthyes: Actinopterygii) in central asia and their biogeochemical implications. In *Sixth Symposium on Mesozoic Terrestrial Ecosystems and Biota: Short Papers*; China Ocean Press: Beijing, China, 1995; pp. 15–21.
60. Benton, R. Fossil butte national monument kemmerer, wyoming: Twenty years of visitor service and research. *Rocks Miner.* **1993**, *68*, 180–182. [[CrossRef](#)]
61. Inoue, J.G.; Miya, M.; Venkatesh, B.; Nishida, M. The mitochondrial genome of indonesian coelacanth *Latimeria menadoensis* (sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. *Gene* **2005**, *349*, 227–235. [[CrossRef](#)] [[PubMed](#)]
62. Van de Peer, Y.; Mizrahi, E.; Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **2017**, *18*, 411–424. [[CrossRef](#)] [[PubMed](#)]
63. Zeng, L.; Zhang, N.; Zhang, Q.; Endress, P.K.; Huang, J.; Ma, H. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytol.* **2017**, *214*, 1338–1354. [[CrossRef](#)]
64. Spinks, P.Q.; Thomson, R.C.; Lovely, G.A.; Shaffer, H.B. Assessing what is needed to resolve a molecular phylogeny: Simulations and empirical data from emydid turtles. *BMC Evol. Biol.* **2009**, *9*, 56. [[CrossRef](#)] [[PubMed](#)]
65. Rokas, A.; Williams, B.L.; King, N.; Carroll, S.B. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **2003**, *425*, 798–804. [[CrossRef](#)] [[PubMed](#)]
66. Birstein, V.J.; Bemis, W.E. How many species are there within the genus *Acipenser*? *Environ. Biol. Fishes* **1997**, *48*, 157–163. [[CrossRef](#)]
67. Chen, M.Y.; Mao, R.L.; Liang, D.; Kuro-o, M.; Zeng, X.M.; Zhang, P. A reinvestigation of phylogeny and divergence times of hynobiidae (Amphibia, Caudata) based on 29 nuclear genes. *Mol. Phylogenet. Evol.* **2015**, *83*, 1–6. [[CrossRef](#)] [[PubMed](#)]
68. Smith, A.G.; Smith, D.G.; Funnell, B.M. *Atlas of Mesozoic and Cenozoic Coastlines*; Cambridge University Press: Cambridge, UK, 2004; ISBN 052-145-155-8.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).