



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

Department of Biology

Ph.D. COURSE IN: BIOSCIENCES

CURRICULUM: EVOLUTION, ECOLOGY AND CONSERVATION

SERIES XXXV

GENETIC AND GENOMIC INVESTIGATIONS FOR THE CONSERVATION OF STURGEONS

Coordinator: Prof. Romualdi Chiara

Supervisor: Prof. Leonardo Congiu

Ph.D. student: Stefano Dalle Palle

CONTENTS

	Pag.
LIST OF MANUSCRIPTS	2
ABSTRACT	3
1. GENERAL INTRODUCTION	5
1.1 PREAMBLE	5
1.2 CONSERVATION GENETICS AND GENOMICS	7
1.3 ACIPENSERIFORMES AND CONSERVATION STATUS	9
1.4 MICROSATELLITE IN POLYPLOIDS STURGEONS MANAGEMENT	11
1.5 INTRONS AS MOLECULAR MARKERS	13
2. AIM OF THE WORK	14
3. RESULT AND DISCUSSION	15
3.1 ASSESSMENT OF THE SEGREGATION PATTERN IN THE ADRIATIC STURGEON	15
3.2 GENETIC CHARACTERISATION OF A CAPTIVE STOCK OF ADRIATIC STURGEONS	18
3.3 GENETIC CHARACTERISATION OF A WILD STOCK OF CHINESE STURGEONS	23
3.4 INVESTIGATION OF MULTI-LOCUS INTRON POLYMORPHISMS (MIPS) POTENTIALITY IN TELEOST SPECIES	27
3.5 PRELIMINARY INVESTIGATION OF MULTI-LOCUS INTRON POLYMORPHISM (MIPS) IN STURGEON SPECIES	32
3.6 PRELIMINARY RESULTS OF THE <i>IN VIVO</i> STUDY TO EVALUATE THE GENETIC LOAD PREDICTIONS IN THE ADRIATIC STURGEON	36
4. CONCLUSIONS	41
ANNEX A	43
ANNEX B	60
ANNEX C	72
ANNEX D	81
ANNEX E	117
LIST OF REFERENCES	141
ACKNOWLEDGEMENTS	148

LIST OF MANUSCRIPTS

The main results of the present thesis are reported in five manuscripts, three of which were published in international journals, one under review and one submitted. The five papers listed here are reported in the Annexes.

Annex A

Dalle Palle S., Boscari E., Bordignon S. G., Muñoz-Mora V. H., Bertorelle G., Congiu L. (2022).

Different Chromosome Segregation Patterns Coexist in the Tetraploid Adriatic Sturgeon *Acipenser naccarii*.

Diversity, 14(9), 745. doi.org/10.3390/d14090745

Annex B

Barca F., Dalle Palle S., Schiavon L., Samassa C., Castaldelli G., Boscari E., Congiu, L. (2022).

Characterization of Captive Breeders to Preserve the Residual Genetic Diversity of Adriatic Sturgeon (*Acipenser naccarii*).

Diversity, 14(10), 829. doi.org/10.3390/d14100829

Annex C

Boscari E., Wu J., Jiang T., Zhang S., Cattelan S., Wang C., Du H., Li C., Li J., Ruan R., Xu J., Zheng Y., Dalle Palle S., Yang J., Congiu L., Wei, Q. (2022).

The last giants of the Yangtze River: A multidisciplinary picture of what remains of the endemic Chinese sturgeon.

Science of The Total Environment, 843, 157011. doi.org/10.1016/j.scitotenv.2022.157011

Annex D

Boscari E., Dalle Palle S., Vitulo N., Scapolatiello A., Schiavon L., Cariani A., Papetti C., Zane L., Marino I. A. M., Congiu L.

Multi-locus Intron Polymorphism (MIP) markers in species identification and population genomics: the first ‘universal’ panel for Teleosts.

Manuscript under review (Methods in Ecology and Evolution)

Annex E

Congiu L., Striebel-Greiter B., Gessner J., Boscari E., Boner M., Dalle Palle S., Ludwing A.
Identification and tracking of sturgeons and paddlefish products in trade: implications for trade control and biodiversity management.

Manuscript submitted (Aquaculture)

ABSTRACT

The order of Acipenseriformes, counting a total of 27 species, is one of the most endangered taxa. A paddlefish species *Psephurus gladius* and another sturgeon species *Acipenser dabryanus* have been recently declared respectively ‘Extinct’ and ‘Extinct in the Wild’ according to the last IUCN Red List assessment. Twenty species of sturgeon are classified as ‘Endangered’ or ‘Critically Endangered’, while the remaining five species are listed as ‘Vulnerable’. Therefore, these populations require adequate and timely conservation plans, both *in situ* and *ex situ*, that need to be informed by relatedness analysis and genetic variability characterisation.

The present Ph.D. thesis will mainly focus on the Adriatic sturgeon (*Acipenser naccarii*), a tetraploid species endemic to the Adriatic region, and historically present in the Po Rivers and its tributaries. Recently, its status has been updated from ‘Possibly extinct in the Wild’ to ‘Critically Endangered’ on the IUCN Red List. However, the preservation and reintroduction in the wild of these species is based on correct *ex situ* management.

To preserve the highest possible level of genetic variability in future generations, the sexually mature farmed progeny (F1) must be allocated to their parents of wild origin (F0) to identify groups of sibs and inform future breeding plans. To this purpose, microsatellite *loci* were used, as they are still the best markers for studying polyploid species, because of their higher allelic diversity and less demanding genotyping investments. However, before applying microsatellite markers to parental allocations, segregation patterns at these *loci* must be previously assessed. In fact, different possible inheritance patterns are possible in tetraploids, namely, disomic and tetrasomic. Segregation analyses of allele transmission were performed in known families composed of mother and father and some tens of fingerlings, allowing for the first time the observation of two coexisting patterns of inheritance in the Adriatic sturgeon.

A similar microsatellite-based relatedness analysis was then applied to another species of sturgeon on the brink of extinction, the Chinese sturgeon (*Acipenser sinensis*), to estimate the number of minimum breeders who participated in the only significant breeding event that occurred in the last 10 years. Even in this case microsatellite *loci* resulted to be powerful molecular markers for parental inferences, allowing to estimate an approximative number of 10 males and 10 females remaining for this species.

The Adriatic sturgeon was also included in a PRIN project called ENDEMIXIT (<https://endemixit.com/>) whose aim is to apply conservation genomics to five endangered Italian endemics. Given the possibility of conducting controlled reproductions, this species was selected to test *in vivo* the potentiality of genomic genetic load predictions, evaluating the relations between the load of bioinformatically identified deleterious mutations and the fitness of the offspring of selected parents. For this task, only preliminary results are currently available, as the data set needs to be cleaned of inaccuracies and biases with respect to the correct identification of genotypes in this tetraploid species.

Another relevant contribution of genetics to the management of captive stocks is the development of suitable markers for the identification of interspecific hybrids, which occur frequently in captive stocks due to the promiscuity between aquaculture for commercial production and aquaculture for conservation. The possibility of using the polymorphisms of intronic regions was explored to develop a new category of multi-locus markers (MIPs, Multi-

locus Intron Polymorphisms) that will allow the detection of multispecies contribution also in hybrids after the first generation. This study followed a previous investigation on Teleosts in which, thanks to the numerous available genomes, a first highly transferable panel of primers for intronic regions amplification was identified, and a protocol for simultaneous amplification followed by sequencing on the Illumina platform was optimised. Introns proved to be promising markers at both species and population levels in different Teleost species. For this reason, the same approach was applied to sturgeons, taking advantage of the recently available genome of *Acipenser ruthenus* and the available transcriptomes of four species: *Acipenser dabryanus*, *Acipenser fulvescens*, *Acipenser naccarii*, and *Acipenser stellatus*. MIPs markers designed for sturgeons promise to be highly versatile with many possible applications from individuals to populations levels in detecting differentiation signals. An important tool for the analysis of genetic diversity *in situ* and *ex situ* conservation plans, and the identification of species and hybrids to contrast the illegal trade of surgeon-derived products.

1. GENERAL INTRODUCTION

1.1 PREAMBLE

The present work arises from the need to implement and reinforce conservation strategies for the protection of species belonging to the order Acipenseriformes, one of the most endangered taxa. Overfishing, habitat degradation, pollution, and competition with alien species have caused the decline of sturgeon and paddlefish populations worldwide (Congiu et al., 2023).

An Italian endemic representative of this taxa is the Adriatic sturgeon (*Acipenser naccarii*). In 1977 a breeding programme started with the transfer of immature wild individuals (F0) from the Po River to the private aquaculture plant ‘Azienda Agricola V.I.P.’ (Orzinuovi, Brescia, Italy). After the first successful reproduction in captivity in 1988, several F1 brood stocks have been obtained by randomly paired F0 breeders and released (Arlati and Poliakova, 2009). Since then, some recaptures have occurred even of possibly wild individuals, but until now, this species is still far from complete recovery. The F0 stock, now reduced to a few individuals, represents the last animals of clear wild origin, and its genetic variability has been explored by Congiu et al. (2011). Additionally, the genetic variability of F1 stocks has been investigated, showing a reduction compared to F0 breeders (Boscari and Congiu, 2014), probably due to the lack of genetic support during past reproductions. The future of this species now depends on the correct *ex situ* management of F1 individuals, with special attention to maximising the level of genetic variation transmitted to future generations. For these purposes, microsatellite *loci* remain useful markers for polyploid species such as sturgeons, given the possibility to observe more than two alleles per *locus*, facilitating relatedness inferences necessary to breeding plans (Boscari et al., 2014).

This same approach is not limited to the management of captive stocks, but it can be used for relatedness analyses in wild individuals and estimation of the number in wild reproductions. This is the case of the ‘Critically Endangered’ Chinese sturgeon (*Acipenser sinensis*). In fact, the findings of 3000 Chinese sturgeon juveniles in the Yangtze estuary were the most relevant spawning signals in the last 10 years, but for which no information was available about the number of breeders and their genetic make-up (Wu et al., 2015).

The Adriatic sturgeon has also been included in the PRIN ENDEMIXIT project (<https://endemixit.com/>), which aims to apply genomics to the study of five endangered Italian endemics. This project involving collaborations among diverse Italian universities aims to understand the effects of genetic load in small populations and its impact on fitness and risk of extinction. Given the possibility of performing controlled crosses in *ex situ* managed Adriatic sturgeon, deleterious mutations were predicted in selected breeders after complete genome resequencing and then Genotyped-by-sequencing (GBS) in their offspring. The purpose is to test the association between load bioinformatic expectation and observed fitness indicators and to investigate the effect of inbreeding in this tetraploid species.

Another important aspect of sturgeon conservation is the identification of species and interspecific hybrids. Individuals can hybridise between different species, and many sturgeon hybrids are fertile. This possibility becomes particularly relevant when different types of interspecific hybrids, backcrosses with the parental species, and multispecies hybrids are generated in aquaculture (Havelka et al., 2011). These animals are very difficult to recognise from pure species, especially in the case of backcrosses, and since the animals in aquaculture

plants are often not kept separate, there is a high risk of accidental admixture. Therefore, it is appropriate that the numerous restocking programmes that are based on the release of aquaculture animals are supported by adequate genetic investigations. The development of markers for this purpose is still ongoing, and, especially for second- or third-generation hybrids, it is necessary to identify multi-locus variable nuclear markers that allow to increase the chance of detecting all the species involved in a hybrid.

This aspect has been addressed by developing a new category of markers based on the hypervariability of intronic sequences, taking advantage of the Next Generation Sequencing (NGS) techniques and the increasingly available genomic resources. Introns are flanked by conserved exonic regions, and therefore it is possible to design transferable primer pair panels with multi-species application, which can be used to characterise intron sequence variability through PCR-based library preparations and Genotyping-by-sequencing techniques (Ortega and Maldonado, 2020; Forcina et al., 2021).

A first exploration of this approach was applied in the wide taxonomic groups of Teleost, which count numerous species and for which several genomes are available. These resources will permit the bioinformatic identification of introns with lengths suitable for NGS techniques and flanked by highly conserved exons. Furthermore, due to the recent publication of the *de novo* genome assembly of the sterlet sturgeon *Acipenser ruthenus* (Du et al., 2020) and the available transcriptomes for other four species of sturgeons, it can be possible to identify and isolate introns in *Acipenseriformes*, leading to the identification of markers for species identification and possibly population-level applications, giving new tools to contrast illegal trade and strengthening conservation strategies.

1.2 CONSERVATION GENETICS AND GENOMICS

Considering the rapid loss of species affecting many different taxa in the last century, it is currently accepted that Earth is now facing the sixth mass extinction (Ceballos et al., 2015; Cowie et al., 2022). Although mass extinctions can be argued to have occurred naturally in the past, there is no doubt that the current decline of species finds its causes in the recent impact of human activities. Furthermore, such a loss of biodiversity would take millions of years to recover, affecting in the immediate future the well-being of humanity itself due to environmental degradation and the loss of important ecosystem services (Ceballos and Ehrlich, 2018). Apart from threats derived from human activities, declining populations are more subject to the effect of demographic and genetic stochastic processes, such as genetic drift, which leads to a random loss in genetic variability. Genetic variants are the fuel for adaptive evolutionary change; thus, a group of individuals with a low level of genetic variability could show a reduced evolutionary potential that acquires even more importance to cope with changing environments (Frankham et al., 2002). Moreover, these stochastic processes can overcome the efficacy of natural selection, possibly leading to the fixation of mildly deleterious mutations (Hedrick, 2001). Finally, in small populations, loss in genetic variability and higher probability of mating among relatives enhance the level of homozygosity, increasing the number of expressed deleterious mutations leading to inbreeding depression and fitness reduction (Keller and Waller, 2002). For all these reasons, maintaining the highest possible level of heritable genetic variation is a prerequisite for the success of species conservation plans.

In the last decades, molecular and genetic techniques have been widely applied and combined with ecology and conservation biology for the study of relevant aspects necessary to the correct conservation of species and biodiversity in a discipline called conservation genetics (Hedrick, 2001; DeSalle and Amato, 2004; Pertoldi et al., 2007).

Conservation genetics relied on neutral markers characterised in both nuclear and mitochondrial DNA through different methodologies. Techniques such as RAPD (Randomly amplified polymorphic DNA), RFLP (Restriction fragment length polymorphisms), AFLP (Amplified fragmented length polymorphisms), microsatellites, DNA sequencing, and SNP identifications were the most used for genetic characterisation (DeSalle and Amato, 2004).

Meanwhile, the advance of sequencing techniques has improved the capacity to explore progressively larger regions of genomes and the computational improvement to process and analyse a greater number of *loci*, paving the way to conservation genomics.

Compared to earlier conservation genetics methods, based on few neutral markers, genome-wide typing of individuals can be used to estimate, with smaller errors, parameters like heterozygosity and inbreeding coefficients (Ouborg et al., 2010). Most importantly, sequencing of large functionally relevant fragments of the genome allows conservation genomics to become closer to adaptive genetic variation that matters in species preservation (Allendorf et al., 2010; Flanagan et al., 2018). From genomic data, it is possible to estimate the genetic load based on bioinformatic predictions considering the predicted negative fitness effect of mutations. Genetic load can be subdivided in masked load and realised load. The masked load quantifies the potential loss of fitness due to deleterious mutation that may be expressed in the future, depending on population demography, while the realised load reduces the fitness of individuals and populations in the current generation (Bertorelle et al., 2022).

Deleterious mutations, and thus the genetic load, can be inferred by aligning conserved genomic regions across diverse lineages, with the assumption that novelties in these regions have probably deleterious effects. The advantage of these methods is that they do not require a reference genome and annotated genes. One of the most widely used comparative approaches, especially in wild animal studies, is Genomic Evolutionary Rate Profiling (GERP). GERP identifies regions that show a nucleotide substitution deficit compared to neutral ones and measures these deficits as ‘rejected substitutions’. Rejected substitutions reflect the intensity of past purifying selection and are used to rank and characterise constrained elements (Cooper et al., 2005). Thus, a high GERP score identifies regions that seem not to tolerate substitution through phylogeny. However, changes in the selection pressures of the sites considered throughout the phylogeny can limit the prediction reliability of these comparative genomic approaches. For example, the presence of lineage-specific substitutions could reflect this turnover in selective pressures and should be evaluated with care (Huber et al., 2020).

Other methods to predict the effects of mutations rely on direct inference based on information from biochemical studies or functional annotations. This information can be categorised as synonymous or non-synonymous, and eventually as low, moderate, or high impact, depending on the type of change. Softwares like SnpEff provide even more details about the effects of the different mutations, and specific variants such as missense, frameshift, stop-gain, and stop-loss can also be predicted (Cingolani et al. 2012). Experimental validations to assess the fitness effects of specific mutations have been performed in well-known model organisms such as *Drosophila melanogaster* (Mukai et al. 1972). Furthermore, a meta-analysis in livestock showed a correlation between increased genomic inbreeding measurement and a median decrease in phenotypic value (i.e., reproduction/survival, weight/growth, conformation, production, and health) (Doekes et al., 2021). Rarer are empirical evaluations for non-model organisms, and even in wild animal studies it would be very useful to connect genomic bioinformatic predictions and experimental validation of fitness effects, allowing more accurate estimates of both the masked and realised load in wild populations (Bertorelle et al., 2022).

For this reason, information on individual life-history traits and, when feasible, *in vivo* experiments are interesting opportunities to confirm *in silico* predictions of genomic approaches and identify mutations significantly associated with metabolic malfunctions and fitness reduction.

1.3 ACIPENSERIFORMES AND CONSERVATION STATUS

Acipenseriformes are an ancient group of fish distributed throughout the northern hemisphere and probably diverged from the actual actinopterygian fish around 345 (295–400) Ma (Du et al., 2020). Evolutionarily speaking, these species present some interesting characteristics; they are considered “Living Fossils” since they exhibit little morphological change if compared to fossil records dated between 250 and 200 Ma. Combined with this morphological stasis, a relatively lower rate of mutation in mitochondrial and nuclear genes and a lower rate of protein evolution are observed in these species (Krieger et al., 2002; Du et al., 2020). These animals are also characterised by distinct levels of ploidy due to several whole genome duplication events that occurred both in their common ancestor and in different branches of this taxonomic group (Kim et al., 2005; Peng et al., 2007).

Unfortunately, this order that counts 27 species is also one of the most endangered taxonomic groups. Recently, an IUCN Red List assessment carried out by the IUCN Sturgeon Specialist Group revealed a further decline in paddlefish and sturgeon populations (Congiu et al., 2023). The Chinese paddlefish (*Psephurus gladius*), endemic to the Yangtze River and considered functionally extinct by 1993, has not been seen since 2003 and is stated as ‘Extinct’ according to the criteria of the IUCN Red List (Zhang et al., 2020). Endemic to the same basin, the Dabry sturgeon (*Acipenser dabryanus*) is classified as ‘Extinct in the Wild’, since all individuals recaptured in nature originated from restocking programmes. The eight European species, despite being under protection since 1992, are classified either as ‘Endangered’ or ‘Critically Endangered’ by the new update of the IUCN Red List. The North American Acipenseriformes present a better situation than the Eurasiatic ones, probably due to more timely conservation efforts, but even here the conservation status of Acipenseriformes species worsened (Congiu et al., 2023).

The causes of Acipenseriformes' decline are well known and are still a current issue for these species. Deterioration and pollution of their original habitats play an important role in affecting reproduction and larvae survival (Williot et al., 2002). Most sturgeons are anadromous species that need to return to a freshwater basin to spawn. Tolerance to salt water, when present, is considered a secondary adaptation to exploit estuarine food resources, since eggs and larvae are physiologically intolerant to high levels of salinity (Sulak and Randall, 2002). The preferred spawn locations differ in depth and current velocity depending on the species and are characterised by hard substrates of pebbles, gravel, or rocks (Billard and Lecointre, 2000). These substrates provide a surface where eggs can adhere and repair to larvae in the first days after hatching (Českleba et al., 1985; Nguyen et al., 2006), although different larvae behaviours are observed depending on the species (Billard and Lecointre, 2000; Kim et al., 2020). In this context, the increasing construction of dams and other barriers are major disturbing factors to wild reproductions because such structures can prevent migration to the spawning sites or even be the cause of their complete disruption (Williot et al., 2002).

Another main factor of sturgeons' decline was their overexploitation for their valuable caviar and meat. Their life-history traits such as late maturation, longevity, and fertility increasing with body size resulted to be adaptive for millions of years (Billard and Lecointre, 2000), but now this characteristic could result unfavourable to cope with the recent anthropic pressures that target the largest individuals.

Since 1 April 1998, all the Acipenseriformes species have been listed under CITES Appendix I or II, resulting in better control of the international trade of these species (Raymakers, 2002). Despite the CITES regulation and the increasing amount of meat and caviar produced in aquacultures (Bronzi et al., 2019), poaching and illegal trade remain serious issues for these species, affecting the recovery of wild populations (Ludwig et al., 2015; Mirrasooli et al., 2019). In this context, molecular markers are essential for the forensic analysis needed to contrast illegal trades and commercial frauds in one of the most lucrative food markets (Ludwig, 2008). Due to the ability of these species to make vital hybrids, nuclear markers should be preferred to mitochondrial ones, as the latter provides only the genetic information of maternal contribution. Moreover, the technique should be relatively inexpensive, easy to use, and fast for routine investigations. To these purposes, quick and easy techniques and protocols have been implemented for the identification of various species of sturgeons exploiting the variability present in the intronic regions of the nuclear S7 ribosomal protein (RPS7) and the nuclear S6 ribosomal protein (RP2S6) (Boscari et al., 2014; Boscari et al., 2017b). Molecular markers are also important tools in genetic and genomic studies to assess the level of genetic variability, inbreeding, and genetic load present in the individuals and populations of these endangered species. With such information, better *in situ* and *ex situ* conservation plans can be organised, enhancing the probability of success (Frankham et al., 2002; Allendorf et al., 2010). Fortunately, there are positive signs that sturgeons can take advantage of adequate protection measures. The ship sturgeon (*Acipenser nudiiventris*) was observed in the Rioni River after years of absence (Beridze et al., 2021), as well as the shovelnose sturgeon (*Pseudoscaphirhynchus hermanni*), reappeared in the Amu Darya River after 20 years (Sheraliev et al., 2021). Even the Adriatic sturgeon has been classified from 'Extinct in the Wild' to 'Critically Endangered' after finding juveniles linked to sporadic natural reproduction in the Po basin, an indicator that long-term conservation efforts for this species may be successful (Congiu et al., 2021).

1.4 MICROSATELLITE IN POLYPLOIDS STURGEONS MANAGEMENT

In sturgeons, genetic analyses need to deal with complex genomes and different levels of ploidy, due to diverse events of whole genome duplication. The first event of duplication took place in the common ancestor possessing 60 chromosomes. Secondary events of whole genome duplication then occur in the Pacific and Atlantic clades, leading to a total of 240 chromosomes (Peng et al., 2007). Finally, a third event led to the unique number of 360 chromosomes observable in *Acipenser brevirostrum* (Kim et al., 2005). The number of chromosomes associated with the different levels of ploidy has been the subject of an extensive debate between the two main positions. The first argues that since the condition with 120 chromosomes is derived from a duplication event in the common ancestor, species with 120 chromosomes must be considered tetraploid, and those with 240 chromosomes octoploid. The second focusses more on its functional aspect attributing to species with 120 and 240 chromosomes a condition of diploidy and tetraploidy, respectively (Fontana et al., 2007).

The two positions are compatible and both correct, as explained by Fontana et al. (2008). If the first defines the degree of nominal ploidy based on the number of duplication events, the second focusses on the functionality of the genome and defines the functional ploidy based on the number of copies of some genes observed to be active with cytogenetic approaches. Adding a layer of complexity, polyploidization can originate through the fusion of unreduced gametes at the intraspecific or interspecific level. In tetraploids, for example, these events lead to two types of condition, called autotetraploidy and allotetraploidy, respectively. In complete autotetraploids, there are four complete homologous chromosomes, and any random pair of bivalents and quadrivalents are possible during meiosis. This condition leads to tetrasomic inheritance, which means that all allelic combinations inside the gametes are possible (Stift et al., 2008). On the contrary, in complete allotetraploids, each tetrad is composed of two sets of diversified homoeologous chromosomes. Homeologous chromosomes cannot form pairs, leading to disomic inheritance, where only four out of six allelic combinations are possible (Stift et al., 2008). Anyway, these are two extreme conditions, and intermediate conditions are observable when chromosomes have different degrees of preferential pairings. For example, cytological diploidization can be promoted by selection against multivalent pairings that can have negative impacts on fertility and karyotype stability. Instead, in allopolyploids, homeologous chromosomes from two distinct species could maintain some degree of genetic affinity that allows homeologous pairing (Soares et al., 2021).

Despite the improvement of NGS techniques and the advancement in genomic analysis, some difficulties remain in the analysis of polyploid species. In general, it is more challenging to reconstruct the haploid genome of a polyploid organism. This is particularly true for sturgeons, where the assembly of the haploid genome of the functional diploid sterlet sturgeon *Acipenser ruthenus* revealed a complex structure with a different degree of homeology among the 60 chromosomes, which originated from the segmental rediploidization process that occurred in this species (Du et al., 2020).

Furthermore, genetic and genomic analyses relying on the characterisations of several biallelic SNPs, which are the most abundant even in polyploids, are complicated due to the difficulties in inferring their correct genotypes and their segregation patterns (Dufresne et al., 2014; Meirmans et al., 2018). In particular, inference of reliable genotypes in polyploids requires higher reads depth compared to those needed in diploids (Matias et al., 2019; Gemenet et al.,

2020), increasing the costs and lowering the number of samples or markers that can be characterised using Genotyping-by-sequencing techniques.

In this context, highly variable microsatellites *loci* remain useful markers to explore the genetic variability of polyploid species such as sturgeons. Microsatellite markers generally show more than two alleles per *locus*, thus can be observed in complete heterozygosis in polyploid species. This characteristic also facilitates inference of segregation patterns that are necessary information in polyploids to reach a correct conclusion during parental and population genetic analyses (Boscari et al., 2011; Boscari et al., 2014) and are useful to shed light on the level of functional ploidy level across the different species of sturgeons (Rajkov et al., 2014). For these reasons, microsatellites are still used and isolated, even with the help of Next Generation Sequencing techniques (Hu et al., 2019; Wang et al., 2022), to perform the studies necessary for the correct management of genetic variability in sturgeon species in both *in situ* and *ex situ* conservation plans (Boscari et al., 2021; White et al., 2021).

1.5 INTRONS AS MOLECULAR MARKERS

Introns are the non-coding regions of DNA present in all eukaryotic genomes and flanked by the more conserved and coding exon sequences. Although introns are functional regions involved in splicing and other processes related to gene expression, their mutation rate is similar to the neutral portions of genomes, and flanking conserved exons are suitable for designing primer transferable in a variety of different species (Ortega and Maldonado, 2020; Forcina et al., 2021).

Introns have been already investigated in the past as molecular markers, and the pioneer study involved the use of EPIC-PCR (Exon-Primed Intron-Crossing PCR) to amplify these portions of DNA with primer pairs design on the flanking exons (Lessa, 1992).

Past studies focused mainly on intronic *loci* length polymorphisms (Côte-Real et al., 1994; Daguin and Borsa, 1999) or intronic single-locus sequencing on a limited number of species (Palumbi and Baker, 1994; Chow and Hazama, 1998). The potentiality of introns has been explored even in sturgeons, resulting in the development of species identification markers currently used in forensic analysis (Boscari et al., 2014; Boscari et al., 2017b).

Thanks to increasing advances and affordability of NGS (Next Generation Sequencing) technologies and methodological improvement in library preparation, new Genotyping-by-sequencing techniques are developed and employed in conservation genomics. In fact, despite the reduction in the price of sequencing techniques, the resequencing of several individuals remains a costly task. Instead, Genotyping-by-sequencing is feasible even in organisms without reference genomes, and amplicons can be easily indexed and pooled during library preparation offering the possibility to characterise hundreds or thousands of microhaplotypes at once in numerous samples (Schweizer et al., 2021). Microhaplotypes, defined as sets of multiple SNPs in a *locus* treated as a single haplotype, have proven to be suitable for many different purposes, such as stocks identification and relationship inference, hybrid identification, estimates of the effective population size, and estimates of heterozygosity and inbreeding (Schweizer et al., 2021; Osborne et al., 2023).

Given that, introns can be interesting genomic regions to identify multi-locus panels of variable microhaplotype-like markers. Starting from the available genomes and genes annotation, it is now possible to isolate bioinformatically introns flanked by highly conserved exons, where to design panels of primers for introns characterisation transferable even in those species for which genomic information is not yet available.

Finally, like other microhaplotype markers, intron markers promise to be suitable for a wide range of biological problems, from hybrid and species identifications at the individual level to the screening of genetic diversity at the population level (Camacho Sanchez et al., 2018; Forcina et al., 2021).

2. AIM OF THE WORK

Considering the conservation status of the Acipenseriformes species, it is extremely important to take timely actions for their preservation and recovery.

In this context, genetic investigations are a fundamental prerequisite to inform both *in situ* and *ex situ* management plans that should aim to maintain the highest level of genetic variation, and thus adaptive potentials. Moreover, the development of new markers for species and hybrid identification can greatly contribute to contrast illegal trade of sturgeon-derived products and to identify same-species conservation units.

The aim of this study is to apply different molecular techniques to solve different biological problems in this taxonomic group characterised by complex genomes and diverse levels of ploidy. Segregation patterns in *Acipenser naccarii* will be investigated and the kinship relationship among individuals managed *ex situ* will be solved using microsatellite markers. The same approach will be applied to the genetic characterisation of wild juvenile individuals of *Acipenser sinensis* and then to estimate the number of breeders that contribute to this last significant reproductive event.

A new approach based on the variability present at intronic *loci* will be explored first in Teleost and then in Acipenseriformes with the aim of identifying new molecular markers applicable for species identification and population studies.

Finally, a genomic approach to estimate the genetic load will be evaluated, comparing bioinformatic predictions with fitness observation in controlled crosses of *Acipenser naccarii*.

Project tasks

In detail, the following aspect will be examined:

1. Assessment of the segregation pattern in the Adriatic sturgeon. ([Annex A](#))
2. Genetic characterisation of a captive stock of Adriatic sturgeons. ([Annex B](#))
3. Genetic characterisation of a wild stock of Chinese sturgeons. ([Annex C](#))
4. Investigation of Multi-locus Intron Polymorphisms (MIPs) potentiality in Teleost species. ([Annex D](#))
5. Preliminary investigation of Multi-locus Intron Polymorphisms (MIPs) in sturgeon species. ([Annex E](#))
6. Preliminary results of the *in vivo* study to evaluate genetic load predictions in the Adriatic sturgeon.

3. RESULT AND DISCUSSION

The five publications that resulted from this study are summarised in the following paragraphs. The detailed original publications (or manuscripts) are annexed (Annexes A, B, C, D, and E).

3.1 ASSESSMENT OF THE SEGREGATION PATTERN IN THE ADRIATIC STURGEON Original publication in Annex A

Genetic analyses on sturgeons deal with complex genomes and various levels of ploidy, originated from independent events of whole genome duplication starting from a common ancestor with 60 chromosomes (Peng et al., 2007; Fontana et al., 2008; Kim et al., 2005). The levels of ploidy have been the subject of extensive debate between two main positions. The first argues that species with 120 and 240 chromosomes must be considered tetraploid and octaploid, respectively. The second position, considering the functional reduction of ploidy, attributes to the two groups a condition of diploidy and tetraploidy (Fontana et al., 2007). The two views use different criteria to define the nominal ploidy, the number of duplications in the first, and the functional activity of genes in the second and can be both considered correct and fully compatible (Fontana et al., 2008). In this context, the Adriatic sturgeon (*Acipenser naccarii*) with 240 chromosomes is generally considered a functional tetraploid (Fontana et al., 1999). This is also confirmed by most of the microsatellites analysed in this species, which mostly show up to 4 alleles per individual.

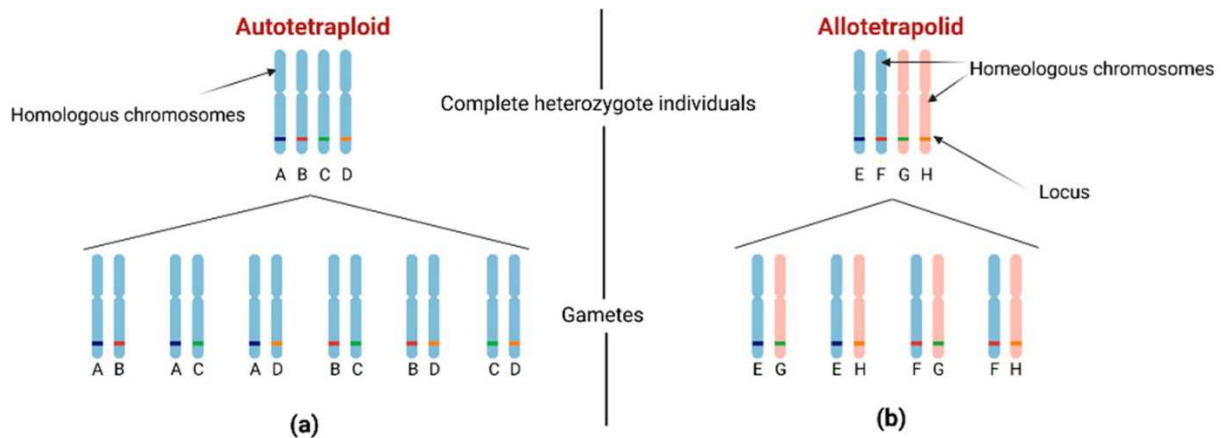


Figure 1. Expected segregation patterns under pure disomy (a) or tetrasomy (b), respectively expected in Autotetraploid and Allotetraploid genomes.

In general, polyploidization can originate through the fusion of unreduced gametes at intraspecific or interspecific levels, leading to two types of conditions in tetraploids, called autotetraploidy and allotetraploidy, which determine the segregation patterns of the alleles within gametes. In complete autotetraploids, there are always four homologous chromosomes, and random pairs of bivalents and quadrivalents are possible during meiosis (Stift et al., 2008), leading to tetrasomic inheritance, where all chromosome combinations within gametes are possible (Figure 1). Instead, in allotetraploids, each tetrad is composed of two homeologous

sets of homologous chromosomes, originating from the two parental species. In this situation, the homeologous chromosomes do not form pairs, and the only possible pairings is between the homologous, leading to double disomic inheritance, where only four out of six chromosome combinations are possible in the gametes (Stift et al., 2008); these are two extreme cases, and intermediate conditions are observable when the chromosomes have different degrees of preferential pairing.

In fact, the inheritance pattern may shift from tetrasomic to disomic or vice versa (Stift et al., 2008). For example, in autotetraploids, fertility and karyotype stability can be negatively impacted by imperfect multivalent pairing, thus promoting diploidization and consequent shifting to disomy. In contrast, in allotetraploids, homeologous chromosomes from two distinct parental species could maintain some degree of genetic affinity, allowing competition with the homologous pair during meiotic interactions and thus showing a certain degree of tetrasomic inheritance (Soares et al., 2021).

Understanding the mechanisms of chromosomal segregation in a tetraploid species can have important conservation implications, for example, when the development of parental allocation methods is required.

In the Adriatic sturgeon, allele segregation was previously investigated at only seven *loci* using microsatellite markers (Boscari et al., 2011). All *loci* showed a tetrasomic inheritance pattern, indicating the probable autopolyploid origin of this species. However, we took advantage of the recent availability of new complete family groups, some of which were obtained within the ENDEMIXIT project (<https://endemixit.com/>), and new 14 isolated and tested microsatellite *loci* not yet explored, to provide a deeper insight into the modality of chromosome segregation in this species. Each *locus* was considered informative only if at least one parent of a given family group was completely heterozygous and the other parent shared no more than one allele with the former, allowing to follow unambiguously the segregation patterns of alleles. To compare the null hypothesis of tetrasomy (H_0) with the disomic pattern and with different intermediate models of inheritance, a likelihood ratio test was applied following the procedure proposed by Stift et al. (2008). More details are reported in Annex A.

The segregation pattern observed at most *loci* confirms that the Adriatic sturgeon can be considered predominantly tetraploid with a tetrasomic inheritance pattern.

However, the tetrasomic segregation pattern was fully rejected for three newly analysed *loci*, indicating that the functional diversification process is at different stages in different parts of the genome, with most regions still functionally tetraploid and some others in which the degree of divergence has reached the condition of double diploidy. We also observed the presence of two *loci* with a marked tendency to disomy, pointing to a possible imperfect preferential pairing that is expected in intermediate stages of functional diploidization (Stift et al., 2008; Soares et al., 2021). Furthermore, some *loci* consistently showed more than 4 alleles in many individuals (Boscari et al., 2015); this could be due to duplication of the region of such microsatellites, but alternatively an incomplete process of functional reduction after genome duplication cannot be excluded, as already proposed also for other sturgeon species (Ludwing et al., 2001; Havelka et al., 2013).

These pieces of evidence show that the level of homology between chromosomes within the Adriatic sturgeon genome is likely to vary from chromosome to chromosome and that different parts of the genome may consequently have different degrees of functional ploidy (from two to

eight) and should be considered when characterising the genome of this species and probably of sturgeons in general.

Whatever the evolutionary origin and implications of our results, which can be better investigated only when the genome of this species will be available, the identification of different inheritance patterns at different markers may have practical consequences for the conservation of the Adriatic sturgeon. Our findings suggest that before developing parental allocation methods, a preliminary analysis of the *loci* used is recommended. This would ease the reconstruction of individual genealogies of animals kept as captive breeders and the reallocation of any individual recaptured after release. Another interesting and relatively unexplored aspect of *ex situ* conservation in which a clear knowledge of the inheritance patterns at different *loci* could be useful is monitoring the genomic impact of breeding protocols. Captive breeding is usually performed following standardised protocols of hormonal induction for eggs and milt release, and fertilisation is carried out in excess of sperm. Then, the resulting viable progeny is released without verifying if the procedure had some effect on their genomic asset, for example, by inducing aneuploidies, which are a common phenomenon in some captive sturgeon stocks (Schreier et al., 2021). The release in nature of genetically anomalous individuals should be avoided, as the consequences that this can have on the following generations and their reproductive efficiency are unknown and, in the case of animals with long generation times, could be revealed after decades. Random screening of familiar groups to verify correct segregation at both disomic and tetrasomic *loci* could significantly contribute to reducing the potential impact of genomic anomalies on natural populations.

3.2 GENETIC CHARACTERISATION OF A CAPTIVE STOCK OF ADRIATIC STURGEONS

Original publication in Annex B

The endemic Adriatic sturgeon (*Acipenser naccarii*) has been part of captive breeding programmes since 1977, starting with the captures of the last wild animals.

After more than 40 years, this wild F0 stock is now reduced to a few individuals, which represent the only living Adriatic sturgeons of certain wild origin. Now, correct management of their F1 progeny is essential for the preservation of this species, with special attention to maximise the level of genetic variation transmitted to future generations in order to maintain the highest possible level of long-term adaptive potential (Boscari et al. 2014).

To this purpose, part of the F1 individuals have been retained in captivity as future breeders by different public authorities and by the ‘Azienda Agricola V.I.P.’ itself. All Adriatic sturgeons bred in captivity in Italy and abroad descend directly from the F0 stock, but the precise genealogy of each individual and the consequent degree of kinship are still unknown for most of them.

Azienda Agricola V.I.P.’s stocks have already been investigated and resulted in a remarkably higher genetic diversity compared to other stocks established from a smaller number of sibling groups (Boscari et al. 2014; Congiu et al., 2011; Boscari and Congiu, 2014).

In this context, a new parental allocation is presented here with the aim of expanding previous investigations on the Azienda Agricola V.I.P.’s stocks, currently acquired by the company ‘Storione Ticino’, including in the analyses those individuals who reached sexual maturity in the meantime. Moreover, a higher number of microsatellite *loci* was used to increase the resolution power.

In particular, the genetic diversity of a total of 42 F0 individuals of wild origin (20 males and 22 females) and 505 individuals of first-generation (F1) of the ‘Storione Ticino’ aquaculture plant were analysed.

A significant fraction of these genotyped animals died in an unfortunate poisoning event some years ago. However, they were included in the analyses because many other animals released as juveniles in the past came from the same crosses. Therefore, a good representation of the diversity of all family groups generated in the last decades can facilitate the genetic tagging of recaptured individuals released by different restocking activities.

The final microsatellite panel included 15 *loci* selected among 24 genotyped, some of which have been excluded due to their disomic segregation patterns or other anomalies and interpretation problems. The parental allocation analyses were then implemented using the software ‘Breeding Sturgeon’, specifically developed for tetraploid sturgeon species (Boscari et al. 2014b).

Two different algorithms are followed: the standard band-sharing measure and a weighted estimation of segregating alleles from different possible parent pairs to their putative offspring. The first and simpler analysis compares the alleles observed in the multi-locus profile of each individual F1 with the ones of all possible pairwise combinations of F0 breeders, and if all F1 alleles are present in a pair of breeders, parental compatibility is accepted. The second method checks for the satisfaction of Mendelian segregation. In particular, for the three individuals under comparison (the two possible breeders and the F1 to be allocated), all possible tetraploid genotypes are inferred at each *locus*, and the inheritance of two alleles from each parent at all

loci is verified. This second approach may also accept the possibility of scoring errors or segregation anomalies tolerating up to two *loci* with imperfect inheritance. Based on the number of anomalies at these two *loci*, an allocation score is calculated and accepted when above the threshold of 0.866 (Boscari et al. 2014b). Both methods were used and the results were compared. Finally, the concordance of mothers and offspring was verified by sequencing and confronting mitochondrial D-loop haplotypes.

Family Group	Parental Pair	N F1 Tot (Alive)	mtDNA Haplotype	Family Group	Parental Pair	N F1 Tot (Alive)	mtDNA Haplotype
01	Pelvienne (F) × NaccS18 (M)	14 (5)	2	17	NaccS8 (F) × 740 (M)	21 (14)	2
02	NaccS8 (F) × Matto (M)	9 (6)	2	18	NaccS3 (F) × Matto (M)	5 (1)	3
03	NaccS7 (F) × NaccS6 (M)	12 (12)	3	19	NaccS33 (F) × 740 (M)	4 (2)	3
04	NaccS8 (F) × NaccS31 (M)	17 (15)	2	20	NaccS7 (F) × VerdeBis (M)	2 (1)	3
05	NaccS33 (F) × NaccS11 (M)	10 (4)	3	21	NaccS12 (F) × NaccS17 (M)	1 (1)	2
06	O2 (F) × Matto (M)	12 (9)	5	22	Indecisa (U) × Matto (M)	2 (2)	2
07	NaccS26 (F) × NaccS29 (M)	27 (18)	5	23	Indecisa (U) × AnS11 (M)	1 (1)	2
08	NaccS19 (F) × NaccS17 (M)	18 (17)	5/7	24	NaccS12 (F) × Matto (M)	1 (0)	2
09	NaccS28 (F) × NaccS17 (M)	51 (30)	3	25	NaccS3 (F)–NaccS29 (M)	1 (0)	3
10	NaccS16 (F) × NaccS23 (M)	39 (15)	3	26	NaccS16 (F) × 740 (M)	1 (1)	3
11	NaccS19 (F) × NaccS31 (M)	30 (21)	5/7	27	NaccS16 (F) × NaccS29 (M)	1 (1)	3
12	NaccS16 (F) × NaccS30 (M)	13 (2)	3	28	NaccS8 (F)–NaccS11 (M)	6 (6)	2
13	NaccS33 (F) × NaccS9 (M)	17 (12)	3	29	Pelvienne (F)–NaccS23 (M)	3 (3)	2
14	NaccS8 (F) × NaccS17 (M)	8 (8)	2	30	Non-allocated	141 (108)	-
15	NaccS12 (F) × NaccS27 (M)	10 (3)	2	31	Multi-allocated	7 (7)	-
16	NaccS3 (F) × 740 (M)	21 (6)	3				

Table 1 The family groups identified with parental allocation analysis based on 15 microsatellite *loci*. For each family group, the parent pair, the total number of allocated animals, and the number of survived ones (in brackets) are reported as well as the maternal mitochondrial haplotype. The families 08 and 11 are composed of F1 animals carrying the mitochondrial haplotypes 5 or 7, and in some cases, animals showed heteroplasmy with both the mitochondrial variants. This was due to the heteroplasmic condition of the shared F0 female.

A total of 357 F1, of which 216 are still alive, were unambiguously assigned to their parental pair, identifying 29 family groups (Table 1).

The contribution of the different breeders to the F1 stock was strongly unbalanced, with only four breeders contributing more than 40% of the still-alive F1. Unfortunately, the size of the families, already small in some cases, was further reduced after poisoning, resulting in two completely lost family groups and 11 composed of three or fewer individuals. Seven individuals remained multi-allocated (compatible with more than one parent pair), indicating that, although the allocation power is expected to increase by adding eight new *loci*, in some cases it may still not be sufficient.

Finally, 141 *A. naccarii* were not allocated, probably because one or both parents died before the first sampling occurred in 2004.

In this case, where the genotypes of potential F0 parents could be missing, an alternative method was evaluated. For this purpose, an estimation of pairwise genetic distances was performed on all F1 animals that shared the same haplotype (possibly full or half-sibs). Due to the tetraploid condition that precludes the possibility of assessing the exact allele dosage, we used the presence/absence information of the different alleles to estimate the distances of Sørensen (Sørensen, 1948) with the R package ‘Ade4’ (Thioulouse et al., 2018). The pairwise distances

were then visualized through Multi-dimensional scaling (MDS) performed with the R package “ggplot2” (Wickham, 2016).

In general, with the Multi-dimensional Scaling (MDS) approach, the results were consistent with previous parental allocations, with animals allocated to the same parental pair forming clear groups of similarity (Figure 2a-d). Families sharing one parent (half-sibs) also tended to cluster together. Moreover, when a second run of the same analysis was performed, investigating each of these clusters of half-sibs singularly, the underlying families became better distinguishable.

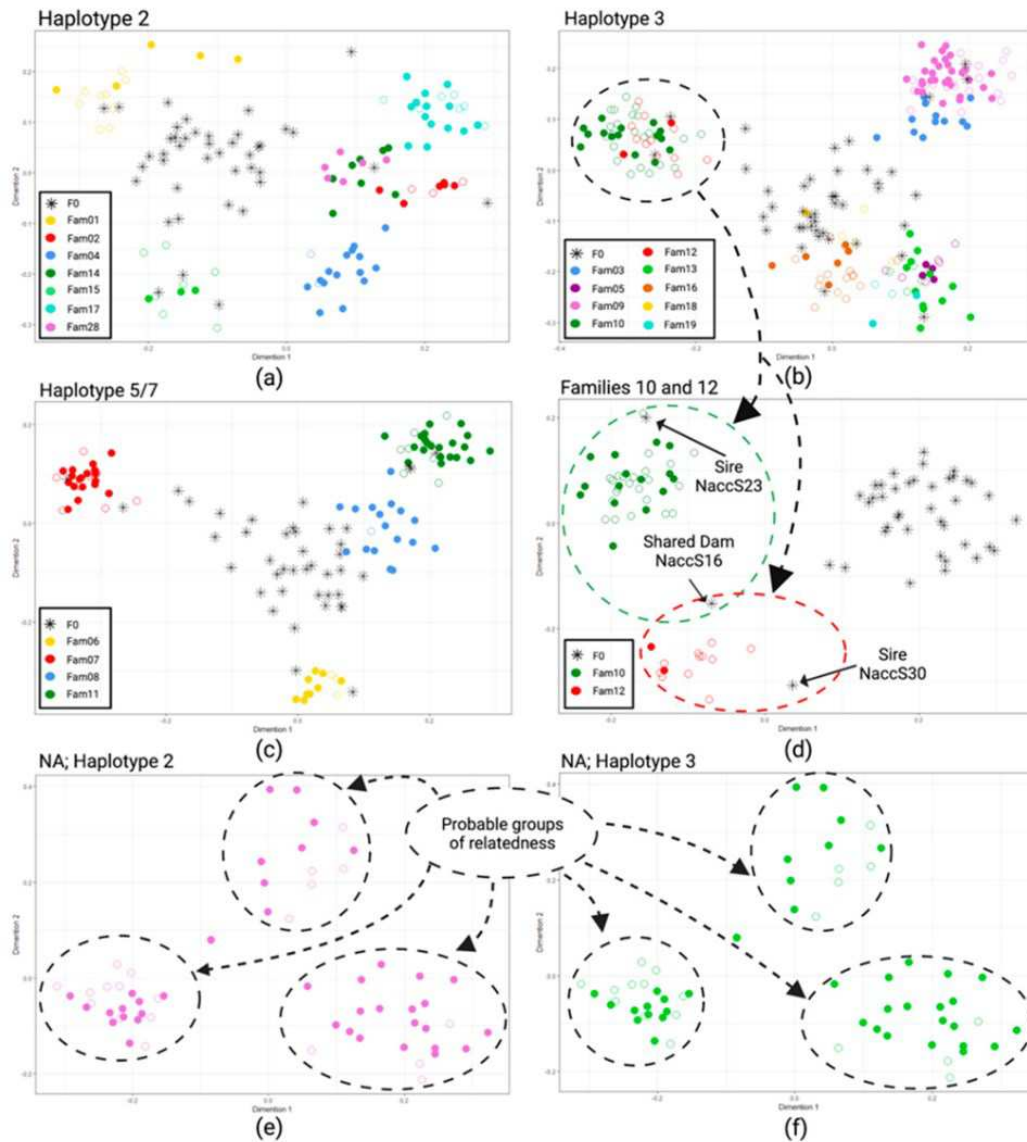


Figure 2 Results of MDS analyses performed on pairwise Sørensen’s distance matrix among F0 and F1 with mitochondrial haplotype 2 (a); F0 and F1 with mitochondrial haplotype 3 (b) and F0 and mitochondrial haplotypes 5 and 7 (c). Individuals of two half-sib family groups (10 and 12), that in the first run were not distinguishable, were included in a second run, excluding all other F1s to show the usefulness of a hierarchical approach in also distinguishing related families the common dam and the sires of the two families are also highlighted (d). The same approach was used on animals for which the parental information was not available, having mitochondrial haplotype 2 and 3 (e, f), respectively.

In all figures, full and empty circles represent alive and dead animals, respectively. NA = non-allocated.

Subsequently, the MDS approach was used in animals of unknown genealogy, after grouped them according to the mitochondrial haplotype (Figure 2e-f). Although it is not possible to

verify the results using paternity analyses, it is evident that there are groups of similarities, probably consisting of groups of siblings or half-siblings. Distinguishing full sibs from half-sibs can be difficult as expected, as the average distance within a family depends on the similarity of the two parents, which can also significantly vary across families. However, in this case, as was shown, a second MDS run, performed on individuals identified as somehow related in the first run, can provide a finer resolution. In the present study, MDS representations were applied to animals with the same haplotypes, allowing for the detection of full sibs and maternal half-sibs (sharing the same mother). However, in doing this, paternal half-sibs (sharing the same father) are not identified if the two mothers had different mitochondrial haplotypes, as they are not included in the same analyses. These types of full sibs could only be identified if all animals were analysed simultaneously. This was not possible, as the MDS analyses of all animals did not allow for the identification of multi-haplotype clusters. This limit was possibly due to the high number of families present in the stock and the consequent homogeneous distribution of pairwise Sørensen's distances. In any case, the method can be useful for preliminary analyses, which in the case of stocks descending from a low number of parents could also allow for the detection of paternal half-sibs.

Regarding the 'Storione Ticino' stocks, even if 29 families are apparently a low number based on the maximum possible families obtainable from 20 F0 males and 22 F0 females, the fact that 26 of 42 adults were reproduced in the past testifies a careful commitment to differentiate the breeders to maximise residual genetic diversity in the offspring. It should also be considered that some F0 individuals never became available for reproduction and that females can ovulate after every two, three or even more years (Tortonese, 1989). Furthermore, given that for 141 F1 individuals, no pairs of F0 parents were found to be compatible, we can reasonably assume that additional breeders were used in the past to generate these families and died before our sample collection, leaving their genetic heritage in their progeny and making them carriers of added value in terms of biodiversity. Unfortunately, some of the family groups were composed of very few individuals, and in some cases by a single animal that cannot be representative of the genetic diversity of their parents. Given the inconstant availability for reproductions, the possibility of permanently losing these genetic lines is relevant (Boscari et al. 2014). Thus, individuals belonging to these residual families should be considered a priority in planning future crossings.

However, there are other stocks of *Acipenser naccarii* (Cabrera-Castro et al., 2018), all descendants of the F0 stock analysed in this study. Therefore, there is hope that representatives of some under-represented or non-represented families are present elsewhere.

However, the stock of the 'Storione Ticino' farm confirmed the presence of a significant degree of genetic diversity (Boscari and Congiu, 2014). This highlights the relevance of saving an adequate number of animals from each crossing and the importance of distributing these animals of different genealogical lines in different plants to decrease the probability of disappearance due to accidental events.

To conclude, the analyses were simplified by the genotypes of most of the wild founders. However, the reliability of the methods used here to identify relatedness groups makes them useful for the analysis of all sturgeon stocks whose composition is unknown.

In general, the approaches adopted here are particularly suitable for polyploid species such as sturgeons, for which standard analyses based on genotype knowledge cannot be used. In fact,

in polyploid organisms a precise inference of the genotype is not always possible, forcing the use of phenotypic information on the presence/absence of the different alleles.

Furthermore, as previously demonstrated, some *loci* can follow a disomic segregation modality (Dalle Palle et al., 2022 – Annex A) indicative of further differentiation of chromosomes toward a diploid condition (Stift et al., 2008). In summary, the coexistence of different levels of ploidy in different regions of the genome challenges the development of analytical methods for sturgeon species belonging to the group with 240 chromosomes (Fontana et al., 2008). Therefore, relying on ‘phenotypic’ data remains the preferred approach for these animals, as will also be demonstrated in the Chinese sturgeon (*Acipenser sinensis*) (Boscari et al., 2022 – Annex C).

3.3 GENETIC CHARACTERISATION OF A WILD STOCK OF CHINESE STURGEONS

Original publication in Annex C

One of the biggest threats to sturgeon species is habitat degradation due to dams and other barriers that impair access or even cause the loss of reproductive sites (Congiu et al., 2023). This problem affects even the Yangtze River (China), the third longest river, which is spanned by the largest hydroelectric dams, the Gezhouba Dam, completed in 1981, and the Three-Georges Dam, completed in 2003 (Huang, 2019). Before the dams' construction, the Yangtze was inhabited by two sturgeon species, the Dabry sturgeon (*Acipenser dabryanus*) and the Chinese sturgeon (*Acipenser sinensis*), and one paddlefish species, the Chinese paddlefish (*Psephurus gladius*). After damming, dramatic loss of suitable hydro-ecological conditions for spawning strongly reduced reproductive activity, negatively affecting recruitment and leading to a worsening of the general conservation status of the two species of sturgeon now classified as 'Critically Endangered' (Zhang et al., 2013, Zhang et al., 2016; Zhou et al., 2014), while Chinese paddlefish was recently declared 'Extinct' (Zhang et al., 2020).

For one of the two remaining sturgeon species present in the area, the 'Critically Endangered' Chinese sturgeon, a 'new' spawning ground was officially recorded in 1981, located just below the Gezhouba Dam (Wei, 2020). Despite this, the number of mature animals that reach this spawning site constantly decreases from about 2000 individuals in the early 1980s to less than 100 individuals per year in the 2010s (Wu et al., 2015).

However, from April to August 2015, more than 3000 juveniles were recorded in the Yangtze estuary, representing the most relevant spawning signal in 10 years, which is also the last recorded so far.

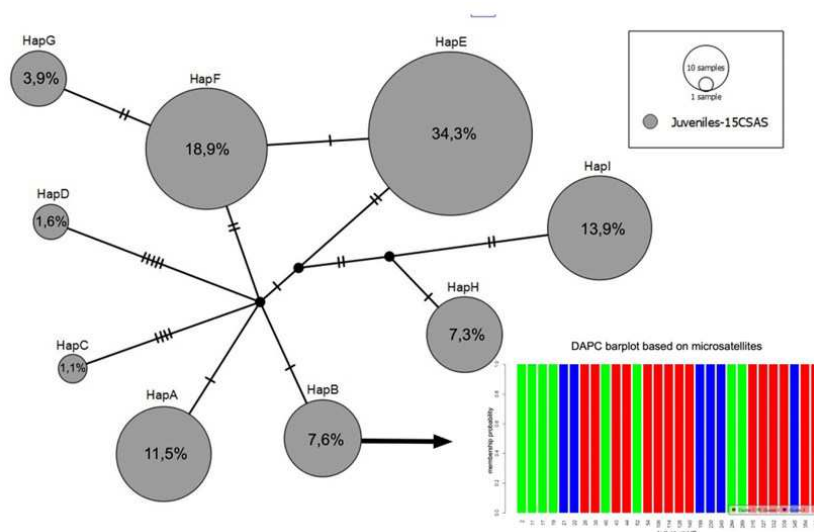


Figure 3 Haplotype network showing relationships among the nine mitochondrial haplotypes detected in the population sample and their relative frequencies expressed in percentages.

The inset bar-plot represents a result example of the DAPC analysis based on microsatellite data and performed to detect differentiated KHn groups within each haplogroup.

The example is referred to Haplotype B for which we considered three differentiated cluster of individuals KB1, KB2, and KB3.

This allowed providing data on the residual effective number of breeders and their genetic make-up, a crucial investigation to inform future *in situ* and *ex situ* conservation strategies.

Genetic characterisation was carried out using the mitochondrial control region and nine nuclear microsatellite markers on a total of 382 of 400 randomly selected individuals for which fin clips were collected. By analysing mitochondrial sequences, 21 variable sites were identified, resulting in only nine haplotypes (A-I) that were unevenly distributed within the population samples (Figure 3).

The data set was then analysed using the discriminant analysis of Principal Components (DAPC; Jombart et al., 2010) implemented in the R package ‘Adegenet’ (Jombart, 2008) within each mitochondrial haplogroup (AI) to capture further genetic differentiation signals and to estimate the best number of genetically differentiated clusters (K).

To run DAPC, microsatellite data were transformed into dominant data, as the exact number of copies of each allele at each *locus* cannot be accurately resolved in tetraploid sturgeons (Boscari et al., 2014a; Boscari et al., 2014b). The DAPC analysis allowed the identification of a total of 31 genetically differentiated groups of animals (KHn) with a sample size ranging from 2 to 34 individuals. Each K group is named with a capital letter identifying the haplogroup (A-I) and a number that refers to a differentiated subgroup within such a haplogroup (e.g., KA1, KA2, KA3, KB1, ..., KHn). Subsequently, the degree of kinship was estimated by comparing the genetic distances between the identified groups (KHn) with the distributions of expected genetic distances between simulated groups with a known degree of relatedness: full-sib (FS), half-sib (HS) and unrelated (UN).

For the simulations, only the five *loci* that did not show any sign of possible duplication (more than four alleles per individual genotype) and for which the tetrasomic segregation pattern of inheritance (Boscari et al., 2011) had previously been verified were used. Simulated progeny was obtained through the previously mentioned ‘Breeding Sturgeon’ software (Boscari et al. 2014b).

Briefly, the software infers all possible tetraploid genotypes for each ‘parental’ individual at each *locus* and then generates random mating pairs by simulating gametes and randomly drawing alleles according to the tetrasomic mode of inheritance. From a panel of 66 parental female and 66 parental male profiles, a total of 198 FS groups of FS with 112 virtual individuals each were generated. These virtual offspring profiles were then compared by obtaining pairwise genetic distances used to assess the reference distributions for FS, HS, and UN.

Subsequently, the simulated distributions were compared with the average observed band sharing (BS) within and between KHn groups to estimate the kinship from the observed data. Statistical significance was obtained by comparing an equal number of randomly selected comparisons within the observed KHn group and the three reference distributions. Therefore, we counted the number of times the expected averages for the three simulated distributions exceeded the observed values of 200 replicates and expressed them as a percentage. More details are given in Annex C.

As an example, comparisons of groups with at least eight individuals are reported in Figure 4. For each comparison, the probability of belonging to each of the three kinship categories was estimated.

Most genetically differentiated groups were composed of FS, suggesting that the number of haplotypes approximates well the real number of breeding females. The different groups with the same haplotype were categorised as HS, pointing out the female as the shared parent. A minor fraction of within-group comparisons was preferentially assigned to HS rather than FS, which could be a signal of further genetic subdivision within those groups, although it is more likely that this is an effect of the small sample size available for these KHn groups.

If the above counting of nine females can be considered to be strongly reliable, estimating the number of males is less straightforward. In many cases, comparisons between groups with different haplotypes (different maternal contributions) were assigned to HS, suggesting that the father could be shared.

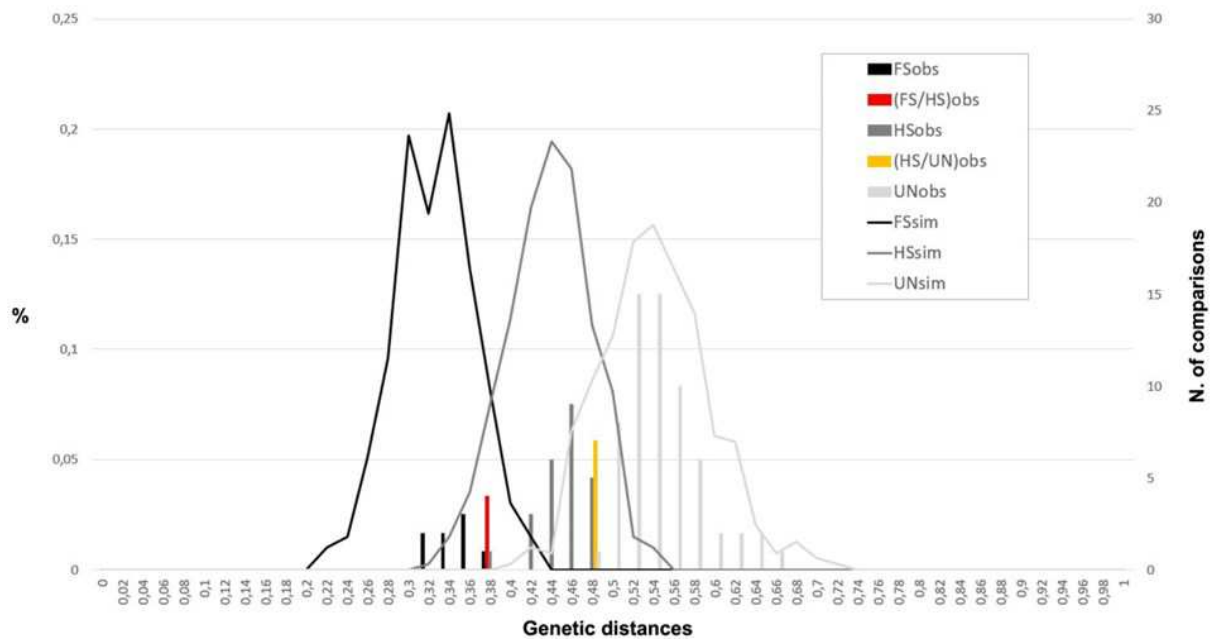


Figure 4 The three curves represented by continuous lines show the pairwise genetic distributions of simulated groups of animals with known degree of relatedness: Full Sibs (FS), Half sibs (HS), and Unrelated (UN). The superimposed bar-plot shows the distribution of the mean observed pairwise distances between KHn groups considering groups with a number of individuals greater than 8 (105 comparisons). The different colours represent the final kinship assignment: FS = black, HS = dark grey, UN = light grey. Comparisons whose probability of assignments have a difference of less than 10 % for two different kinship categories remains unresolved and are represented in red (FS/HS) and orange (HS/UN).

Considering that estimating the exact number of males remains a speculation, the significant fraction of groups that share the male can be taken as a good indication that the overall number of male breeders might be considerably lower than the number of KHn groups. This allows us to hypothesise almost the same order of magnitude for the number of female and male breeders, indicating a total of approximately 20 animals participating in the spawning.

However, since on average each Chinese sturgeon female spawns every 2 to 7 years (Yue et al., 2015), it can be reasonably hypothesised that additional animals exist that did not reproduce in 2014. Furthermore, we cannot exclude the presence of additional animals participating in spawning without reproductive success or whose progeny was not sampled.

To date, even focussing strictly on the spawning stock here estimated, the presence of nine different haplotypes in such a small population suggests a discrete residual genetic diversity, of the same order of magnitude as what is observed in other sturgeon species such as the Adriatic sturgeon (Boscari et al., 2014a, Boscari et al., 2014b). Given the rapid demographic collapse of this species, this precious heritage should be the target of conservation measures that must be implemented urgently. The only long-term way to protect Chinese sturgeon is through the construction of a bypass channel that connects the medium-high stretch of the river with the ocean (Wei, 2020). Meanwhile, it is extremely urgent to implement short-term conservation strategies that include direct actions at both the ecosystem and species levels. First, locate and protect the spawning area predicted to be between 20 km below the original spawning ground and the mouth of the Qingjiang River, as well as restoring suitable environments for reproduction according to what has already been done for other sturgeon species (Českleba et

al., 1985). Second, establish *ex situ* conservation programmes through informed management of the more than 3000 F1 kept in aquaculture (Wei, 2020). This includes the evaluation of purity and genetic diversity management according to strategies already developed for other polyploid sturgeons (Congiu et al., 2011; Boscari et al., 2014a; Boscari et al., 2014b, Boscari et al., 2017a).

3.4 INVESTIGATION OF MULTI-LOCUS INTRON POLYMORPHISMS (MIPS) POTENTIALITY IN TELEOST SPECIES Original publication in Annex D

The new Next Generation Sequencing (NGS) techniques and the advance in protocols for library preparation allow the Genotyping-by-sequencing (GBS) of multiple *loci* in several individuals. Thanks to the ever-increasing genomic resources, it is also possible to bioinformatically identify specific regions of interest and design panels of primers for their amplifications across species. This procedure is facilitated when the sites of interest are flanked by conserved regions. This is the case for introns that are bordered by coding exons (Ortega and Maldonado, 2020). Based on this, the investigations of their potentiality and transferability have been evaluated in the present study starting from the Teleost taxa, for which several genomes distributed across their phylogeny are available.

A bioinformatic pipeline has been created to extract the exonic regions, 100 bp before and after the intron junctions, from data available in 48 fish genomes. Clusters of these sequences with at least 90% identity and containing information of at least 15 genomes have been assembled. The 2441 clusters were examined for highly transferable primer design. Clusters were selected to have an average intron length between 200 and 800 bp, the highest possible number of species, and highly conserved exonic regions. These criteria allowed the selection of 121 *loci* with predicted lengths compatible with Illumina sequencing, easily amplifiable with PCR, and potentially transferable between species for which the genome is not yet available. Subsequently, the primer pairs were used to characterise 384 individuals belonging to 65 different species of Teleost. Libraries were prepared by multiplex PCR and PCR-indexing, and then sequenced in 300 pb paired-end Illumina MiSeq v3. Fastq reads have been demultiplexed and filtered by quality, and the paired-end reads merged intra-sample for each *locus*. At this point, *loci* showing coverage of fewer than 30 high-quality reads or too long to be merged were removed at the species level. Finally, the allelic variants at each *locus* were identified based on an intra- and inter-individual analysis and, for each *locus* and sample combination, the first alleles in frequency whose cumulative abundance was superior to 75% of the total count were retained. In addition, the presence of no more than two alleles per genotype was checked. Individual replicated showed a percentage of *loci* with identical genotypes of 88.1% on average, and replicated within or between runs are comparable with a mean of genotype identity of 89,65% $\pm$ 6,05 intra-run and 84,18% $\pm$ 7,42 inter-run, showing good replicability of the methodology. A more detailed description of the experimental procedures, sequencing performances, and bioinformatic filtering is provided in Annex D.

Regarding the estimation of the transferability of *loci* across species, it must be considered that 42 of the 65 species analysed in this work are represented by only one or two individuals of different origins and states of conservation. This possibly implies an underestimation of the degree of transferability, since *loci* could fail the amplification due to poor conservation, while they may give better results if amplified on more individuals or better-quality DNAs. Despite this, MIPS markers have proven to be highly transferable, as more than 60% of *loci* are amplified in more than half of the species.

The best results were obtained in species belonging to the Eupercaria, Ovalentaria, and Carangiarina clades (Betancur et al., 2016; Nelson et al., 2016), with a mean of 76 *loci* (62.8%), 72 (59.6%) and 71 (58.7%), respectively.

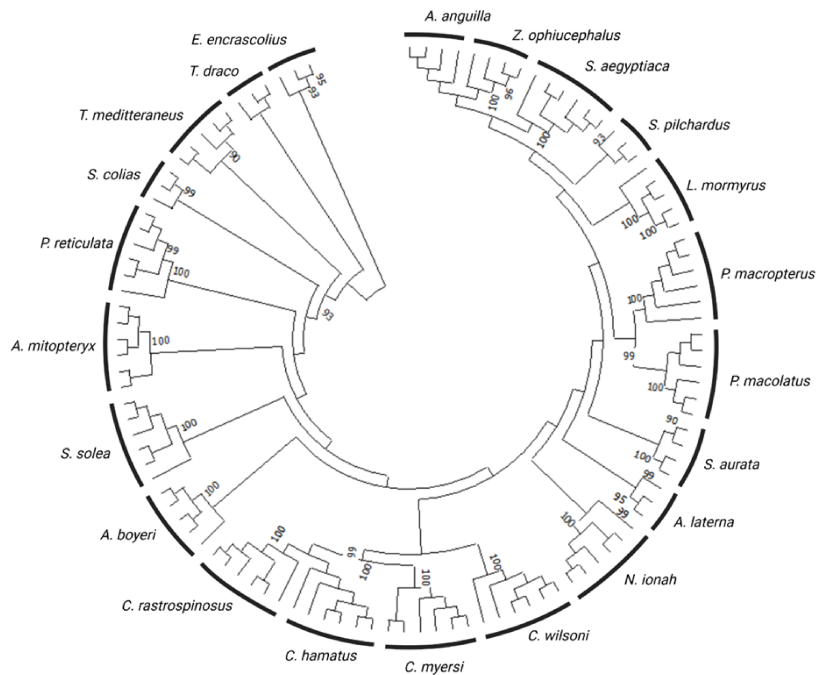


Figure 5 Graphical representation by UPGMA of the allele sharing across a subsample of genotyped individuals. Only species with at least three individuals have been considered and, for highly represented species, a maximum of six individual with the highest coverage have been included.

On the order of Clupeiformes, Gadiformes, Cypriniformes, and Anguilliformes, the last two having only one species tested, the genotypes characterisation was successful at 51 *loci* (42.4%) on average. Overall, given the wide taxonomic range of Teleost species tested, this can be considered a very satisfying result. These *loci* also revealed a strong potential for species identification. The pairwise analysis of allele sharing between samples is represented graphically in Figure 5. In this analysis, the species with at least three individuals showed high coverage at almost all *loci* to avoid missing data. The fact that samples of the same

species clustered together suggests a strong identification power of MIPs markers. To validate this hypothesis, MIPs have been applied to two case studies, the first aimed at testing the species identification power in the suborder Notothenioids, and the second to analyse the genetic differentiation within the genus *Solea*, at both species and population level. *Loci* were selected separately for both sole and icefish case studies considering only the species under analysis. Each *locus* was selected if the genotype was characterised in at least 75% of individuals per species or population, avoiding homopolymer elongation or microsatellites that fall in the merging region of the paired-end reads.

Focussing first on Notothenioids, some of these species, besides having a close phylogenetic position, show subtle morphological differences that make them prone to misidentification based on morphology. Furthermore, for species of the genus *Chionodraco*, interspecific hybridisation was hypothesised based on microsatellite analysis (Marino et al. 2013).

A subsample of 79 individuals belonging to seven diverse species was genotyped, some of which have already been analysed (Schiavon et al., 2021). Of the 121 intron *loci* tested, 54 (44%) passed selection for the analysis on these Notothenioids species. The differentiation between samples within and between population samples based on the 54 intron *loci* is represented by the Principal Component Analysis (PCA) (Figure 6). The first, second, and third principal components (PC1, PC2, and PC3) explain 16.14%, 10.19%, and 8.75% of the variance in the dataset, respectively, and almost all samples cluster according to the declared species. The only exception was represented by a single individual (Cham_PS82_2711) classified as *Chionodraco hamatus*, appearing completely differentiated from all other samples included in the dataset. Sequencing of the mitochondrial gene, Cytochrome Oxidase I (COI), revealed a misclassification and assigned this individual to the *Trematomus scotti* species, belonging to the

suborder Notothenioidei. The eleven animals that exhibited ambiguous phenotypic traits that prevented a clear morphological classification were pure assigned to different species of the genus *Chionodraco*. Even here, the results were supported by mitochondrial barcoding.

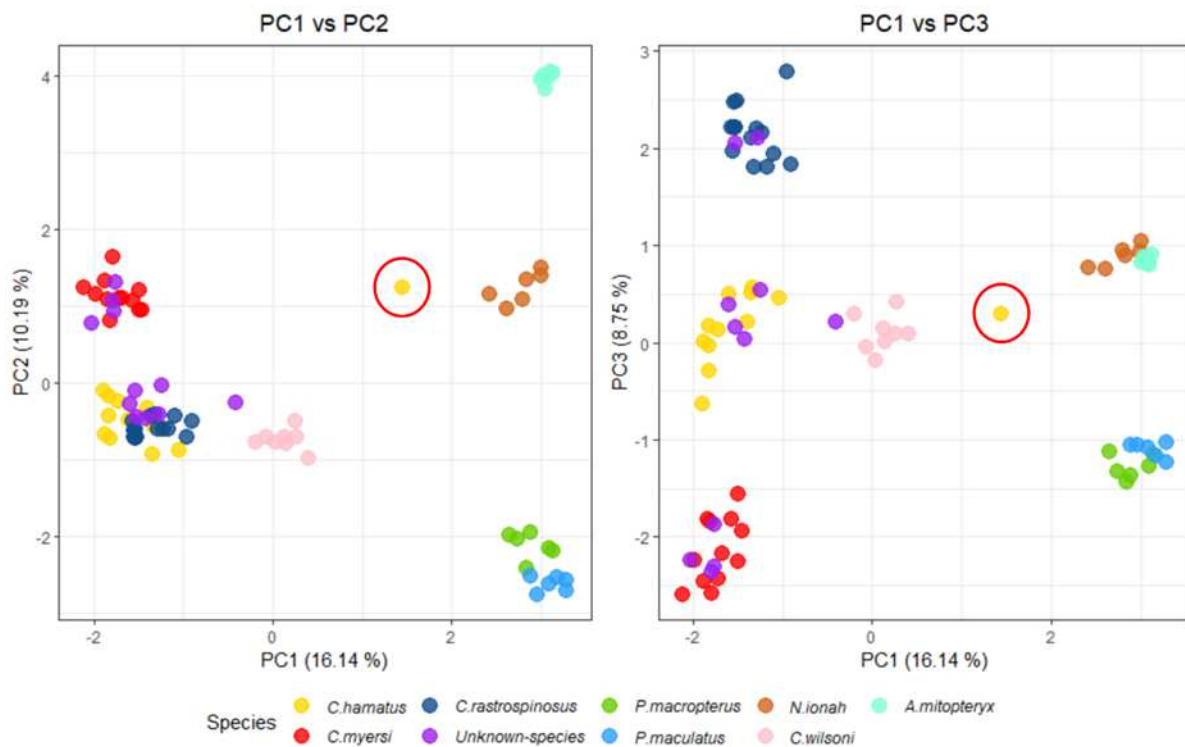


Figure 6 Plot of the principal component analysis of variation based on 54 intron *loci* genotyped at a total of 81 icefish individuals. The three first axes explain 16.14, 10.19 and 8.75% of the variation. Colours indicate species. The outlier individual which has been genetically assigned to the species *Trematomus scotti* (originally misassigned to *C. hamatus*) is highlighted by a red circle. A PC1 vs PC2. B PC1 vs PC3.

In the soles case study, 106 individuals were considered, 71 *Solea solea* and 35 *Solea aegyptiaca* sampled from 6 and 5 different sites, respectively.

A total of 43 MIPs markers passed all filters and were selected for the analyses. Almost all *loci* were polymorphic, 100% and 93.1%, respectively, in *S. solea* and *S. aegyptiaca*, and 39 of them showed alternative alleles for the two species, and each of these could be used to develop rapid single *locus* tools to differentiate the two species.

In this case, without considering a priori knowledge based on morphology, the first, second and the third principal components (PC1, PC2, and PC3) explain 42.5, 4.45, and 3.77% of the variance in the dataset, respectively. The variation of PC1 explains species divergence while PC2 and the PC3 explain intraspecific subdivisions in *S. aegyptiaca* and *S. solea* (Figure 7B). In *S. aegyptiaca*, all analyses identify the best cluster as concordant with 6 populations subdivision (Figure 7B to the left). In species of *S. solea*, the MIPs markers substantially replicate what has already been assessed in previous works (Corti, 2022; Corti et al., 2022): a continuous gradient of divergence from the west to the east side of the Mediterranean basin (Figure 7B on the right).

Even here, MIPs markers identify incorrect morphological classifications. Three individuals captured in the North Adriatic, sampled at the ‘Chioggia Fish Market’ and morphologically classified as *S. solea*, instead clustered with the Adriatic population of *S. aegyptiaca* in all

analyses (Figure 7A). The misclassification has been confirmed again by the mitochondrial barcode (COI) that assigned all samples to the putative species *S. aegyptiaca*.

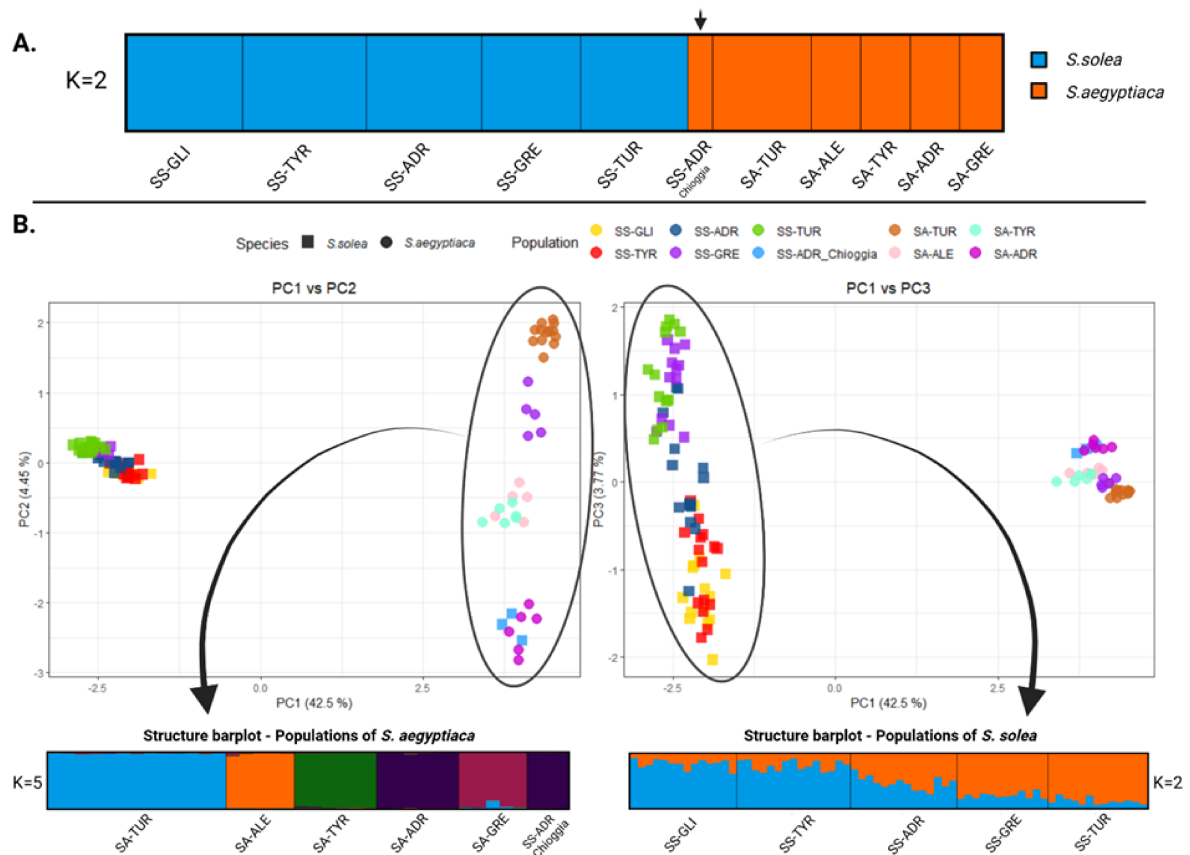


Figure 7 Distribution of genetic differentiation patterns between species and populations of the genus *Solea* based on the genotype information at 43 intron *loci*.

A. Structure bar-plot showing the membership probabilities of cluster assignment for each of the 106 individuals of the two species (71 *S. solea* and 35 *S. aegyptiaca*). The best number of cluster ($K=2$) that overlaps the species subdivision is reported. The black arrow indicates the three individuals morphologically classified as *S. solea*, which cluster with *S. aegyptiaca* according to their mitochondrial barcode information. B. Plot of the Principal Component Analysis of variation (PC1 vs PC2 on the left, PC1 vs PC3 on the right). Loading on the three first axes explain 42.5%, 4.45% and 3.77% of the variation. Shapes indicate species while colours indicate populations. Below each PCoA, the best cluster subdivision obtained by structure analysis performed within species is also reported as bar-plot showing individual membership probabilities.

Overall, the potentiality of MIPs markers has been clearly demonstrated by these two case studies in which both sole and icefish were easily distinguishable and by the correct species clustering of individuals belonging to 22 different taxa (Figure 5). The cases of discrepancy between morphological classification and genetic analysis turned out to be due to errors based on an unclear morphological phenotype, and the genetic identification of the MIPs was always consistent with the mitochondrial barcoding.

MIPs markers are characterised by highly polymorphic multi-locus codominant information with an expected lower level of homoplasy and ambiguity compared to microsatellites.

In fact, the informative power of a single MIPs *locus* can be very elevated, since a single intronic region can harbour an elevated number of haplotypes, each of which can be considered as an allele containing a distinctive set of SNPs or indels. As an example, the *Solea* samples showed *loci* with up to 40 alleles. This represents an undoubted advantage of MIPs, which, like

microsatellites, are promising to be a powerful analysis tool both at individual and population levels, due to their mendelian inheritance, applicable even in paternity or relatedness analysis. As regards the comparison with other Genotyping-by-sequencing methods, the approaches are quite similar for the experimental effort involving the preparation of genomic libraries and subsequent sequencing. Nevertheless, differently from other targeted GBS techniques, once the *loci* have been bioinformatically identified and the primer designed in the conserved exons, these transferable primers can be applied to a wide range of species for which no previous information is required. This property is shared by other methods, such as Restriction site associated DNA (RAD) sequencing, which also do not require a priori information to perform the genotyping. However, the already mentioned high conservation of the flanking exonic regions where the primers attach also ensures a higher efficiency, greatly reducing the percentage of amplified *loci* that are discarded because they are not shared among individuals in the downstream analysis. As a result, sequencing data waste is also reduced, with a consequent benefit in costs and the potential to increase the number of samples that can be characterised in a sequencing run. Moreover, it is always possible to increase the average coverage per *locus* performing new sequencing runs, an important option that we have applied, to efficiently genotype polyploid sturgeons that are also characterised by a high number of duplicated genes (Du et al., 2020) as illustrated in the following session. Therefore, the MIPs approach combine the advantage to focus on a reduced number of highly variable *loci* with the possibility to be applied to a wide range of non-model species without a priori information. Moreover, our methodology provides MIPs sequences length up to 600bp, much greater than that of the RAD methods. These long sequences that capture genetic variations across multiple adjacent SNPs allowing to identify a large number of alleles per *locus*, also permit the development of quick, inexpensive, and easily applicable diagnostic PCR-based methodology for forensic analyses and species identification. The development of these simple diagnostic techniques may seem surpassed by highly efficient methods, where it is possible to simultaneously genotype hundreds of SNPs in hundreds of individuals, such as SNP-Chip methodologies. However, the convenience of these approaches is lost when a small number of individuals need to be analysed, such as in routine support to *ex situ* conservation programmes or in forensic analyses. In these cases, having a quick and effective approach based on a simple, fast, and cheap PCR can be very beneficial. Both these aspects are of primary relevance in Acipenseriformes that, as already mentioned, are greatly impacted by poaching and illegal trade and for which the identification of pure-species individuals is fundamental for their *ex situ* management. A preliminary investigation on this group of species is described in the following paragraph.

3.5 PRELIMINARY INVESTIGATION OF MULTI-LOCUS INTRON POLYMORPHISM (MIPS) IN STURGEON SPECIES

Original publication in Annex E

In sturgeons, conservation, the development of tools to identify species and hybrid is fundamental to contrast illegal trade of derived products and, in conservation, to ensure the correct identification of individuals used to establish brood stocks for *ex situ* conservation. Sturgeons, in fact, are characterised by the possibility of producing a fertile hybrid when species with the same ploidy level are crossed (Havelka et al., 2011); moreover, most of them are difficult to identify, especially in the case of similarity of the parental species, such as in the so-called “gueldenstaedtii-complex” composed by *Acipenser gueldenstaedtii*, *Acipenser persicus*, *Acipenser baerii*, and *Acipenser naccarii* (Boscari et al. 2014a). An identification protocol has already been established exploiting the variability present in the intron present in the intronic regions of the nuclear S7 ribosomal protein (RPS7) and the nuclear S6 ribosomal protein (RP2S6) (Boscari et al., 2014a; Boscari et al., 2017b). However, identification becomes more complicated when interspecific hybrids backcross with the parental species. Consequently, the development of new markers is still necessary, especially for second- or third-generation hybrids, where multi-locus nuclear markers will allow to increase the chance of detecting all the species involved in a hybrid.

To this purpose, the same approach previously described and use for Teleost species (Annex D) has been applied to Acipenseriformes with some adjustments. In this case, the information on the only available Acipenseriformes genome, belonging to the Sterlet sturgeon, *Acipenser ruthenus*, was compared with the available transcriptomes of *Acipenser dabryanus*, *Acipenser fulvescens*, *Acipenser naccarii*, and *Acipenser stellatus*.

For the genome of *Acipenser ruthenus*, the GTF files, which contain information on the structure of annotated genes and contain their coordinates, were downloaded from the Ensemble database (<https://www.ensembl.org/>). According to the GTF files, the fasta exon-flanking sequences of each intron were extracted, and 100 bp before and after the intron junction were maintained for further analysis. Then, the transcriptomic and genomic sequences were clustered with at least 90% identity. The 72014 sequence clusters obtained were examined for highly transferable primer construction. Clusters were selected to have an intron length between 400 and 500 bp in the sterlet genome and highly conserved exonic consensus sequences. These criteria allowed to concentrate the investigation efforts on a first set of 96 *loci*. Then the primer pairs have been used for the preparation of the PCR-based library in 192 individuals of 17 different species, with the same protocols used for Teleost species (Annex D).

After the first quality filters, the SeekDeep v 3.0.1 (Hathaway et al., 2018) pipeline was implemented to generate the clustering of allelic variants at each *locus*. Unlike the settings used for Teleost, in this case, the pipeline parameters have been empirically modified to obtain the highest level of concordance in the replicates for these species characterised by different levels of ploidy. In the final set of parameters called ‘ERCACutOff40’ all the reads that differ for a single low-quality base have been ignored and collapsed into a single haplotype cluster. Then the low-coverage haplotypes differing for a single base from higher-coverage (at least five-fold) ones collapsed with the latter. These settings are intended to reduce the possibility of artefacts produced by the amplification and sequencing steps. Additionally, low-frequency haplotypes were retained in one individual if the same haplotype was well supported in other samples.

Finally, all clusters composed of less than 40 sequences or recognised as chimeric were removed. The parameters used in the SeekDeep ProcessClusters package are shown in Box 1. The replicates of the five samples belonging to five different species, which are *Acipenser brevirostrum*, *Acipenser fulvescens*, *Acipenser naccarii*, *Acipenser sinensis* and *Huso huso*, reached levels of allele concordance of 66%, 49%, 64%, 54% and 73%, respectively.

Box 1: 'ERCAcutOff40'

List of parameters empirically modified in the ProcessClusters package of SeekDeep to increase the allelic concordance between the library replicates during the bioinformatic calling of alleles at individual level.

--strictErrors

Collapse parameters with a several low-quality mismatches in population clustering

--sampleMinTotalReadCutOff=30

Sample minimum total reads cut off, if the total read count for the sample is below this threshold it will be thrown out

--excludeCommonlyLowFreqHaplotypes

Remove commonly low frequency haplotypes

--majorHaplotypeFracForRescue=0.05

Fraction in order to be considered a major haplotype in a sample for comparing during rescue

--allowHomopolymerCollapse

Allow homopolymer slippage collapse between clusters

--excludeLowFreqOneOffs

Collapse any haplotypes that are low frequency compared to another haplotype (determined by --oneOffLowFreqMultiplier) and only differs by 1 base

--oneOffLowFreqMultiplier=5

Low frequency multiplier used for --excludeLowFreqOneOffs, considered low frequency if haplotype frac is less than its fraction times this number than the other haplotype

--rescueExcludedLowFreqHaplotypes

Rescue excluded haplotypes for falling below frequency cut off if they appear as a major haplotype in another sample

--clusterCutOff=40

Clusters of less than 40 sequences will be removed

Notwithstanding the above selection criteria, the low percentage of maximum identity obtained between replicates and a preliminary analysis of the library highlighted several problems. In all species, for example, the presence of extra-numerary alleles was found in diverse *loci* compared to what was expected based on their ploidy. This seems to be due to an unexpectedly high frequency of duplicate *loci*. Duplications are inferred by the presence of two or more

distinct allelic clusters identifiable in an individual single *locus* alignment. This hypothesis is supported by the fact that once the two (or more) distinct sets of alleles have been identified, the number of alleles per each sub-locus were coherent to the expected functional ploidy of the species. These observations are also consistent with the higher level of duplicated genes observed in the *de novo* reference genomes of *Acipenser ruthenus* (Du et al., 2020).

Unfortunately, even if the preliminary investigation is promising, it has not been possible to perform a detailed analysis of the characterised *loci*, nor have it been possible to identify particularly informative single SNPs for all the species analysed. These impediments were caused by various problems encountered during the processing of the raw data, which made the present results very preliminary. The high frequency of duplicate *loci*, combined with the polyploidy that characterises many species here analysed, resulted in a drastic drop in the average coverage observed for the different alleles, leading to the impossibility of always discriminating the real haplotypes from the artefacts with certainty, making the data at diverse *loci* unusable. Therefore, it is evident that the protocol that worked well for Teleost species, with the same type of sequencing platform, needs to be further improved to cope with the genetic peculiarities of sturgeon species.

Further sequencing runs could greatly mitigate the coverage dilution caused by gene duplications. This would ensure more reliable post-processing of sequencing data, leading to an increase in the concordance between replicates. It would therefore be possible to reconsider the discarded *loci* and then proceed to a detailed screening of the data.

However, the best characterised and concordant *loci* were used to perform allele presence-absence profiles, which allowed to clearly group the different individuals according to their species (Figure 8). Profiles were visualised in a Neighbor-Joining's tree obtained through the MEGA-X software (Kumar et al., 2018). More precisely, the 18 *loci* that were present in at least 90% of the individuals and at least 10 species were selected. The three species for which only one individual was present in the library were excluded from this analysis.

The allelic information present at these *loci* was sufficient to correctly group individuals of 12 species out of the total of 14 analysed, with the only exceptions of two species, *Acipenser schrenckii* and *Huso dauricus*, which clustered in a single group. This result was partially expected, as the same samples analysed with different markers had shown a certain degree of admixture (Boscari et al., 2017a). In fact, the distinction between these two species is complex, probably due to past hybridizations that occurred both in captivity and in the wild. Hybrids of these two species are fertile and can be crossed with each other or backcrossed with the parent species. Additionally, the caviar produced from the hybrid between these two species is highly valued and represents the majority of Chinese aquaculture caviar production. Until a few years ago, when breeding in floating cages was still allowed in China, most of these hybrids were reared directly in the river inside floating structures, and on several occasions, due to extreme weather phenomena, tens of thousands of individuals (including hybrids) escaped into the wild. For this reason, probably, both in captivity and in the wild, animals with different degrees of hybridisation are present, like our reference samples that were shown to be not pure species individuals.

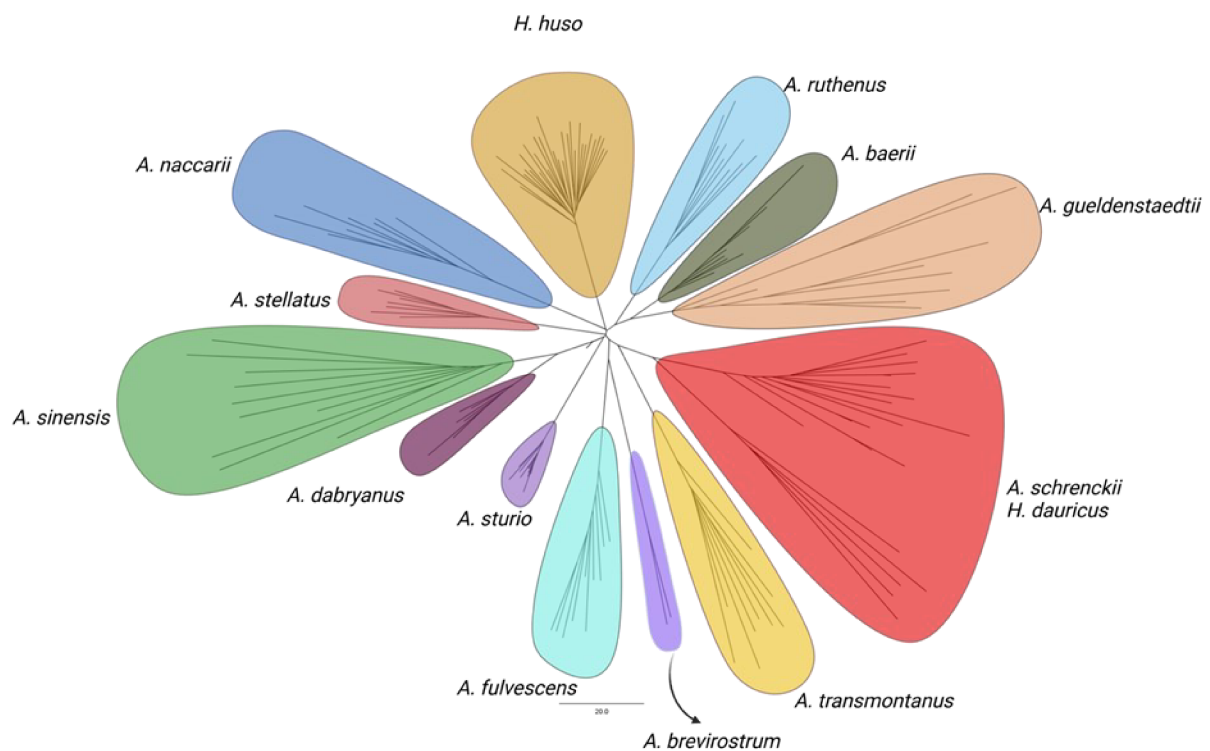


Figure 8 Neighbor-Joining tree based on the number of allelic differences among species for the 18 best characterize intronic *loci*. The genotypes at each *locus* were converted into presence/absence data for each allele. The data were then transformed into a sequence consisting of the T and A bases, to be an input file for the MEGAX software.

The MIPs markers characterised in this work have confirmed, although through preliminary analyses, the expected potential for this type of hypervariable nuclear *loci* in the identification of species and interspecific hybrids. From the MIPs sequences, after further sequencing runs, it will be possible to identify diagnostic SNPs, which will be useful for the development of rapid PCR-based markers for the identification of species and hybrids.

3.6 PRELIMINARY RESULTS OF THE *IN VIVO* STUDY TO EVALUATE THE GENETIC LOAD PREDICTIONS IN THE ADRIATIC STURGEON

The endemic species *Acipenser naccarii* was included in the PRIN project ENDEMIXIT (<https://endemixit.com/>), with the aim of applying the genomic approach to the study of five endangered Italian endemics. One of the main goals is to understand the dynamics of the accumulation of deleterious mutations in declining populations and their impact on fitness and risk of extinction.

From genomic data, it is possible to estimate the genetic load based on bioinformatically predicted deleterious variants in coding genes. However, functional experiments necessary to obtain mechanistic insight and to test the predictions obtained through genomic approaches are still rare in non-model organisms. In this context, the previously characterised stocks kept in captivity at the aquaculture plant ‘Storione Ticino’ allowed an *in vivo* study to assess the relationship between fitness indicators observed in controlled crosses and the bioinformatically estimated genetic load. Six pairs of first-generation (F1) full sibs have been selected to perform 12 different crosses, obtaining six distinct inbred and six outbred families. This cross-breeding plan was designed to have groups of offspring with various expected levels of homozygosity and to verify differences in fitness between inbred and outbred individuals. Subsequently, a first batch of about 500 eggs from each cross was sampled and, after hatching, the larvae have been kept in a controlled environment for the first two months. At this stage, all dead larvae were sampled and stored in ethanol for further analysis. Unfortunately, saprolegnia infections, which often affect eggs even in aquaculture plants, greatly reduced the success of hatching in some batches and prevented the possibility of extracting good quality DNA from eggs. In addition to this, mortality in the first days was very high, reducing some of the family groups to a few individuals, insufficient to continue with the following assessments.

Being aware of this risk, the remaining larvae produced by the 12 reproductions were kept in separate ponds at the ‘Storione Ticino’ plant, allowing the collection of 100 randomly sampled animals from each family and their translocation to a specific facility in the ‘Oasi le Sorgenti’ (Dueville, Vicenza, Italy). Some behavioural responses were evaluated as fitness indicators, such as light responses, resistance in a flow channel, and the ability to explore an environment to find food sources. Although it was not possible to measure the propensity to explore a new environment, some signals were found for the first two tests. However, these indicators were not suitable to be included as fitness proxy. The responses were inconsistent even at the individual level, while the results between individuals did not show great variability. Therefore, it would have been necessary to perform many repetitions to eventually find a pattern, with the risk of stressing the animals. Clearer were the differences in growth rates. Although born the same day, the individuals started to show a significant difference in body size, and even within the same group, an individual could be many times larger than a sibling. From January 2021, the length and weight of all the individuals were measured. Measurements were repeated every six weeks until July for a total of five assessments. Analyses revealed a non-significant difference in size and growth rate between inbred and outbred individuals, while a significant difference emerges over time between the different families (Figure 9). An explanation might lie in the fact that after only one generation of inbreeding, the effect of inbreeding depression due to the complete unmasking of recessive deleterious mutations could be mitigated by polyploidy (Clo and Kolar, 2022).

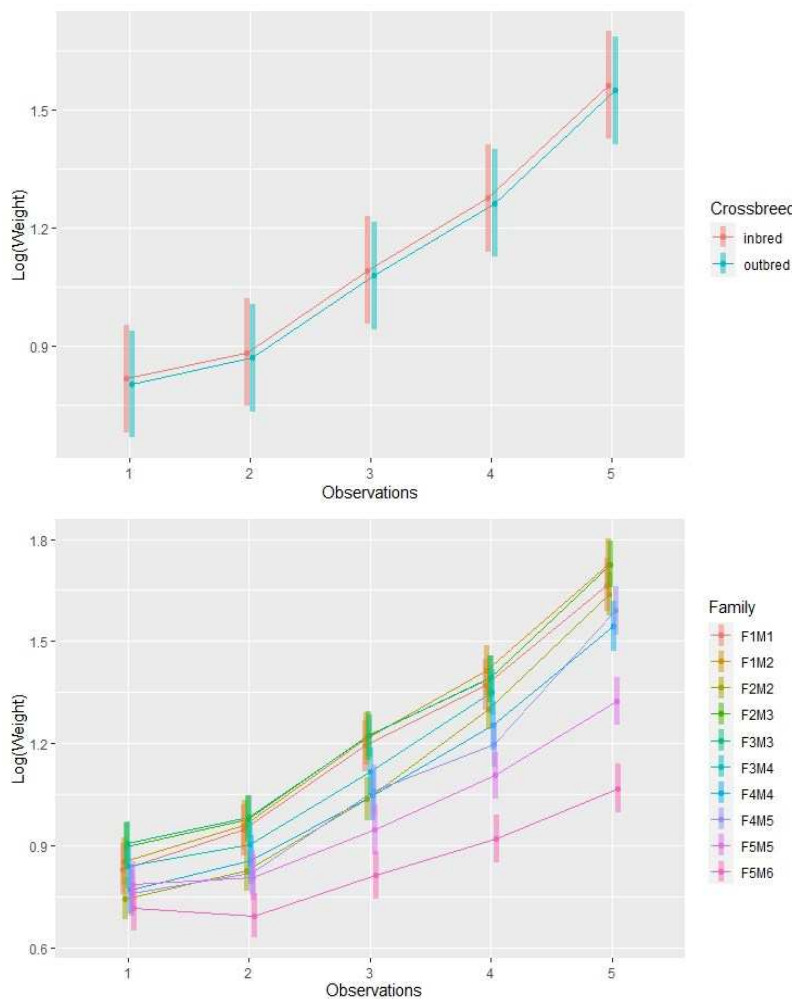


Figure 9 Graphical representation of the difference in growth-rate between inbred and outbred individuals (above) and among the different families (below). From January 2021 to July 2021 all individuals were weighed, every six weeks, for a total of five observations. The weights measured in grams were log-transformed to assume a normal distribution. A linear mixed model was used to compare the inbred and outbred average weights over time with family as a random effect. Model's formula:

$$weight = crossbreed + time + (family)$$

No significant differences were observed between inbred and outbred along the five observations.

A liner model was used instead to compare the growth among the different families. Data are missing for family F6M6 and F6M1 since were lost before measurements. Model's formula:

$$weight = time * family$$

A significative interaction was present (time*family: $p < 0,001$) indicating that the families have different growth-rate over time. Family F5M6 significantly differ in average weight ($p < 0,001$) respect to the other families in the 4th observation. Both family F5M5 and family F5M6 significantly differ in average weight ($p < 0,001$) respect to the other families in the 5th observation.

Furthermore, the impact of deleterious mutations could be buffered by the presence of duplicated genes in polyploids (Fox et al., 2020), which appears to be abundant in sturgeon species (Du et al., 2020). However, the differences among families could reflect the effect of a different combination of genes, and thus a different load of deleterious mutations. Although it is not possible to exclude some parental effects, for example, the different quality of the eggs in female 5 (F5), which could have impacted the early stage of development and the successive performances of the offspring (Green, 2008).

The genomes of the twelve parents were sequenced in the Illumina NovaSeq 6000 system at the University of Florence, producing paired-end reads with an expected coverage of 30X per sample. The actual average sequencing coverage ranged from 21.8X in female breeder 2 (F2) to 62.6X in male breeder 6 (M6) with an average of 34.9X (SD = 9.6).

The ENDEMIXIT project also included the assembly of a *de novo* genome for *Acipenser naccarii*, but this task has not been completed in time due to delays in sequencing processing at the 'Vertebrate Genome Project' laboratory (New York, USA). Now, the assembly, currently in progress, requires the design of a custom pipeline to reconstruct the haploid scaffolds of this highly complex tetraploid genome.

For these reasons, the family breeder sequencings have been mapped on the only available sturgeon genome, of the species *Acipenser ruthenus* (Du et al. 2020). The mean depth ranged

from 19.1X in female breeder 2 (F2) to 54.0X in male breeder 6 (M6) with an average of 30.3X (SD=8.2). After mapping performed with Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009), alternative variants to the reference genome have been ‘called’ and ‘hard filtered’ following the best practises recommended using the bioinformatic Genome Analysis Toolkit (GATK) (McKenna et al., 2010). At this stage, SNPs with a GATK-inferred genotype quality greater than 10 (GQ>10) were selected. Then the effect of each mutation present in the breeders has been predicted using the bioinformatic tool SnpEff (Cingolani et al. 2012).

A total of 1000 biallelic SNPs were chosen: 400 SNPs predicted to have a high deleterious impact (i.e. stop-gain, stop-loss), 400 SNPs predicted to be missense, and 100 SNPs predicted to be synonymous as control. The chosen SNPs have been filtered to not map in repeated regions of the reference genome and not be in linkage, imposing a minimum distance of 50000 pb from each other during the selection. The frequency of mutations in the entire group of parents has been taken into account, avoiding variants with frequencies greater than 70% of the total possible alleles. This was done to limit the selection of putatively deleterious variants highly represented in the parent stock, without a real negative impact on fitness. Finally, variants having at least two mutated alleles in at least one parental individual were selected. In this way, mutations in complete homozygosity in the offspring could possibly be observed. Almost all mutations predicted to have a high deleterious impact were selected for genotyping, while the missense and synonymous mutations, which were in a higher number, were randomly taken after the previously described selections.

An even number of two differentially performing progeny groups was selected to study the respective levels of genetic loads for these three categories of predicted mutations. The first group consists of those larvae that died early before the yolk sac absorbent, while the second group contains those individuals who show a higher growth rate and body size at the end of the described observations. The offspring belonging to these two categories and evenly selected among families, were Genotyped-by-sequencing at ‘Floodlight Genomics LLC’ (Knoxville, TN, USA), for a total of 384 individuals. After sequencing, the reads were mapped again with BWA to the reference genomes of *Acipenser ruthenus*.

A total of 798 SNPs were successfully amplified, sequenced, and recovered. The average read depth per sample ranges from 8.1X to 224.5X (only one outlier showed this high value) with an average of 43.6X (DS=18.3). The variants were called and the genotypes were inferred with the GATK bioinformatic tool.

Identification of the correct genotype was not straightforward due to the tetraploid status of the species. In fact, five genotypes are possible for biallelic SNPs, namely *nulliplex* (AAAA), *simplex* (AAAB), *duplex* (AABB), *triplex* (ABBB), and *quadruplex* (BBBB). Although the inference of the *nulliplex* and *quadruplex* is relatively simpler, the inference of the three types of heterozygosity that biallelic SNPs can present is more uncertain. In particular, *duplex* genotypes suffer from most misidentifications being wrongly classified as *simplex* or *triplex* (Matias et al., 2019).

In general, most bioinformatic analysis tools are designed to work with diploid data and less support exist when working with polyploid datasets (Meirmans et al., 2018), making it difficult to adopt straightforward and highly validated protocols for polyploid data. In fact, an unequal distribution of genotype quality (GQ) and read depth (DP) was observed between the different genotypes inferred with GATK (Figure 10).

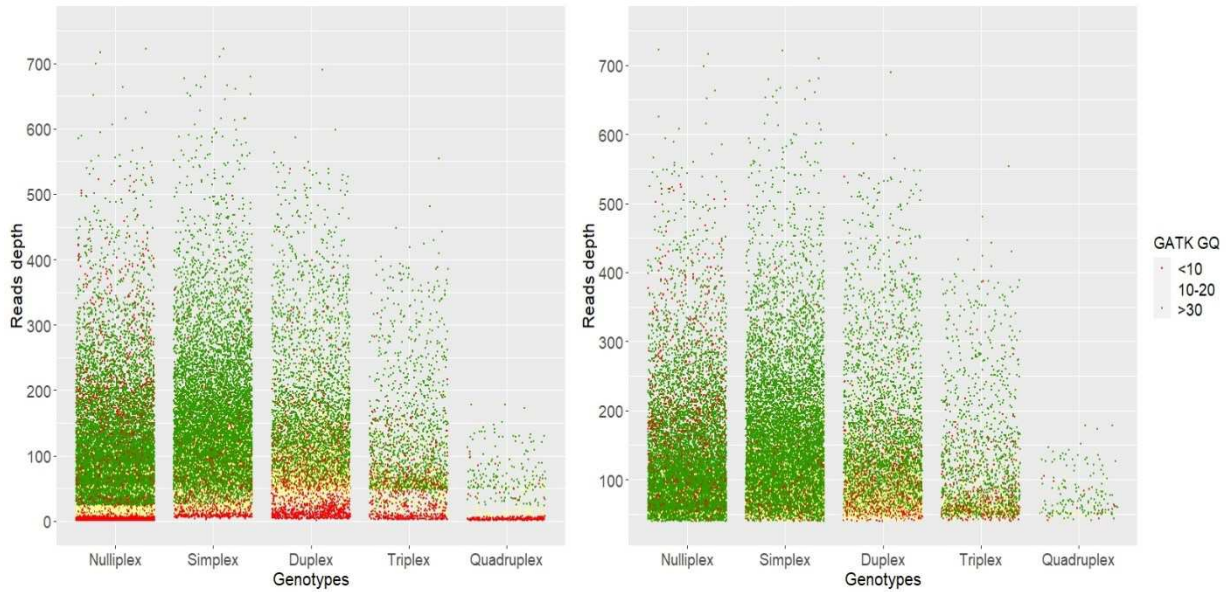


Figure 10 Graphical representation of the reads depth and genotype quality (GQ) values for each genotype inferred with GATK. A set of 95 samples (average read depth $43,6 \pm 5$) and all the 798 *loci* have been selected for the graphical representation (on the left). Another representation removing all the genotype with read depth lower than 40, the same threshold used on the preliminary analysis, has been performed (on the right). Each dot represents a genotype associated to a specific sample/*locus* combination. Red dots represent genotypes with a low estimate genotype quality (GQ <10), yellow dots represent genotypes with intermediate estimate genotype quality (GQ 10-20), and green dots represent genotypes with high estimate genotype quality (GQ >30). Reads depth being equals, it is possible to note as *duplex* and *triplex* genotypes are disfavoured in terms of genotype quality estimate by GATK. The pattern is clearer when all the genotypes with reads depth lower than 40 are removed (on the right). The GATK's genotype quality estimates that affect differentially the distinct class of genotypes need to be investigate and solved before to proceed with further analyses.

A very preliminary analysis was performed by filtering the individual SNPs for $GQ > 30$ and $DP > 40$ avoiding as much uncertainty about the genotype but without losing an excessive number of informative SNPs. The *loci* shared in at least 70% of all individuals were kept. A total of 361 SNPs were selected: 169 high impact, 150 missense, and 42 synonymous.

A significantly higher genetic load was observed in larvae for both high impact and missense SNP mutations, as expected. However, the same pattern was present for synonymous mutations that were selected as a control and for which no significant difference in mutation was expected. This could be due to an erroneous prediction of the effects of synonymous mutations identified on the annotated genes of a different species, or alternatively, in the incorrect inferences of genotypes as previously described. Confirming the unbalanced distribution, the GQ quality values are especially low for *duplex* and *triplex* genotypes, greatly influencing the estimations of genetic load for each of the three classes of variants (Figure 10). To proceed with the analyses, it is thus necessary to identify the causes that lead some genotypes to be disadvantaged in terms of their quality parameters and, therefore, to be less represented after the filtering steps. In this context, it will probably be necessary to evaluate alternative methods or to set up *ad hoc* strategies for a more reliable prediction of the genotypes, on which the result strongly depends. Moreover, the completion of the *de novo* assembly of the Adriatic sturgeon genome will permit a new, more reliable prediction of the effects of our selected mutations in its annotated genes, providing the possibility to confirm or correct the previous inferences and eventually to identify new mutations on the 100-150bp long regions sequenced during this genotyping. This new analysis would add more robustness to the final evaluation of the methodology, allowing to find

a better estimate of the relationship between predicted genetic load and the fitness observed in the individuals.

4. CONCLUSIONS

The Order of Acipenseriformes is one of the most endangered taxa according to the last IUCN Red List reassessment released by the IUCN Sturgeon Specialist Group.

For this reason, adequate management of residual genetic diversity in sturgeon and paddlefish represents a priority.

The present study demonstrates how the highly variable microsatellite *loci* still represent a valuable approach for the genetic analysis of these species that are characterised by various levels of ploidy. However, their application requires in some cases a preliminary assessment of their segregation patterns. In fact, two different inheritance patterns (disomic and tetrasomic) are possible in tetraploid species. A specifically designed experiment in which segregation was followed in the progeny of preliminarily genotyped breeders, maintained *ex situ* in the aquaculture plant ‘Storione Ticino’, revealed the coexistence of two different inheritance patterns in the tetraploid Adriatic sturgeon (*Acipenser naccarii*). This evidence shows that the level of homology between chromosomes within the Adriatic sturgeon genome can vary from chromosome to chromosome, and consequently different parts of the genome can have different degrees of functional ploidy, giving insight into the complexity of the genome structure in this species.

After this first assessment, the microsatellite *loci* that showed a tetrasomic inheritance pattern were successfully applied for parental allocation of first-generation offspring (F1) to their wild-origins progenitors (F0). A total of 357 F1 offspring were successfully allocated to 29 F0 parental pairs, while for 141 individuals no allocation was reached, probably because one or both F0 parents died before the first sampling occurred in 2004 on the parental stock. Nevertheless, for many of these non-allocated individuals, microsatellites allowed to infer full-sibs or half-sibs relatedness.

In general, the ‘Storione Ticino’ stock analysed with the above approach was characterised by a higher level of genetic diversity compared to other stocks established by public entities from a smaller number of siblings’ groups. These results allowed us to highlight the need of comparative transversal analyses across captive stocks, aimed at optimising the management of available genetic diversity and considering them as a single meta-stock for the designing of breeding plans.

The same approach based on microsatellite *loci* was successful in characterising a wild stock of new-born Chinese sturgeon (*Acipenser sinensis*), also allowing the estimate of a total number of about 10 females and 10 males participating in the reproductive event and possibly remaining for this species.

Relevant problems for sturgeon species are poaching and illegal trade, which compromise the successful recovery of wild populations. Sturgeons are also characterised by the possibility of producing many vital fertile hybrids, whose identification is not possible using standard approaches, usually based on mitochondrial markers. Therefore, the development of new markers for species and hybrid identification can greatly contribute to contrast the illegal trade of sturgeon-derived products. To this purpose, the genetic information harboured in intronic regions was explored in the Teleost taxa, for which more genomic resources are available. MIPs (Multi-locus Intron Polymorphisms) are simultaneously codominant and multiallelic nuclear markers characterised by high allelic variability comparable to that of microsatellite markers.

This high allelic variability also permits to compensate the lower number of *loci* characterised. Moreover, being the primers designed in highly conserved exonic regions, these markers allow to characterise a wide range of species even if no a priori genomic information is available.

Lastly, the sequence information permits to capture a higher genetic variation across multiple adjacent SNPs, providing a more detailed and accurate genetic profiling compared to individual SNPs. At the same time, the long sequences available facilitate the development of allele-specific PCR on single SNPs, which can be employed for genotyping, population allocation, and species identification.

After exploring the potentiality of intronic markers in Teleost species and demonstrating their efficacy at both species and population levels, the same approach was applied to sturgeons.

MIPs markers designed for sturgeons promised once again to be highly versatile, with many possible applications from species to populations levels in detecting differentiation signals, even if, for technical reasons, the correct genotype of analysed individuals was reliably recovered for only 18 *loci* out of the 96 tested. The problems were caused by the high frequency of duplicate *loci*, combined with the polyploidy that characterises many species of sturgeon. This resulted in a drastic drop in the average coverage observed for the different alleles, making it difficult to discriminate the real haplotypes from the artefacts. This line of research is still under development, and a second sequencing run of this intronic library must be performed to enhance the average coverage.

In general, genomic approaches based on Next Generation Sequencing (NGS) technologies still encounter several difficulties when used in the characterisation of polyploid species. As seen, the average sequencing coverage remains a critical parameter for the successful identification of alleles and inferences of genotypes. This last issue is especially true for less variable *loci*, such as biallelic SNPs, for which inference of the correct genotype, starting from the relative quality and quantity of reads that account for each allele, remains a difficult task. These problems were crucial to preventing a reliable estimate of the genetic load that had to be compared with the fitness indicators in our selected *Acipenser naccarii* individuals. Furthermore, the unavailability of the reference genome for this species does not yet allow a fully reliable prediction of deleterious mutations.

Therefore, it is not yet possible to achieve definitive results for this task, and *ad hoc* strategies must be implemented to obtain a more reliable genetic load on which ultimately the outcome of this study depends.

ANNEX A

Article

Different Chromosome Segregation Patterns Coexist in the Tetraploid Adriatic Sturgeon *Acipenser naccarii*

Stefano Dalle Palle ^{1,†}, Elisa Boscari ^{1,†} , Simone Giulio Bordignon ¹, Víctor Hugo Muñoz-Mora ², Giorgio Bertorelle ²  and Leonardo Congiu ^{1,*}

¹ Department of Biology, University of Padova, Via Ugo Bassi 58b, 35121 Padova, Italy

² Department of Life Sciences and Biotechnology, University of Ferrara, Via Luigi Borsari 46, 44121 Ferrara, Italy

* Correspondence: leonardo.congiu@unipd.it

† These authors contributed equally to this work.

Abstract: The Adriatic sturgeon, *Acipenser naccarii* (Bonaparte, 1836), is a critically endangered tetraploid endemism of the Adriatic region; it has been targeted, over the last 20 years, by different conservation programs based on controlled reproduction of captive breeders followed by the release of their juvenile offspring; its preservation would greatly benefit from the correct and coordinated management of the residual genetic variability available in the different captive stocks. In this sense, the setup of an efficient parental allocation procedure would allow identifying familiar groups and establishing informed breeding plans, effectively preserving genetic variation. However, being the species tetraploid, the analyses often deal with complex genome architecture and a preliminary evaluation of allele segregation patterns at different chromosomes is necessary to assess whether the species can be considered a pure tetraploid, as previously observed at some loci, or if a more complex situation is present. Here we study the segregation at 14 microsatellites loci in 12 familiar groups. Results support in different families the tetrasomic segregation pattern at 11 markers and the disomic segregation at three markers. The Adriatic sturgeon thus shows a mixed inheritance modality. In this species, and likely in other sturgeons, accurate knowledge of the loci used for paternity analysis is therefore required.

Keywords: acipenseridae; autotertraploid; allopolyploid; Adriatic sturgeon; disomic segregation; inheritance; microsatellites; tetrasomic segregation



Citation: Dalle Palle, S.; Boscari, E.; Bordignon, S.G.; Muñoz-Mora, V.H.; Bertorelle, G.; Congiu, L. Different Chromosome Segregation Patterns Coexist in the Tetraploid Adriatic Sturgeon *Acipenser naccarii*. *Diversity* **2022**, *14*, 745. <https://doi.org/10.3390/d14090745>

Academic Editors: Michael Wink and Simon Blanchet

Received: 13 August 2022

Accepted: 7 September 2022

Published: 10 September 2022

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



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sturgeons are the most endangered group of species, according to the International Union for Conservation of Nature (IUCN, July 2021) (<http://www.iucnredlist.org>, accessed on 8 August 2022). For this reason, sturgeons are targeted by several conservation efforts that often include restocking programs with juveniles produced in captivity; these ex-situ conservation activities must necessarily be supported by studies aimed at preserving the residual genetic diversity through long-term breeding programs [1]. However, genetic analyses on sturgeons deal with complex genomes and various levels of ploidy, due to independent events of whole-genome duplication [2,3]. The first event of duplication took place in the *Acipenseriformes*' common ancestor starting from sixty chromosomes. Then, secondary events of duplication occurred in the Pacific and Atlantic clades leading to a total of 240 chromosomes [2]. Finally, a third event led to the unique number of 360 chromosomes observable in *Acipenser brevirostrum* (Lesueur, 1818) [4]. The number of chromosomes associated with the distinct levels of ploidy has been the subject of an extensive debate between two main positions. The first argues that, since the condition with 120 chromosomes results from a duplication event in the common ancestor, the species with 120 and 240 chromosomes must be considered tetraploid and octaploid, respectively. The second position, taking into account the functional reduction of ploidy that follows whole

genome duplications, attributes to the two groups a condition of diploidy and tetraploidy, respectively [5]. The two views use different criteria to define the nominal ploidy, the number of duplications and the functional activity of genes, respectively, and are both correct and fully compatible [3].

The Adriatic sturgeon, *Acipenser naccarii* (Bonaparte, 1836), a species with 240 chromosomes, is in general considered a functional tetraploid based on the analyses of 28S and 5S rDNA through in situ hybridization [6]; this is also confirmed by most of the microsatellites analysed in this species which mostly show up to 4 alleles per individuals. However, some loci consistently showed more than 4 alleles in many individuals [7]; this could be due to the duplication of the region of such microsatellites but, a still incomplete process of functional reduction following genome duplication cannot be excluded as already proposed also for other sturgeon species [8,9].

For these reasons, loci to be used for applied purposes, such as parental allocation and kinship analysis, should be carefully selected and their functional ploidy should be preliminarily assessed. Moreover, these applications require a careful preliminary investigation also on the segregation modalities which, in tetraploids, can be of different types. In fact, polyploidization can originate through the fusion of unreduced gametes at intraspecific or interspecific levels, leading to two types of conditions in tetraploids, called autotetraploidy and allotetraploidy, respectively. The origins of polyploidization have important implications on the segregation patterns of the alleles within gametes. In complete autotetraploids, there are always four homologous chromosomes, and random pairs of bivalents and quadrivalents are possible during meiosis (see Figure 1 for graphical support) [10]; this condition leads to tetrasomic inheritance, which means that all allelic combinations within gametes are possible; this is not the case in complete allotetraploids, where each tetrad is composed of two sets of homeologous chromosomes, originating from the two parental species. In this situation, the homeologous chromosomes do not form pairs, leading to disomic inheritance, where only four out of six allelic combinations are possible [10]; these are two extreme cases and intermediate conditions are observable when the chromosomes have different degrees of preferential pairing. Indeed, the inheritance may shift from tetrasomic to disomic or vice versa [10]. For example, in autotetraploids, fertility and karyotype stability can be negatively impacted by imperfect multivalent pairing, thus promoting diploidization and consequent shifting to disomy; on the contrary, in allotetraploids, the homeologous chromosomes from two distinct parental species could maintain some degree of genetic affinity permitting the competition with the homologous pair during meiotic interactions with a certain degree of tetrasomic inheritance [11]. Understanding the mechanisms of chromosomal segregation in a tetraploid species can have important conservation implications, for example when the development of parental allocation methods is required.

In the Adriatic sturgeon, allele segregation was previously investigated at only 7 loci using microsatellite markers [12]. All loci showed a tetrasomic inheritance pattern, pointing at the probable autopolyploidization origin of this species. However, as secondary differentiation of some homologous chromosomes cannot be excluded and different segregation patterns can be followed by different chromosomes, we took advantage of the recent availability of new complete family groups and new isolated and tested microsatellite loci not yet explored, to provide a deeper insight into the mode of chromosome segregation in this species.

This is a key step not only to have a better understanding of the karyotype in the Adriatic sturgeon but also for the correct interpretation of the genetic analysis and parental allocation which in this species can have multiple applications. Firstly, the distinction of extremely rare individuals of wild origin from released ones. Secondly, the identification of groups of siblings existing in the different farms by assigning everyone to his pair of parents of the F0 generation. Finally, knowing the segregation patterns is crucial when the generation of virtual genotypes starting from observed genetic profiles is needed [13,14].

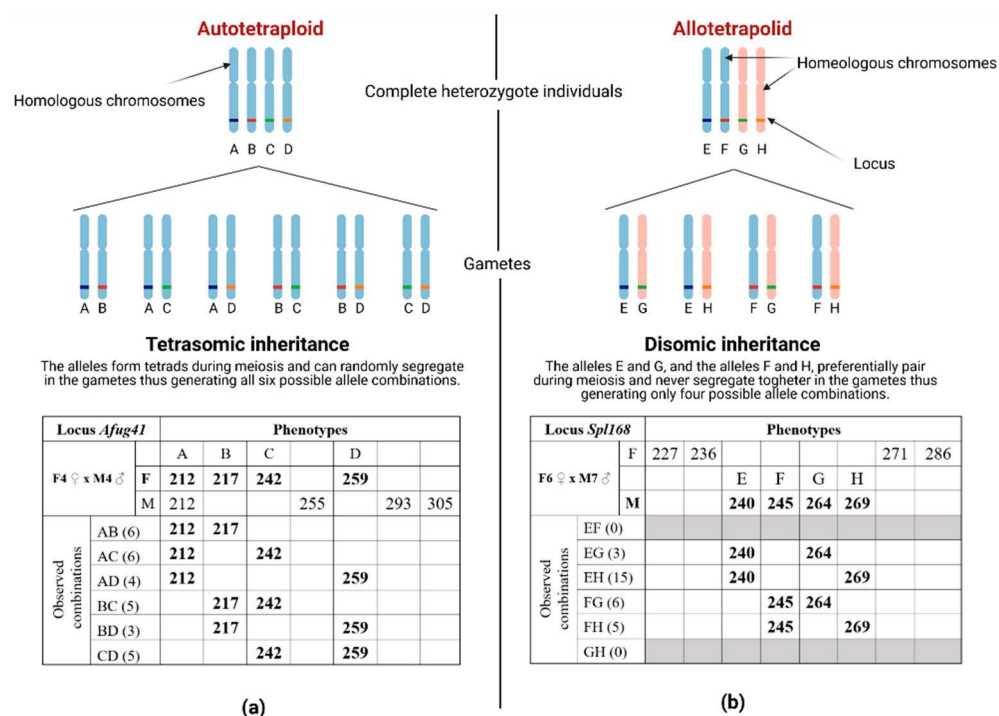


Figure 1. Expected segregation patterns under pure disomy (a) or tetrasomy (b), respectively expected in Autotetraploid and Allotetraploid genomes.

Thanks to the possibility of making 12 distinct crosses and raising their progeny separately within the project ENDEMIXIT (<https://endemixit.com/>, accessed on 8 August 2022), many new informative familiar groups became available for the analysis of segregation patterns at a much higher number of loci than those available in the past. The main purpose of the present study is therefore to exhaustively describe the modalities of microsatellite alleles inheritance in the Adriatic sturgeon with the following objectives: (a) verifying if the inferred pure tetrasomy is confirmed on a high number of loci, (b) providing a significant contribution to the management of the residual genetic diversity of this critically endangered species, and (c) shedding light on the functional ploidy level across its genome.

2. Materials and Methods

2.1. Samples and DNA Purification

Most samples analyzed in the present study were obtained from the aquaculture plant Storione Ticino (Cassolnovo, Italy). The reproduction of six females (F2, F4, F5, F6, F7, and F8) and six males (M2, M4, M5, M6, M7, and M8) was performed in different combinations. For each parent, a fin clip was collected for genotyping and, for each cross, the progeny was reared in captivity and spontaneously dead animals were collected and stored in ethanol. Moreover, five familiar groups (Nacc7 ♀ × Nacc5 ♂, Nacc8 ♀ × Nacc31 ♂, Nacc19 ♀ × Nacc17 ♂, Nacc19 ♀ × Nacc30 ♂, Nacc28 ♀ × Nacc23 ♂) derived from crosses performed in the past and for which fingerlings were available were also used. In total, 21 adult parents (12 parent pairs) and 376 fingerlings (with about 30 individuals per family) were collected (Table 1).

Genomic DNA was extracted from breeder's fin clips (10–100 mg) and from offspring's muscular tissue, using Euroclone spinNAker Universal Genomic DNA mini kit (Euroclone) and stored at -20°C till their processing for microsatellite analysis.

2.2. Selection of Loci and Genotyping

Loci analysed in the present study were selected based on the possibility of unambiguously tracing the genetic contribution of at least one of the two parents in at least one of the

available families. For this reason, all loci in which there were no complete heterozygotes or there were individuals with more than four alleles were discarded. In fact, it is known that the Adriatic sturgeon presents a minority of loci at which more than four alleles were can be observed in different individuals; thus, segregation anomalies cannot be detected [7]; these extra numerary alleles could originate from duplication or from a locally unreduced octaploid condition. In fact, we recall that the Adriatic sturgeon is functionally a tetraploid but evolutionarily it is an octaploid [3]. Nevertheless, segregation anomalies were also observed in some individuals of each family that were accordingly excluded from the analysis.

Table 1. Screening of microsatellite loci.

Locus	Ref	Ta	Size	Family	Parents	Genotypes	Genotyped Fingerling (Y/N)	#
LS-39	15	52 °C	116–155	Not informative	-	-	N	-
AfuG113	16	Td	326–364	Not informative	-	-	N	-
AfuG132	16	61 °C	259–346	F5 ♀ × M6 ♂	F5 ♀ M6 ♂	293/309/318/330 313/318/322	Y	32
				F6 ♀ × M7 ♂	F6 ♀ M7 ♂	313/318/342 293/309/318/322	Y	32
				F8 ♀ × M2 ♂	F8 ♀ M2 ♂	304/318/322 313/338/342/346	Y	30
AfuG41	16	58 °C	156–198	F8 ♀ × M2 ♂	F8 ♀ M2 ♂	212/234/293 203/255/259/285	Y	30
				F2 ♀ × M4 ♂	F2 ♀ M4 ♂	247/255/259/285 212/255/293/305	Y	30
				F4 ♀ × M4 ♂	F4 ♀ M4 ♂	212/217/242/259 212/255/293/305	Y	30
				F6 ♀ × M7 ♂	F6 ♀ M7 ♂	212/242/259/278 203/225/229/259	Y	32
				Nacc7 ♀ × Nacc5 ♂	Nacc7 ♀ Nacc5 ♂	225/229/252/259 203/225/247/305	Y	32
				Nacc8 ♀ × Nacc31 ♂	Nacc8 ♀ Nacc31 ♂	212/242/293/305 203/234/305	Y	32
An20 *	17	62 °C	159–213	F7 ♀ × M8 ♂	F7 ♀ M8 ♂	160/164/194 160/172/182/186	Y	30
Anac-15214	7	61 °C	259–285	Not informative	-	-	N	-
Anac-2589	7	63 °C	224–294	Not informative	-	-	N	-
Anac-6784	7	62 °C	311–346	Nacc19 ♀ × Nacc17 ♂	Nacc19 ♀ Nacc17 ♂	311/322/330/334 322/326	Y	32
Anac-3133	7	56 °C	164–178	F7 ♀ × M8 ♂	F7 ♀ M8 ♂	164/174 164/166/170/172	Y	30
AnacA6 *	18	62 °C	289–313	F5 ♀ × M6 ♂	F5 ♀ M6 ♂	307/313 293/297/301/307	Y	32
AnacB11	18	60 °C	132–162	F6 ♀ × M7 ♂	F6 ♀ M7 ♂	148/150 138/144/148/162	Y	32
				Nacc19 ♀ × Nacc17 ♂	Nacc19 ♀ Nacc17 ♂	132/136/138/144 132/150/162	Y	32
AnacB7	18	60 °C	152–198	F6 ♀ × M7 ♂	F6 ♀ M7 ♂	166/172/174/176 154/156/164	Y	32
				Nacc19 ♀ × Nacc30 ♂	Nacc19 ♀ Nacc30 ♂	156/166/174 154/170/174/176	Y	32
AnacC11	18	50 °C	167–193	Not informative	-	-	N	-
AnacE4 *	18	58 °C	326–354	F5 ♀ × M6 ♂	F5 ♀ M6 ♂	332/340/346/354 336/346	Y	32
AoxD161	19	60 °C	111–155	F4 ♀ × M5 ♂	F4 ♀ M5 ♂	123/127/131/139 131/135/155	Y	30
				F8 ♀ × M2 ♂	F8 ♀ M2 ♂	123/131/135/139 127/131	Y	30
				Nacc7 ♀ × Nacc5 ♂	Nacc7 ♀ Nacc5 ♂	123/127/131/143 131/135/139	Y	32

Table 1. Cont.

Locus	Ref	Ta	Size	Family	Parents	Genotypes	Genotyped Fingerling (Y/N)	#
AoxD234 *	19	52 °C	215–275	F2 ♀ × M4 ♂	F2 ♀ M4 ♂	219/223/243/255 227/243	Y	30
				F4 ♀ × M5 ♂	F4 ♀ M5 ♂	227/243/247/251 235/239/255/259	Y	30
				Nacc28 ♀ × Nacc23 ♂	Nacc28 ♀ Nacc23 ♂	219/243/247/263 239/243/251/255	Y	24
				F6 ♀ × M7 ♂	F6 ♀ M7 ♂	227/243/247/255 219/239/247	Y	32
AoxD241 *	19	57 °C	156–196	F7 ♀ × M8 ♂	F7 ♀ M8 ♂	168/176/180 164/172/176/184	Y	30
AoxD64 *	19	60 °C	216–260	Not informative	-	-	N	-
Spl-120	20	55 °C	263–303	Not informative	-	-	N	-
Spl-168	10	63 °C	200–314	F2 ♀ × M4 ♂	F2 ♀ M4 ♂	207/215/220 166/215/224/229	Y	30
				F7 ♀ × M8 ♂	F7 ♀ M8 ♂	232/240/294 218/236/271/273	Y	30
				F4 ♀ × M5 ♂	F4 ♀ M5 ♂	209/232/257/273 218/232/279/282	Y	30
				F6 ♀ × M7 ♂	F6 ♀ M7 ♂	227/236/271/286 240/245/264/269	Y	32
				Nacc19 ♀ × Nacc17 ♂	Nacc19 ♀ Nacc17 ♂	218/227/269/294 227/240/286/314	Y	32
				Nacc7 ♀ × Nacc5 ♂	Nacc7 ♀ Nacc5 ♂	232/245/264/294 214/240/273/279	Y	32
				Nacc8 ♀ × Nacc31 ♂	Nacc8 ♀ Nacc31 ♂	209/236/271/279 218/249/273/279	Y	32

Loci were used for genotyping of breeders for the selection of informative family/locus combinations for which progeny was also processed. References, annealing temperatures used, and size are reported. For each informative family/locus combination, the genotypes of each parent at such loci are also reported and the parent's genotypes for which the allele segregation was followed in the progeny is in bold. The total number of fingerlings processed for each informative family is reported in the last column. * Loci for which the segregation inheritance pattern was already explored in other familiar groups of the same species [12]. Y = Yes, N = Not processed. # = number of analyzed individuals.

All breeders were genotyped at 21 microsatellite loci (Table 1) [7,15–20] to select the informative Family/Locus combinations for the assessment of chromosomal segregation. For each selected combination, the progeny was also amplified and genotyped. Tracking segregation in the progeny requires the satisfaction of some features, such as (i) the complete heterozygosity (four different alleles) of at least one parent to ensure that each allele can unambiguously mark the segregation of its own chromosome and (ii) no more than an allele shared by the two parents to avoid ambiguity in following the alleles transmission to the progeny [12].

A total of 33 Family/Locus combinations were finally selected and approximately 30 fingerlings each were genotyped (Table 1). In 22 out of 33 case studies only one breeder of the parent pair was completely heterozygote and therefore informative to follow the segregation of the alleles, while for the remnant 11 Family/Locus combinations the segregation pattern was followed for both breeders of the parent pair as both they showed completely heterozygous genotypes with a single allele shared at most (Table 1).

Microsatellite loci were amplified from genomic DNA in a 10 µL reaction: 1X Master Mix Buffer (QIAGEN), 0.2 µM of each primer and about 50 ng of genomic DNA. Amplifications were performed on SimpliAmp Thermal Cycler (Thermo Fisher Scientific, Bologna, Italy). Amplifications were checked on 1.8% agarose gel in TBE1X stained with GelRed (BIOTIUM, Fremont, Canada, GelRed™ Nucleic Acid Stain). Genotyping was performed on ABI PRISM 310 Genetic Analyzer (external service, BMR Genomics, Padova, Italy). For microsatellite scoring the software GeneMarker Version 1.95 (SoftGenetics LLC, State College, PA, USA) was used.

2.3. Data Analyses

Scoring was performed by two operators, independently, with GeneMarker Version 1.95 (SoftGenetics LLC). The final dataset was carefully checked, and each individual was controlled for perfect Mendelian inheritance within his Family/Locus combination under the assumption that the observed genotypes at a given locus should be compatible with the inheritance of two allele copies from each parent. Individuals that for different possible reasons discussed later did not match this assumption were discarded. Since discordances with the above-accepted criterion can be attributed to anomalies in chromosomal segregation in the gametes of one parent, which are independent of what happened in the mating partner's gametes, it was decided not to discard those animals whose segregation anomaly was due to a problem in the uninformative parent. Limiting our observations to those animals that met standard inheritance of two out of four alleles, we consciously excluded possible genotypes that were not necessarily derived from segregation anomalies such as the case of double reduction, a phenomenon widely observed in autotetraploid plants in which the four homolog chromosomes may form multivalents [21,22] in which identical alleles carried on the sister chromatids may enter the same gamete with consequent segregation of two copies of the same allele even if it was present in a single copy in the parental genome [23]. We have decided to limit ambiguities by disregarding this phenomenon and excluding all the individuals that could be compatible with it focusing on animals in which alleles were unambiguously traceable.

As proposed by Stift et al. (2008) [10], Likelihood Ratio Tests (LRT) with 1 df were applied to compare the null model of tetrasomy with the other alternative models intermediate between disomic and tetrasomic. For each parent-locus combination and each alternative inheritance model, the log-likelihood was estimated from constrained nonlinear regression models, using SPSS syntax as reported in the original reference [10]. The Sequential Bonferroni correction [24] was applied to adjust significance levels ($p < 0.05$) for multiple comparisons across loci and families.

3. Results

From 22 to 32 individuals for each family-locus combination were successfully genotyped. In some cases, a few animals were discarded for unreliable profiles or, after the allele scoring, due to the presence of segregation anomalies of alleles inherited from the informative parent (Table 2).

At 11 (AfuG132, AfuG41, An20, Anac6784, Anac3133, AnacA6, AnacB11, AnacB7, AnacE4, AoxD234, AoxD241) out of 14 informative loci the LRT did not lead to the rejection of the null model of inheritance thus supporting the hypothesis of tetrasomy. At almost all these loci, indeed, all six allele combinations were observed in almost all tested families (Figure 1a, Table 2, Figure A1 of Appendix A). The only exceptions were the loci AoxD234 and Anac-3133 at a single family each showing only five allele combinations. However, in both cases, the strict disomic inheritance was excluded after correction for multiple tests (Table 2). In these two families, the sixth combination is expected to be detectable by increasing the sample size. Five (An20, AnacA6, AnacE4, AoxD234, and AoxD241) of the 11 loci showing tetrasomic inheritance were already tested to assess the inheritance pattern in different familiar groups of Adriatic sturgeon [12] and their tetrasomy has been here confirmed. Specifically, the locus AnacA6 here analysed at a single family showed only four allele combinations but, missing combinations are not compatible with disomic inheritance modality and even in this case the null model was not rejected.

On the contrary, a significant rejection of the null model was observed at three loci, Spl163, AoxD161, and Spl168, analysed respectively at one, three and six families and never analysed before in other studies. In some cases, the analysed families have both parents informative for segregation and agree in suggesting a disomic mode of inheritance (Table 2, Figure 1b, Figure A1 Appendix A). The disomic model fitted the observed allele combination frequencies significantly better than the null model, and the parameter τ equal to zero indicates a full disomic inheritance at almost all Family/Locus combinations. The

only two exceptions were observed in the segregation of the alleles of female F8 at locus AoxD161 and male Nacc31 at locus Spl168 (Table 2). In these two cases, five combinations of alleles were present, the parameter τ assumed a very low value confirming a strong degree of preferential pairing during meiosis but the rejection of the null model was not significant after the Bonferroni correction, thus suggesting possible imperfect preferential pairing.

Table 2. Likelihood Ratio Test.

Locus	Family	Informative Parent	N (Nd)	Null Model (T = 1) Likelihood Obs	Best Intermediate Model			Model Comparison: LRT	p-Values
					Pairing Alleles	T	Likelihood Obs		
Afug132	F5 ♀ × M6 ♂	F5 ♀	22 (8)	39.42	AC/BD	0.82	39.23	0.19	0.3322
	F6 ♀ × M7 ♂	M7 ♂	29 (1)	51.96	AD/BC	0.83	51.74	0.22	0.3185
	F8 ♀ × M2 ♂	M2 ♂	29 (1)	51.96	AB/CD	0.83	51.74	0.22	0.3185
Afug41	F8 ♀ × M2 ♂	M2 ♂	27 (3)	48.38	AC/BD	0.89	48.29	0.09	0.3853
	F2 ♀ × M4 ♂	F2 ♀	27 (3)	48.38	AB/CD	0.78	48.03	0.35	0.2776
		M4 ♂	27 (3)	48.38	AB/CD	0.44	45.98	2.39	0.0609
	F4 ♀ × M4 ♂	F4 ♀	29 (1)	51.96	AC/BD and AD/BC	0.93	51.93	0.03	0.4259
		M4 ♂	29 (1)	51.96	AD/BC	0.83	51.74	0.22	0.3185
	F6 ♀ × M7 ♂	F6 ♀	28 (2)	50.17	AD/BC	0.86	50.02	0.15	0.3509
		M7 ♂	28 (2)	50.17	AB/CD	0.86	50.02	0.15	0.3509
	Nacc7 ♀ × Nacc5 ♂	Nacc7 ♀	28 (4)	50.17	AD/BC	0.64	49.21	0.96	0.1631
		Nacc5 ♂	28 (4)	50.17	AD/BC	0.75	49.71	0.46	0.2489
	Nacc8 ♀ × Nacc31 ♂	Nacc8 ♀	31 (1)	55.54	AD/BC	0.77	55.13	0.41	0.2603
An20 *	F7 ♀ × M8 ♂	M8 ♂	30 (0)	53.75	AD/BC	0.70	53.03	0.72	0.1984
Anac6784	Nacc19 ♀ × Nacc17 ♂	Nacc19 ♀	30 (2)	53.75	AC/BD and AD/BC	0.91	53.68	0.08	0.3912
Anac3133	F7 ♀ × M8 ♂	M8 ♂	29 (1)	51.96	AB/CD	0.41	49.06	2.90	0.0444 ^a
AnacA6 *	F5 ♀ × M6 ♂	M6 ♂	29 (1)	51.96	AB/CD and AC/BD	0.72	51.38	0.58	0.2225
AnacB11	F6 ♀ × M7 ♂	M7 ♂	26 (4)	46.59	AB/CD	0.58	45.31	1.28	0.1290
	Nacc19 ♀ × Nacc17 ♂	Nacc19 ♀	26 (6)	46.59	AB/CD	0.81	46.34	0.25	0.3088
AnacB7	F6 ♀ × M7 ♂	F6 ♀	27 (3)	48.38	AB/CD	0.67	47.57	0.80	0.1849
	Nacc19 ♀ × Nacc30 ♂	Nacc30 ♂	32 (0)	57.34	AD/BC	0.94	57.30	0.03	0.4295
AnacE4 *	F5 ♀ × M6 ♂	F5 ♀	28 (2)	50.17	AB/CD and AD/BC	0.96	50.16	0.01	0.4622
AoxD161	F4 ♀ × M5 ♂	F4 ♀	29 (1)	51.96	AB/CD	0.00	40.20	11.76	0.0003
	F8 ♀ × M2 ♂	F8 ♀	30 (0)	53.75	AC/BD	0.10	45.28	8.47	0.0018 ^a
	Nacc7 ♀ × Nacc5 ♂	Nacc7 ♀	32 (0)	57.34	AD/BC	0.00	44.36	12.97	0.0002
AoxD234 *	F2 ♀ × M4 ♂	F2 ♀	27 (3)	48.38	AD/BC	0.89	48.29	0.09	0.3853
	F4 ♀ × M5 ♂	F4 ♀	28 (2)	50.17	AD/BC	0.86	50.02	0.15	0.3509
		M5 ♂	28 (2)	50.17	AD/BC	0.64	49.21	0.96	0.1631
	Nacc28 ♀ × Nacc23 ♂	Nacc28 ♀	23 (1)	41.21	AB/CD	0.78	40.93	0.28	0.2972
		Nacc23 ♂	23 (1)	41.21	AC/BD	0.26	37.29	3.92	0.0239 ^a
	F6 ♀ × M7 ♂	F6 ♀	29 (1)	51.96	AC/BD	0.52	50.07	1.89	0.0844
	F7 ♀ × M8 ♂	M8 ♂	29 (1)	51.96	AC/BD	0.72	51.38	0.58	0.2225
AoxD241 *	F2 ♀ × M4 ♂	M4 ♂	28 (2)	50.17	AC/BD	0.00	38.82	11.35	0.0004
Spl168	F7 ♀ × M8 ♂	M8 ♂	26 (4)	46.59	AB/CD	0.00	36.04	10.54	0.0006
	F4 ♀ × M5 ♂	F4 ♀	30 (0)	53.75	AB/CD	0.00	41.59	12.16	0.0002
		M5 ♂	30 (0)	53.75	AB/CD	0.00	41.59	12.16	0.0002
	F6 ♀ × M7 ♂	F6 ♀	29 (1)	51.96	AB/CD	0.00	40.20	11.76	0.0003
		M7 ♂	29 (1)	51.96	AB/CD	0.00	40.20	11.76	0.0003
	Nacc19 ♀ × Nacc17 ♂	Nacc19 ♀	23 (9)	41.21	AB/CD	0.00	31.88	9.33	0.0011
		Nacc17 ♂	23 (9)	41.21	AB/CD	0.00	31.88	9.33	0.0011
		Nacc17 ♂	23 (9)	41.21	AB/CD	0.00	31.88	9.33	0.0011
	Nacc7 ♀ × Nacc5 ♂	Nacc7 ♀	32 (0)	57.34	AB/CD	0.00	44.36	12.97	0.0002
		Nacc5 ♂	32 (0)	57.34	AB/CD	0.00	44.36	12.97	0.0002

Table 2. Cont.

Locus	Family	Informative Parent	N (Nd)	Null Model (T = 1) Likelihood Obs	Best Intermediate Model			Model Comparison: LRT	p-Values
					Pairing Alleles	T	Likelihood Obs		
Nacc8 ♀ × Nacc31 ♂		Nacc8 ♀	29 (3)	51.96	AB/CD	0.00	40.20	11.76	0.0003
		Nacc31 ♂	29 (3)	51.96	AB/CD	0.10	43.86	8.10	0.0022 ^a

Results of Likelihood Ratio Test between the null model of tetrasomy (TAU = 1) and the best fitting intermediate one (TAU estimated). Significant *p*-values (after Bonferroni correction) that reject the null hypothesis of tetrasomy are highlighted in grey. ^a not significant values after Bonferroni correction for multiple tests. * Loci for which the segregation inheritance pattern was already explored in other familiar groups of the same species [12].

Expected allele combinations in tetrasomic and disomic mode of inheritance in tetraploids. (a) Observed allele combinations inherited from a complete heterozygote parent in autotetraploids in which tetrads are generated during meiosis and random segregation of allele pairs is expected. In brackets, the number of individuals carrying the relative allele combination is reported. (b) Observed allele combinations inherited from a complete heterozygote parent in allotetraploids in which a preferential pair between homologous chromosomes occurs. Only four possible gametes are expected. Alleles of the parent for which the segregation is reported in each table are marked in bold and by a letter used to indicate the observed combinations. Complete tables for all loci are reported in Appendix A. (Created with [BioRender.com](https://www.biorender.com), accessed on 8 August 2022).

4. Discussion

The segregation pattern observed in the present study at most loci indicates that the Adriatic sturgeon can be considered predominantly tetraploid with a tetrasomic inheritance pattern. The presence of three disomic loci, however, indicates that the functional diversification process is at different stages in different parts of the genome, with some regions possibly still octaploid, most tetraploid, and some others in which the degree of divergence has gone up to a condition of double diploidy. We also observed the presence of two loci with a marked tendency to disomy with a few unexpected allele combinations, pointing to a possible imperfect preferential pairing expected at intermediate stages of functional diploidization [10,11]; this coexistence of different ploidy levels was previously described in other organisms. In plants, for example, the co-existence of different segregation patterns (e.g., tetrasomic and disomic) with different intermediate degrees of preferential pairing among chromosomes provides evidence of a process of functional reduction of ploidy which reasonably cannot occur simultaneously throughout the genome, but which is the result of a progressive differentiation [11,25].

As for the sturgeons, in other species the presence of different degrees of ploidy has been deduced based on the maximum number of alleles per individual present at the different loci [16] and, in some cases, the Mendelian transmission in the progeny has also been verified [15,26,27]. However, the selection of the cross/locus combinations to allow the traceability of every single allele and consequently to distinguish between the different modes of tetraploid segregation (e.g., disomic, tetrasomic or intermediate) has been conducted to our knowledge only on the Adriatic sturgeon.

The evidence that the level of homology between chromosomes within the Adriatic sturgeon genome is likely to vary from chromosome to chromosome and that different parts of the genome may consequently have different degrees of functional ploidy (2 to 8) should be considered when characterizing the genome of this species and probably of sturgeons in general. Genome assembly procedures must contemplate the possibility that different regions are present with a variable number of copies.

Whatever the evolutionary origin and implications of our results, which can be better investigated only when the genome of this species will be available, the identification of different inheritance pattern at different markers may have practical consequences for the conservation of the Adriatic sturgeon. In fact, our findings suggest that before developing parental allocation methods, a preliminary analysis of the loci used is recommended; this

would ease the reconstruction of individual genealogies of animals kept as captive breeders and the reallocation of any individual recaptured after release. Another interesting and relatively unexplored aspect of sturgeon *ex situ* conservation in which a clear knowledge of the inheritance patterns at different loci could be useful is the monitoring of the genomic impact of breeding protocols. Captive breeding is usually performed following standardized protocols of hormonal induction of egg and milt release and fertilization is done in an excess of sperms. Then, the resulting viable progeny is released without verifying if the procedure used had some effect on their genomic asset, for example, by inducing aneuploidies, which are a common phenomenon in some captive sturgeon stocks [28]. The release in nature of genetically anomalous individuals should be avoided as the consequences that this can have on the following generations and on their reproductive efficiency is unknown and, in the case of animals with long generation times, could be revealed after decades. Random screening of familiar groups to verify the correct parents-to-progeny segregation at both disomic and tetrasomic loci could significantly contribute to reducing the potential impact of genomic anomalies on natural populations.

5. Conclusions

Thanks to the availability of some family groups not previously analyzed, it was possible to better investigate the patterns of chromosomal segregation in the polyploid Adriatic sturgeon. The picture that emerged is that of an extremely dynamic genome in which it is possible to find the co-existence of regions with different degrees of ploidy, some of which retain the legacy of ancient duplications and others show a dynamic reduction of functional ploidy; this pattern is probably shared with other polyploid sturgeons in which different patterns of segregation have been observed [27], but it is not known whether the same genomic regions are involved.

Additional studies are required to better characterize the distribution of ploidy across the genome; this will likely contribute to explaining some specific features of the sturgeons, such as the ability of species with different degrees of ploidy to hybridize and produce viable offspring [29]; this extreme genomic plasticity could somehow be linked to a high degree of genomic redundancy; it would also be very interesting to verify whether the regions with different segregation patterns are somehow related to the size of the chromosomes, which in sturgeons is known to be very variable, with a small number of large chromosomes and a high number of micro-chromosomes. However, until complete genomes of good quality are available, it is difficult to move beyond speculation on these topics.

Author Contributions: Conceptualization, S.D.P., E.B. and L.C.; methodology, S.D.P., E.B. and L.C.; investigation, S.D.P. and S.G.B.; writing—original draft preparation, S.D.P. and L.C.; writing—review and editing, E.B., V.H.M.-M. and G.B.; project administration, G.B. and L.C.; funding acquisition, G.B. and L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Italian PRIN 2017, grant number 201794ZXTL.

Institutional Review Board Statement: The animal study protocol was approved by the Institutional Ethics Committee of the University of Padova (OPBA)—(protocol code 421922 of 14 October 2020).

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank the Province of Vicenza and the “Bacino Pesca B (VI)” for the hospitality in the “Oasi le Sorgenti” Natural Reserve and for the logistical support and the Storione Ticino aquaculture farm for performing all crosses that allowed this study.

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

Appendix A

Locus <i>Afug132</i>		Phenotypes					
F5 ♀ x M6 ♂		A	B		C		D
		293	309		318		330
				313	318	322	
Observed combinations	AB (2)	293	309				
	AC (3)	293			318		
	AD (6)	293					330
	BC (2)		309		318		
	BD (3)		309				330
	CD (6)				318		330
Locus <i>Afug132</i>		Phenotypes					
F6 ♀ x M7 ♂				313	318		342
		E	F		G	H	
		293	309		318	322	
Observed combinations	EF (7)	293	309				
	EG (7)	293			318		
	EH (4)	293				322	
	FG (4)		309		318		
	FH (4)		309			322	
	GH (3)				318	322	
Locus <i>Afug132</i>		Phenotypes					
F8 ♀ x M2 ♂		304		318	322		
			E			F	G
			313			338	342
Observed combinations	EF (4)		313			338	
	EG (1)		313				342
	EH (8)		313				346
	FG (4)					338	342
	FH (8)					338	346
	GH (4)						342
Locus <i>Afug41</i>		Phenotypes					
F8 ♀ x M2 ♂			212	234			293
		E			F	G	H
		203			255	259	285
Observed combinations	EF (3)	203			255		
	EG (4)	203				259	
	EH (4)	203					285
	FG (5)				255	259	
	FH (4)				255		285
	GH (7)					259	285
Locus <i>Afug41</i>		Phenotypes					
F2 ♀ x M4 ♂			A	B	C	D	
			247	255	259	285	
		E		F			G
Observed combinations	AB (4)		247	255			
	AC (4)		247		259		
	AD (7)		247			285	
	BC (3)			255	259		
	BD (6)			255		285	
	CD (3)				259	285	
Locus <i>Afug41</i>		Phenotypes					
F2 ♀ x M4 ♂			A	B	C	D	
			247	255	259	285	
		E		F			G
Observed combinations	EF (2)	212		255			
	EG (5)	212					293
	EH (7)	212					305
	FG (5)			255			293
	FH (6)			255			305
	GH (2)						293
Locus <i>Afug41</i>		Phenotypes					
F4 ♀ x M4 ♂		A	B	C		D	
		212	217	242		259	
		E			F		G
Observed combinations	AB (6)	212	217				
	AC (6)	212		242			
	AD (4)	212				259	
	BC (5)		217	242			
	BD (3)		217			259	
	CD (5)			242		259	
Locus <i>Afug41</i>		Phenotypes					
F4 ♀ x M4 ♂		A	B	C		D	
		212	217	242		259	
		E			F		G
Observed combinations	EF (6)	212			255		293
	EG (5)	212					293
	EH (3)	212					305
	FG (5)				255		293
	FH (6)				255		305
	GH (4)						293

Figure A1. Cont.

Locus <i>Afug132</i>		Phenotypes					
F5 ♀ x M6 ♂		A	B		C		D
		293	309		318		330
Observed combinations	AB (2)	293	309				
	AC (3)	293			318		
	AD (6)	293					330
	BC (2)		309		318		
	BD (3)		309				330
	CD (6)				318		330
Locus <i>Afug132</i>		Phenotypes					
F6 ♀ x M7 ♂				313	318		342
		E	F		G	H	
Observed combinations	EF (7)	293	309				
	EG (7)	293			318		
	EH (4)	293				322	
	FG (4)		309		318		
	FH (4)		309			322	
	GH (3)				318	322	
Locus <i>Afug41</i>		Phenotypes					
F2 ♀ x M4 ♂		A	B	C	D		
		247	255	259	285		
Observed combinations	AB (4)	247	255				
	AC (4)	247		259			
	AD (7)	247			285		
	BC (3)		255	259			
	BD (6)		255		285		
	CD (3)			259	285		
Locus <i>Afug41</i>		Phenotypes					
F4 ♀ x M4 ♂		A	B	C		D	
		212	217	242		259	
Observed combinations	AB (6)	212	217				
	AC (6)	212		242			
	AD (4)	212			259		
	BC (5)		217	242			
	BD (3)		217		259		
	CD (5)			242	259		
Locus <i>Afug41</i>		Phenotypes					
F8 ♀ x M2 ♂		304		318	322		
		E			F	G	H
Observed combinations	EF (4)		313			338	342
	EG (1)		313				
	EH (8)		313				346
	FG (4)					338	342
	FH (8)					338	346
	GH (4)					342	346
Locus <i>Afug41</i>		Phenotypes					
F8 ♀ x M2 ♂			212	234			293
		E			F	G	H
Observed combinations	EF (3)	203			255	259	285
	EG (4)	203				259	
	EH (4)	203					285
	FG (5)				255	259	
	FH (4)				255		285
	GH (7)					259	285

Figure A1. Cont.

Locus <i>A6</i>		Phenotypes				
F5 ♀ x M6 ♂					307	313
		E	F	G	H	
		293	297	301	307	
Observed combinations	EF (1)	293	297			
	EG (7)	293		301		
	EH (0)					
	FG (15)		297	301		
	FH (0)					
	GH (6)			301	307	

Locus <i>B11</i>		Phenotypes				
F6 ♀ x M7 ♂				148	150	
		E	F	G		H
		138	144	148		162
Observed combinations	EF (2)	138	144			
	EG (6)	138		148		
	EH (3)	138				
	FG (4)		144	148		
	FH (8)		144			162
	GH (3)				148	162

Locus <i>B11</i>		Phenotypes				
Nacc19♀ x Nacc17♂		A	B	C	D	
		132	136	138	144	
		132				150 162
Observed combinations	AB (4)	132	136			
	AC (7)	132		138		
	AD (3)	132			144	
	BC (5)		136	138		
	BD (4)		136		144	
	CD (3)			138	144	

Locus <i>B7</i>		Phenotypes						
F6 ♀ x M7 ♂					A	B	C	D
					166	172	174	176
		154	156	164				
Observed combinations	AB (3)				166	172		
	AC (8)				166		174	
	AD (6)				166			176
	BC (5)					172	174	
	BD (2)					172		176
	CD (3)						174	176

Locus <i>B7</i>		Phenotypes					
Nacc19♀ x Nacc30♂			156	166		174	
		E			F	G	H
		154			170	174	176
Observed combinations	EF (8)	154			170		
	EG (8)	154				174	
	EH (4)	154					176
	FG (6)				170	174	
	FH (3)				170		176
	GH (3)					174	176

Locus <i>E4</i>		Phenotypes				
F5 ♀ x M6 ♂		A		B	C	D
		332		340	346	354
			336		346	
Observed combinations	AB (3)	332		340		
	AC (5)	332			346	
	AD (3)	332				354
	BC (6)			340		
	BD (5)			340		354
	CD (6)				346	354

Locus <i>Aox161</i>		Phenotypes				
F4 ♀ x M5 ♂		A	B	C		D
		123	127	131		139
				131	135	155
Observed combinations	AB (0)					
	AC (10)	123		131		
	AD (7)	123				139
	BC (6)		127	131		
	BD (6)		127			139
	CD (0)					

Locus <i>Aox161</i>		Phenotypes				
F8 ♀ x M2 ♂		A		B	C	D
		123		131	135	139
			127	131		
Observed combinations	AB (7)	123		131		
	AC (1)	123			135	
	AD (6)	123				139
	BC (5)			131	135	
	BD (0)					
	CD (11)				135	139

Figure A1. Cont.

Locus <i>AoxD161</i>		Phenotypes					
Nacc7 ♀ x Nacc5 ♂		A	B	C			D
		123	127	131			143
				131	135	139	
Observed combinations	AB (24)	123	127				
	AC (6)	123		131			
	AD (0)						
	BC (0)						
	BD (7)		127				143
	CD (5)			131			143

Locus <i>Aox234</i>		Phenotypes				
F2 ♀ x M4 ♂		A	B		C	D
		219	223		243	255
				227	243	
Observed combinations	AB (8)	219	223			
	AC (5)	219			243	
	AD (4)	219				255
	BC (4)		223		243	
	BD (4)		223			255
	CD (2)				243	255

Locus <i>Aox234</i>		Phenotypes							
F4 ♀ x M5 ♂		A			B	C	D		
		227			243	247	251		
			E	F				G	H
Observed combinations			235	239				255	259
	AB (6)	227			243				
	AC (4)	227				247			
	AD (6)	227					251		
	BC (2)				243	247			
	BD (7)				243		251		
	CD (3)					247	251		

Locus <i>Aox234</i>		Phenotypes							
F4 ♀ x M5 ♂		A			B	C	D		
		227			243	247	251		
			E	F				G	H
Observed combinations			235	239				255	259
	EF (8)		235	239					
	EG (4)		235					255	
	EH (4)		235						259
	FG (2)			239				255	
	FH (4)			239					259
	GH (6)							255	259

Locus <i>Aox234</i>		Phenotypes						
Nacc28 ♀ x Nacc23 ♂		A		B	C			D
		219		243	247			263
			E	F		G	H	
Observed combinations			239	243		251	255	
	AB (4)	219		243				
	AC (7)	219			247			
	AD (3)	219						263
	BC (4)			243	247			
	BD (3)			243				263
	CD (2)				247			263

Locus <i>Aox234</i>		Phenotypes						
Nacc28 ♀ x Nacc23 ♂		A		B	C			D
		219		243	247			263
			E	F		G	H	
Observed combinations			239	243		251	255	
	EF (6)		239	243				
	EG (2)		239			251		
	EH (6)		239				255	
	FG (5)			243		251		
	FH (0)							
	GH (4)					251	255	

Locus <i>Aox234</i>		Phenotypes					
F6 ♀ x M7 ♂			A		B	C	D
			227		243	247	255
		219		239		247	
Observed combinations	AB (3)		227		243		
	AC (1)		227			247	
	AD (8)		227				255
	BC (5)				243	247	
	BD (4)				243		255
	CD (8)						255

Locus <i>AoxD241</i>		Phenotypes					
F7 ♀ x M8 ♂			168		176	180	
				F	G		H
		164		172	176		184
Observed combinations	EF (2)	164		172			
	EG (2)	164			176		
	EH (5)	164					184
	FG (9)			172	176		
	FH (5)			172			184
	GH (6)				176		184

Locus <i>Spl163</i>		Phenotypes				
F2 ♀ x M4 ♂			207	215	220	
		E		F		G H
		166		215		224 229
Observed combinations	EF (11)	166		215		
	EG (0)					
	EH (5)	166				229
	FG (7)			215		224
	FH (0)					
	GH (5)					224 229

Locus <i>Spl168</i>		Phenotypes				
F7 ♀ x M8 ♂			232		240	
		E		F	G	H
		218		236		271 273
Observed combinations	EF (0)					
	EG (5)	218			271	
	EH (6)	218				273
	FG (7)			236	271	
	FH (8)			236		273
	GH (0)					

Figure A1. Cont.

Locus <i>Spl168</i>		Phenotypes						
F4 ♀ x M5 ♂		A		B	C	D		
		209		232	257	273		
			E	F			G	H
			218	232			279	282
Observed combinations	AB (0)							
	AC (10)	209			257			
	AD (6)	209				273		
	BC (7)			232	257			
	BD (7)			232		273		
	CD (0)							

Locus <i>Spl168</i>		Phenotypes						
F4 ♀ x M5 ♂		A		B	C	D		
		209		232	257	273		
			E	F			G	H
			218	232			279	282
Observed combinations	EF (0)							
	EG (7)		218				279	
	EH (2)		218					282
	FG (13)			232			279	
	FH (8)			232				282
	GH (0)							

Locus <i>Spl168</i>		Phenotypes						
F6 ♀ x M7 ♂		A	B				C	D
		227	236				271	286
				E	F	G	H	
				240	245	264	269	
Observed combinations	AB (0)							
	AC (6)	227					271	
	AD (4)	227						286
	BC (9)		236				271	
	BD (10)		236					286
	CD (0)							

Locus <i>Spl168</i>		Phenotypes						
F6 ♀ x M7 ♂		A	B				C	D
		227	236				271	286
				E	F	G	H	
				240	245	264	269	
Observed combinations	EF (0)							
	EG (3)			240		264		
	EH (15)			240			269	
	FG (6)				245	264		
	FH (5)				245		269	
	GH (0)							

Locus <i>Spl168</i>		Phenotypes						
Nacc19 ♀ x Nacc17 ♂		A	B		C		D	
		218	227		269		294	
			E	F		G		H
			227	240		286		314
Observed combinations	AB (0)							
	AC (4)	218			269			
	AD (7)	218					294	
	BC (3)		227		269			
	BD (9)		227				294	
	CD (0)							

Locus <i>Spl168</i>		Phenotypes						
Nacc19 ♀ x Nacc17 ♂		A	B		C		D	
		218	227		269		294	
			E	F		G		H
			227	240		286		314
Observed combinations	EF (0)							
	EG (4)		227			286		
	EH (5)		227					314
	FG (8)			240		286		
	FH (6)			240				314
	GH (0)							

Locus <i>Spl168</i>		Phenotypes						
Nacc7 ♀ x Nacc5 ♂		A		B	C			D
		232		245	264			294
		E		F		G	H	
		214		240		273	279	
Observed combinations	AB (0)							
	AC (12)		232		264			
	AD (9)		232					294
	BC (5)			245	264			
	BD (6)			245				294
	CD (0)							

Locus <i>Spl168</i>		Phenotypes						
Nacc7 ♀ x Nacc5 ♂		A		B	C			D
		232		245	264			294
		E		F		G	H	
		214		240		273	279	
Observed combinations	EF (0)							
	EG (8)	214				273		
	EH (7)	214					279	
	FG (7)			240		273		
	FH (10)			240			279	
	GH (0)							

Figure A1. Cont.

Locus <i>Spl168</i>		Phenotypes							
		A		B		C		D	
Nacc8♀ x Nacc 31♂		209		236		271		279	
			E		F		G		H
			218		249		273		279
Observed combinations	AB (0)								
	AC (9)	209				271			
	AD (7)	209						279	
	BC (6)			236		271			
	BD (7)			236				279	
	CD (0)								

Locus <i>Spl168</i>		Phenotypes							
		A		B		C		D	
Nacc8♀ x Nacc 31♂		209		236		271		279	
			E		F		G		H
			218		249		273		279
Observed combinations	EF (1)		218		249				
	EG (8)		218				273		
	EH (4)		218					279	
	FG (11)				249		273		
	FH (5)				249			279	
	GH (0)								

Figure A1. Total segregation results. Single-locus microsatellite inheritance for each studied sturgeon family. Microsatellite alleles of parents are reported as sizes in bp. For families in which both parents are informative, two different schemes are shown. The 4 alleles of the informative parents are highlighted in bold and marked with capital letters A, B, C, D for females and E, F, G, H for males, also used to label allele combinations observed in the progeny. The number of F1 in which each combination has been observed is reported in brackets. Missing combinations are marked in grey.

References

- Ludwig, A. A Sturgeon View on Conservation Genetics. *Eur. J. Wildl. Res.* **2005**, *52*, 3–8. [\[CrossRef\]](#)
- 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. Phylogenetics Evol.* **2007**, *42*, 854–862. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fontana, F.; Congiu, L.; Mudrak, V.A.; Quattro, J.M.; Smith, T.I.; Ware, K.; Doroshov, S.I. Evidence of Hexaploid Karyotype in Shortnose Sturgeon. *Genome* **2008**, *51*, 113–119. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kim, D.S.; Nam, Y.K.; Noh, J.K.; Park, C.H.; Chapman, F.A. Karyotype of North American Shortnose Sturgeon *Acipenser brevirostrum* with the Highest Chromosome Number in the Acipenseriformes. *Ichthyol. Res.* **2005**, *52*, 94–97. [\[CrossRef\]](#)
- Fontana, F.; Zane, L.; Pepe, A.; Congiu, L. Polyploidy in Acipenseriformes: Cytogenetic and Molecular Approaches. In *Fish Cytogenetics*; Pisano, E., Ozouf-Costaz, C., Foresti, F., Kapoor, B.G., Eds.; Science Publisher, Inc.: Enfield, NH, USA, 2007; pp. 385–403.
- 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. [\[CrossRef\]](#)
- 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. Ichth.* **2015**, *31*, 977–983. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Havelka, M.; Hulák, M.; Bailie, D.A.; Prodöhl, P.A.; Flajšhans, M. Extensive Genome Duplications in Sturgeons: New Evidence from Microsatellite Data. *J. Appl. Ichthyol.* **2013**, *29*, 704–708. [\[CrossRef\]](#)
- Stift, M.; Berenos, C.; Kuperus, P.; van Tienderen, P.H. 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. [\[CrossRef\]](#)
- Soares, N.R.; Mollinari, M.; Oliveira, G.K.; Pereira, G.S.; Vieira, M.L.C. Meiosis in Polyploids and Implications for Genetic Mapping: A Review. *Genes* **2021**, *12*, 1517. [\[CrossRef\]](#)
- Boscari, E.; Barbisan, F.; Congiu, L. Inheritance pattern of Microsatellite Loci in the Polyploid Adriatic Sturgeon (*Acipenser naccarii*). *Aquaculture* **2011**, *321*, 223–229. [\[CrossRef\]](#)
- Congiu, L.; Pujolar, J.M.; Forlani, A.; Cenadelli, S.; Dupanloup, I.; Barbisan, F.; Galli, A.; Fontana, F. Managing Polyploidy in Ex Situ Conservation Genetics: The Case of the Critically Endangered Adriatic Sturgeon (*Acipenser naccarii*). *PLoS ONE* **2011**, *6*, e18249. [\[CrossRef\]](#) [\[PubMed\]](#)
- Boscari, E.; Pujolar, J.M.; Dupanloup, I.; Corradin, R.; Congiu, L. Captive Breeding Programs Based on Family Groups in Polyploid Sturgeons. *PLoS ONE* **2014**, *9*, e110951. [\[CrossRef\]](#) [\[PubMed\]](#)
- May, B.; Krueger, C.C.; Kincaid, H.L. Genetic Variation at Microsatellite Loci in Sturgeon: Primer Sequence Homology in *Acipenser* and *Scaphirhynchus*. *Can. J. Fish. Aquat. Sci.* **1997**, *54*, 1542–1547. [\[CrossRef\]](#)
- Welsh, A.B.; Blumberg, M.; May, B. Identification of Microsatellite Loci in Lake Sturgeon, *Acipenser fulvescens*, And Their Variability in Green Sturgeon, *A. medirostris*. *Mol. Ecol. Notes* **2002**, *3*, 47–55. [\[CrossRef\]](#)
- Zane, L.; Patarnello, T.; Ludwig, A.; Fontana, F.; Congiu, L. Isolation and Characterization of Microsatellites in the Adriatic Sturgeon (*Acipenser naccarii*). *Mol. Ecol. Notes* **2002**, *2*, 586–588. [\[CrossRef\]](#)

18. Forlani, A.; Fontana, F.; Congiu, L. Isolation of Microsatellite Loci from the Endemic and Endangered Adriatic Sturgeon (*Acipenser naccarii*). *Conserv. Genet.* **2007**, *9*, 461–463. [[CrossRef](#)]
19. Henderson-Arzapalo, A.; King, T.L. Novel Microsatellite Markers for Atlantic Sturgeon (*Acipenser oxyrinchus*) Population De-Lineation and Broodstock Management. *Mol. Ecol. Notes* **2002**, *2*, 437–439. [[CrossRef](#)]
20. McQuown, E.C.; Sloss, B.L.; Sheehan, R.J.; Rodzen, J.; Tranah, G.J.; May, B. Microsatellite Analysis of Genetic Variation in Sturgeon: New Primer Sequences for Scaphirhynchus and Acipenser. *Trans. Am. Fish. Soc.* **2000**, *129*, 1380–1388. [[CrossRef](#)]
21. Sybenga, J. Chromosome Pairing Affinity and Quadrivalent Formation in Polyploids: Do Segmental Allopolyploids Exist? *Genome* **1996**, *39*, 1176–1184. [[CrossRef](#)]
22. Lloyd, A.; Bomblies, K. Meiosis in Autopolyploid and Allopolyploid Arabidopsis. *Curr. Opin. Plant Biol.* **2016**, *30*, 116–122. [[CrossRef](#)] [[PubMed](#)]
23. Choudhary, A.; Wright, L.; Ponce, O.; Chen, J.; Prashar, A.; Sanchez-Moran, E.; Luo, Z.; Compton, L. Varietal Variation and Chromosome Behaviour During Meiosis in Solanum Tuberosum. *Heredity* **2020**, *125*, 212–226. [[CrossRef](#)] [[PubMed](#)]
24. Rice, W.R. Analyzing tables of statistical tests. *Evolution* **1989**, *43*, 223–225. [[CrossRef](#)] [[PubMed](#)]
25. Scott, A.D.; Van de Velde, J.D.; Novikova, P.Y. Inference of Polyploidy Origin and Inheritance Mode from Population Genomic Data. *BioRxiv*. [[CrossRef](#)]
26. Pyatskowitz, J.D.; Krueger, C.C.; Kincaid, H.L.; May, B. Inheritance of Microsatellite Loci in the Polyploid Lake Sturgeon (*Acipenser fulvescens*). *Genome* **2001**, *44*, 185–191. [[CrossRef](#)]
27. Rodzen, J.A.; May, B. Inheritance of Microsatellite Loci in the White Sturgeon (*Acipenser transmontanus*). *Genome* **2002**, *45*, 1064–1076. [[CrossRef](#)]
28. Schreier, A.D.; Van Eenennaam, J.P.; Anders, P.; Young, S.; Crossman, J. Spontaneous Autopolyploidy in the Acipenseriformes, with Recommendations for Management. *Rev. Fish Biol. Fish.* **2021**, *31*, 159–180. [[CrossRef](#)]
29. Birstein, V.J. Sturgeon Species and Hybrids: Can Hybrids Produce Caviar. *Environ. Policy Law* **2002**, *32*, 210–214.

ANNEX B

Article

Characterization of Captive Breeders to Preserve the Residual Genetic Diversity of Adriatic Sturgeon (*Acipenser naccarii*)

Federica Barca ^{1,†}, Stefano Dalle Palle ^{1,†} , Luca Schiavon ¹, Chiara Samassa ¹, Giuseppe Castaldelli ² ,
Elisa Boscari ^{1,*}  and Leonardo Congiu ^{1,3} 

¹ Department of Biology, University of Padova, Via Ugo Bassi 58b, 35121 Padova, Italy

² Department of Environmental and Prevention Sciences, University of Ferrara, Via L. Borsari 46, 44121 Ferrara, Italy

³ National Biodiversity Future Center (NBFC), 90133 Palermo, Italy

* Correspondence: elisa.boscari@unipd.it

† These authors contributed equally to this work.

Abstract: Since 1996, the Adriatic sturgeon (*Acipenser naccarii*) has been inscribed on the IUCN Red List as “Critically Endangered and possibly extinct in the wild”. Nowadays, its survival totally depends on restocking programs conducted by releasing juveniles generated from adult breeders reared in aquaculture. Conducting accurate genetic characterizations of all individuals potentially involved in reproduction activities is therefore of primary importance to avoid inbreeding and to maximize the genetic diversity transmitted to following generations. Since all animals reared in captivity descend from a single stock of wild origin, this offers the ideal condition for carrying out relatedness analysis based on parentage allocations. In this study, we provided the most complete characterization of about 500 individuals representing the most diverse extant stock of Adriatic sturgeon. Through the analyses of mitochondrial d-loop and 15 microsatellite loci selected from 24 genotyped loci, we identified about 30 different familiar groups, updating data on breeding stocks, increasing the genetic information already available, and extending the analyses to animals never genotyped before. Given its completeness, it will represent a reference database for any future parental allocation of recaptured animals for the inclusion of all other stocks present, as well as for the development of a long-term breeding plan. The approach used has also been proven useful on individuals of unknown genealogy, allowing for the identification of family groups and thus being proven to be promising for the analysis of stocks of other tetraploid sturgeon species.

Keywords: acipenseridae; aquaculture; biodiversity; conservation genetics; critically endangered; microsatellites; parental allocation; pedigree; relatedness; stock assessment



Citation: Barca, F.; Dalle Palle, S.; Schiavon, L.; Samassa, C.; Castaldelli, G.; Boscari, E.; Congiu, L. Characterization of Captive Breeders to Preserve the Residual Genetic Diversity of Adriatic Sturgeon (*Acipenser naccarii*). *Diversity* **2022**, *14*, 829. <https://doi.org/10.3390/d14100829>

Academic Editor: Michael Wink

Received: 13 September 2022

Accepted: 29 September 2022

Published: 2 October 2022

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



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The latest IUCN (International Union for Conservation of Nature) sturgeon assessment, released on 21 July 2022, showed a worsening of the group of species, which was already at greatest risk of extinction in the world since 2010. Among these, the Adriatic sturgeon (*Acipenser naccarii*, Bonaparte 1836), endemic to the North Adriatic region, showed an opposite trend if compared to other sturgeon species. Indeed, it has had an improvement in its reclassification from “Critically Endangered and possibly extinct in the wild” to simply “Critically Endangered”. Its extinction has so far been avoided also thanks to captive breeding programs started in 1977, just before the species almost disappeared in nature, by transferring about 90 immature wild individuals (hereafter F0) from the Po River to a private aquaculture plant, the Azienda Agricola V.I.P. (Orzinuovi, Brescia, Italy). Fifty of these animals survived and reached sexual maturity after about 10 years. In 1988, the first successful reproduction of the species in captivity was obtained [1,2], paving the way for a constant propagation effort that allowed generating more than 30 F1 stocks in more than 30 years by crossing different F0 breeders [3]. Juveniles obtained were mostly

used for reintroduction purposes, and over 500,000 animals were released under several independent and uncoordinated restocking programs [4]. Since then, occasional captures have been reported, but up to now, very few and indirect pieces of evidence of natural reproduction are available [5], and the persistence of the species in the wild still depends on reintroduction programs and on the proper management of residual genetic diversity.

After more than 40 years from its establishment, the F0 stock is now reduced to a few individuals, which represent the only living animals of certain wild origin. Now, mature F1s are being used for reproductions. In fact, part of the progenies from each F0 cross performed over the years has been retained in captivity as future breeders by different public authorities as well as by the Azienda Agricola V.I.P. itself. As expected, this stock resulted in a remarkably higher genetic diversity compared with all other stocks established by only using a very small number of sibling groups [3,6,7]. In any case, all animals bred in captivity in Italy and abroad descend directly from the F0 stock, potentially easing the genetic analysis required to reconstruct their pedigree but, nevertheless, the precise genealogy of each individual and the consequent degree of kinship is still unknown for most stocks. Therefore, a careful survey of all Adriatic sturgeons held in captivity in the different plants and an exhaustive assessment of the number and size of existing family groups are a priority to inform optimal strategies for the management of a living bank of Adriatic sturgeon.

In fact, the future of this species now relies on the correct management of F1 stocks with special attention to maximizing the level of genetic variation transmitted to future generations [3], thus maintaining the highest possible level of long-term adaptive potential. Recently, the Pan European Action Plan [8], approved by the Berne Convention and included in the Habitats Directive of the European Community, highlighted the importance of a coordinated effort to optimize conservation strategies, including the management of genetic heritage. Fortunately, the owners of the above-mentioned most biodiverse stock of this species, previously reared by the V.I.P. but currently acquired by the company “Sturgeon Ticino”, agreed to carry out a complete genotyping of their stocks, hopefully setting the basis for a complete genetic characterization of all *A. naccarii* individuals present in captivity. Most of these stocks are in Italy, but a few of them are in other European countries such as Spain [9]. The present work, by proposing a standardized approach to the analyses of genetic diversity in polyploid sturgeons, reports the results of this characterization aimed at identifying the number of family groups and their size, providing the information for a long-term crossbreeding plan.

2. Materials and Methods

2.1. Sampling and Molecular Markers Used

In the present work, the genetic diversity of a total of 42 individuals of wild origin (F0; 20 males and 22 females) and 505 individuals of first-generation (F1) of *A. naccarii* were analyzed through mitochondrial and nuclear markers.

For the historical 42 F0 and for 418 F1 individuals born along the last 20 years in captivity and sampled in 2011, the purification of genomic DNAs and a preliminary parental allocation, which was based on sequencing of the mitochondrial control region and genotyping of seven microsatellite loci (Table S1), were already performed [3]. Here, the genotype information at an additional 17 microsatellite loci (Table S1) was obtained, eight of which (AnacB11, AnacB7 [10], Spl120 [11], AfuG132, AfuG112 [12], Anac_c15214 [13], AoxD161, and AoxD241 [14]; Table S1) yielded suitable information to be used in parental allocation. These were therefore added to the previously used seven loci to provide updated genealogy reconstructions based on 15 loci. A significant fraction of these 418 animals died in an unfortunate poisoning event some years ago, but we have decided to include them in the analyses anyway because the hundreds of thousands of animals released as juveniles in the past came from the same crosses, and a good representation of the diversity of all family groups generated in the last decades can facilitate genetic tagging of the recaptures.

For the remaining 87 individuals born after 2011 that were never characterized before, all 24 microsatellite loci and mitochondrial d-Loop were analyzed (Table S1).

Genomic DNA was purified from fin clips with the EuroGOLD Tissue Mini Kit (EuroClone, Milan, Italy) and stored at -4°C before the amplification. The mitochondrial control region was amplified using the Pro1F-Phe1R primer pair following the conditions described in the original reference [6]. After enzymatic purification with ExoSAP-ITTM (Usb), sequencing reactions were performed availing the Applied Biosystem Automatic Sequencer at the Eurofins Genomics sequencing service (Ebersberg, Germany, <https://eurofinsgenomics.eu/>, accessed on 13 September 2022), and chromatograms were analyzed with MEGAX software [15]. For microsatellite analyses, the seven loci previously used were amplified in single PCR reactions and then assembled for genotyping (Multiplex M0 in Table S1) following the same protocols used during past genetic characterization of the same stock [3]. The panel of 17 new loci was amplified in two multiplex reactions, M1 and M2 (Table S1), using the Multiplex PCR Master Mix (QIAGEN); the thermal profile was 15' at 95°C , 35 cycles of 30'' at 94°C , 90'' at 60°C , 60'' at 72°C , and 30' at 60°C . Locus Afug113 was amplified in single touchdown PCR reaction and then assembled in M1; loci AoxD241, Afug112, and AnacB7 were also individually amplified and assembled to M2. Further information about loci used (e.g., primer sequences, size ranges, multiplexing, locus-specific thermal profiles, original references) are detailed in Table S1.

PCR products were genotyped at the BMR Genomics (external service, Padova, Italy, <https://www.bmr-genomics.it/>, accessed on 13 September 2022), and allele scoring was performed using GENEMARKER software version 1.95 (SoftGenetics LLS, State College, PA, USA).

2.2. Parental Allocation Analysis

For a total of 505 individuals successfully genotyped, parental allocation was carried out to reconstruct genealogy and identify full sibs or half sibs. The final microsatellite panel included 15 loci selected among the 24 genotyped. Besides the seven old loci already used in the past and also included in the hereafter analyses, nine loci of the new panel of 17 were finally not included (Table S1) in the parental allocation for different reasons: (i) the loci 2589, Anac_c14336, and Spl168 showed a number of alleles greater than four in many individuals, suggesting duplication of the analyzed region or unreduced ploidy, and in any case were not suitable for one of the parental allocation procedures used; (ii) loci Anac_c3133, Afug113, Afug41, and Spl-163 showed discordances between the genotypes obtained in multiplex and in single amplification; and (iii) the loci Anac_c12159 and Anac_c6784 presented an electropherogram difficult to interpret in part of the animals and were prone to bias in allele scoring. The recovery of these discarded loci would have required further fine-tuning and single-locus amplification on many animals, but given the good informative power of the retained loci, it was considered not necessary.

Given the tetraploid genome of the Adriatic sturgeon, the inference of the real genotype at a certain locus is not always possible, and the allelic frequencies can only be estimated; therefore, parental allocation analyses were implemented by using the software *BreedingSturgeons*, specifically developed for tetraploid sturgeon species [3]. Two different algorithms are followed: the standard band-sharing measure and a weighted estimation of segregating alleles from different possible parent pairs and putative offspring. The first and simpler one compares the alleles observed in the multi-locus profile of each F1 individuals with the ones of all possible pairwise combinations of breeders (one male and one female). If all the F1 alleles are present in a breeders' pair, the parental compatibility is accepted, or otherwise is rejected. This first method, however, does not check for the satisfaction of Mendelian segregation, which is on the contrary assessed by the second approach in which for the three individuals under comparison (the two possible breeders and the F1 to be allocated), all possible tetraploid genotypes are inferred from the allele profiles at each locus and the inheritance of two alleles from each parent at all loci is verified. Breeders' pairs that at each locus equally contribute to the generation of the F1 profile are

considered totally compatible with an allocation score of 1. This second approach may also accept the possibility of scoring errors or segregation anomalies tolerating up to two loci with imperfect inheritance. On the basis of the number of anomalies at these two loci, an allocation score is accordingly calculated and accepted above a threshold of 0.866 [3]. Accepting discrepancies at a small fraction of loci, the software indirectly accepts allocations of those F1 that show (at those loci) alleles absent in the profile of the parent pair (a scenario which is not tolerated by the first method). However, in these cases the compatibility score is decreased. The two methods were both used, and the results were compared.

Finally, the mitochondrial haplotype of the identified mother was compared with the mitochondrial haplotype of the associated offspring to verify its concordance.

2.3. Analyses and Visualization of Pairwise Genetic Distances

Brotherhood among F1 animals can be inferred through parental allocation to the same parent pairs when parents are available; on the contrary, when they are not, alternative methods are needed. With the aim of testing an alternative empirical approach, we performed an estimation of pairwise genetic distances among all F1 animals, assuming that the more related are two genomes the lower is their genetic distance. Due to the tetraploid condition that precludes the possibility of assessing the exact allele dosage, we used the presence/absence information of the different alleles to estimate Sørensen's distances [16] with the R package "Ade4" [17]. Pairwise distances were then visualized through multi-dimensional scaling performed with R package "ggplot2" [18]. Groups composed of three or more individuals of known relatedness were analyzed to validate the reliability of the approach, before applying it on animals with unknown genealogies.

3. Results

3.1. Pedigree Reconstruction

All 505 animals were successfully analyzed at 15 loci, and 357 of which, 216 still alive, were unambiguously allocated to their parental pair, identifying a total of 29 family groups (Tables 1 and S2).

Table 1. Résumé of family groups identified with parental allocation analysis based on 15 microsatellite loci. For each family group, the parent pair, the total number of allocated animals, and the number of survived ones (in brackets) are reported as well as the maternal mitochondrial haplotype. The families 08 and 11 are composed of F1 animals carrying the mitochondrial haplotypes 5 or 7, and in some cases, animals showed heteroplasmy with both the mitochondrial variants. This was due to the heteroplasmic condition of the shared F0 female. Further details for each individual analyzed are reported in Table S2.

Family Group	Parental Pair	N F1 Tot (Alive)	mtDNA Haplotype	Family Group	Parental Pair	N F1 Tot (Alive)	mtDNA Haplotype
01	Pelvienne (F) × NaccS18 (M)	14 (5)	2	17	NaccS8 (F) × 740 (M)	21 (14)	2
02	NaccS8 (F) × Matto (M)	9 (6)	2	18	NaccS3 (F) × Matto (M)	5 (1)	3
03	NaccS7 (F) × NaccS6 (M)	12 (12)	3	19	NaccS33 (F) × 740 (M)	4 (2)	3
04	NaccS8 (F) × NaccS31 (M)	17 (15)	2	20	NaccS7 (F) × VerdeBis (M)	2 (1)	3
05	NaccS33 (F) × NaccS11 (M)	10 (4)	3	21	NaccS12 (F) × NaccS17 (M)	1 (1)	2
06	O2 (F) × Matto (M)	12 (9)	5	22	Indecisa (U) × Matto (M)	2 (2)	2
07	NaccS26 (F) × NaccS29 (M)	27 (18)	5	23	Indecisa (U) × AnS11 (M)	1 (1)	2
08	NaccS19 (F) × NaccS17 (M)	18 (17)	5/7	24	NaccS12 (F) × Matto (M)	1 (0)	2
09	NaccS28 (F) × NaccS17 (M)	51 (30)	3	25	Naccs3 (F)–Naccs29 (M)	1 (0)	3
10	NaccS16 (F) × NaccS23 (M)	39 (15)	3	26	NaccS16 (F) × 740 (M)	1 (1)	3
11	NaccS19 (F) × NaccS31 (M)	30 (21)	5/7	27	NaccS16 (F) × NaccS29 (M)	1 (1)	3
12	NaccS16 (F) × NaccS30 (M)	13 (2)	3	28	Naccs8 (F)–Naccs11 (M)	6 (6)	2
13	NaccS33 (F) × NaccS9 (M)	17 (12)	3	29	Pelvienne (F)–Naccs23 (M)	3 (3)	2
14	NaccS8 (F) × NaccS17 (M)	8 (8)	2	30	Non-allocated	141 (108)	-
15	NaccS12 (F) × NaccS27 (M)	10 (3)	2	31	Multi-allocated	7 (7)	-
16	NaccS3 (F) × 740 (M)	21 (6)	3				

Of these, 26 confirm the previous allocation, even increasing the number of loci from 7 [3] to 15, further supporting past results. The only exceptions are five animals of the familiar groups 09, 18, and 25, which, compared to their past allocation, are now assigned to the same dam but to a different sire (Table S2). In these cases, the compatibility index of the previously selected parent pair decreased below the threshold due to some incompatibilities in the newly genotyped loci, and the compatibility index of the new parent pair increased as alleles at all newly genotyped loci resulted in being compatible. This is somewhat expected in re-analyzing relatedness and reconstructing pedigree information by adding loci.

The remnant two family groups (28 and 29, in Table 1) were never identified previously, even though the respective F0 individuals involved were already represented in other familiar groups. Unfortunately, the size of families, already small in some cases, was further reduced after poisoning, resulting in two family groups completely lost and 11 composed of three or less individuals, at least in the Storione Ticino Stock.

Seven individuals remained multi-allocated (compatible with more than one parent pair), indicating that, although the allocation power is expected to increase by adding eight new loci, in some cases it may still not be sufficient.

Finally, 141 *A. naccarii* were not allocated, probably because progeny of one or both parents died before our first sampling occurred in 2004.

The contribution of the different breeders to the F1 stock was strongly unbalanced, with only four breeders contributing to than 40% of the still alive F1, and some other breeders with no progeny at all or very few F1 in this stock (Figure 1). Out of the 26 wild breeders (14 males out of 20 and 12 females out of 22) that were successfully reproduced in the past and for which at least one F1 is still available, 9 and 15 were, respectively, the parents of about 50% and 70% of the stock. Approximately the same percentages were valid for both the total and the alive stocks.

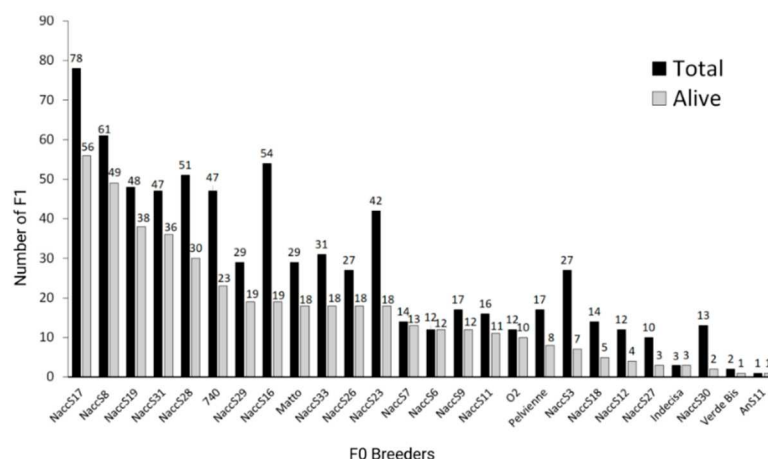


Figure 1. Contribution of the different F0 breeders to the F1 generation. For each breeder reproduced in the past, the total number of F1 in the analyzed stock and the ones still alive are reported (created with Biorender.com, accessed on 13 September 2022).

3.2. Multi-Dimensional Scaling and Relatedness Inference

MDS analysis (Figure 2) was performed to visualize pairwise genetic distances among individuals belonging to F1 families with at least three animals. For a simple matter of clarity, individuals were separately analyzed according to their mitochondrial haplotypes, and the results were reported in three different panels (Figure 2a–c). The graph with the simultaneous analysis of all individuals is reported in Figure S1 of the Supplementary Materials. Individuals with haplotypes 5 and 7 were grouped, as haplotype 7 was sporadically observed mainly in heteroplasmy with haplotype 5.

In general, the MDS graphical representations were concordant with the results of parental allocations, and animals allocated to the same parental pair formed clear groups of similarity. Families sharing one parent (half-sibs) also tended to cluster together, such as families 02, 14, and 28 (Figure 2a); families 10 and 12 (Figure 2b); families 05, 13, and 19 (Figure 2b); and families 08 and 11 (Figure 2c). However, when a second run of the same analysis was performed, considering separately each of these cluster of half-sibs, the underlying families became better distinguishable, as in the example reported in Figure 2d in which families 10 and 12, overlapping when analyzed with all individuals with haplotype 3 (Figure 2b), were clearly separated; moreover, the two fathers and the common mother of these families took a coherent position in the graph.

The results of the same approach used on animals of unknown genealogy, after having grouped them according to the mitochondrial haplotype, are shown in Figure 2e,f. Although it is not possible to verify the results using paternity analyzes for these individuals where the parents are unknown, it is evident that there are groups of similarities, probably consisting of groups of siblings or half-siblings.

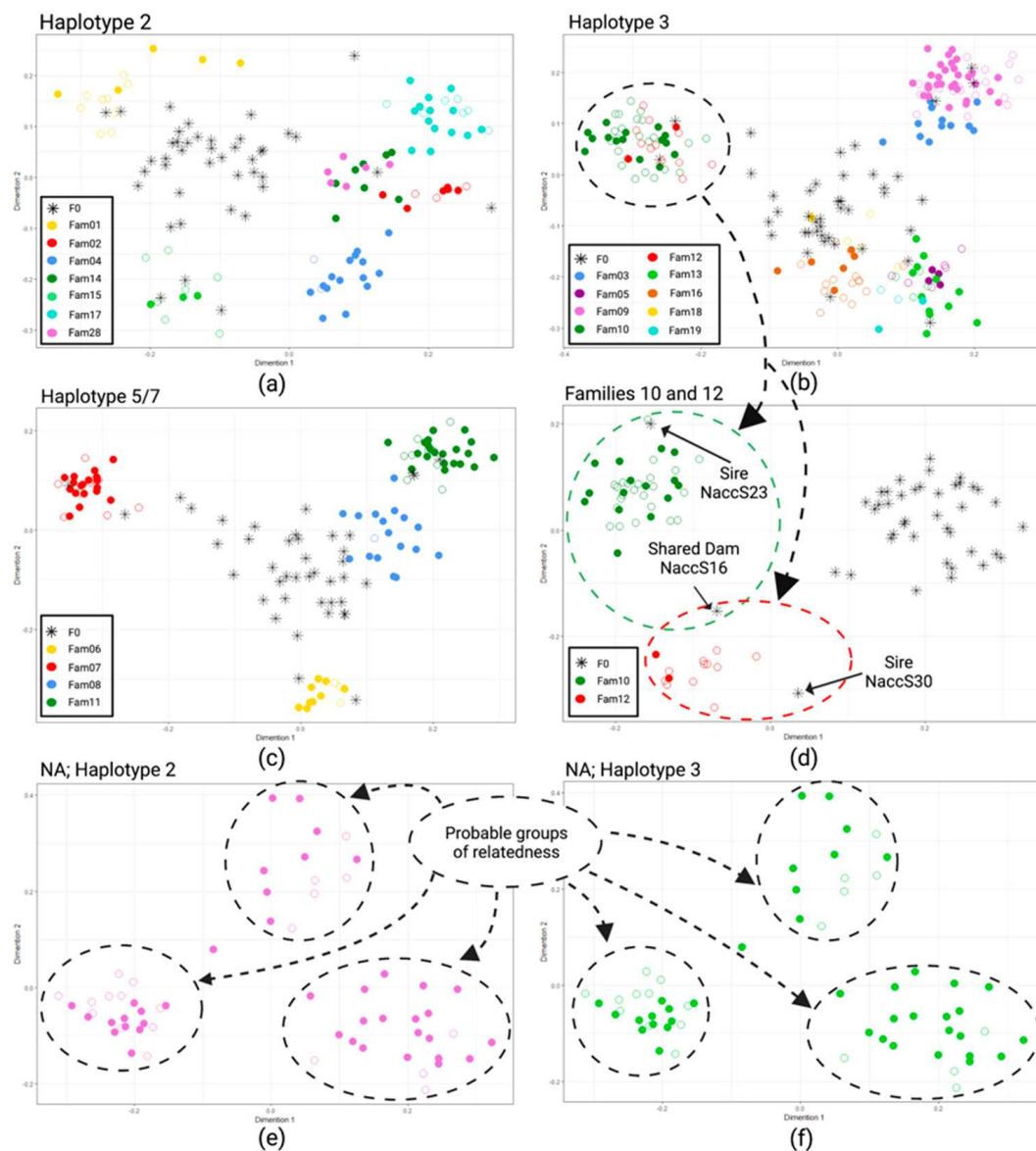


Figure 2. Results of MDS analyses performed on pairwise Sørensen's distance matrix among F0 and F1 with mitochondrial haplotype 2 (a); F0 and F1 with mitochondrial haplotype 3 (b) and F0 and mitochondrial haplotypes 5 and 7 (c). Individuals of two half-sib family groups (10 and 12), that in the first run were not distinguishable, were included in a second run, excluding all other F1s to show the usefulness of a hierarchical approach in also distinguishing related families; the common dam and the sires of the two families are also highlighted (d). The same approach was used on animals for which the parental information was not available, having mitochondrial haplotype 2 and 3 (e,f), respectively. In all figures, full and empty circles represent alive and dead animals, respectively (created with Biorender.com, accessed on 13 September 2022). NA = non-allocated.

4. Discussion

This genetic characterization allowed for the deepening of the research already conducted in the past with a lower number of loci and extended it to all mature individuals reared at the Storione Ticino farm that had never been characterized before. To our knowledge, this is the most complete genetic characterization performed so far on a captive stock of a sturgeon species, with the aim of planning an informed production of juveniles to be released into the wild.

The stock of the Storione Ticino farm confirmed the presence of a significant degree of genetic diversity when compared to other F1 *A. naccarii* stocks [7]. Even if 29 families

were apparently low on the basis of the theoretically maximum possible number of families obtainable from 20 F0 males and 22 F0 females, the fact that 26 out of 42 adults were reproduced in the past testifies a careful commitment in differentiating the breeders used, in an attempt to maximize residual genetic diversity. It should also be considered that some of the animals never became available for reproduction. Moreover, given that for 141 F1 individuals, no pairs of F0 parents were found to be compatible, we can reasonably assume that additional breeders were used in the past to generate these families and died before our sample collection, leaving their genetic heritage in their progenies and making them bearers of added value in terms of biodiversity. Unfortunately, some of the family groups were composed of very few individuals, and in some cases by a single animal. This represents a serious problem, firstly because one single individual is not representative of the genetic diversity of their parents, and second because not all animals become available for reproduction, especially females that in this species ovulate every two, three, or sometimes more years [19], and the probability of permanently losing these genetic lines is relevant [3]. This highlights the relevance of saving an adequate number of animals from each crossing and the importance of distributing the animals of the different genealogical lines in different plants to decrease the probability of disappearance due to accidental events [8]. As already mentioned, however, there are other stocks of *A. naccarii* [9], all descendants of the F0 stock analyzed in this study. There is, therefore, hope that representatives of some under- or non-represented families are present elsewhere. In any case, individuals belonging to these residual families should be considered a priority in planning future crossings.

The information provided here can be used to select the breeders to pair as well as to perform genetic allocations to parents of origin. For this reason, we decided to also perform the characterization on the 170 animals already dead due to an accidental poisoning [3] as they might have been used to generate fingerlings released in the wild. Moreover, since part of the animals of the analyzed broodstock were used to produce commercial caviar, the method proposed could also be suitable to certify the origin of products in trade.

The fact that full- and half-sibs are often not completely distinguishable is somewhat expected as the average distance within a family depends on the similarity of the two parents, which can significantly vary across families. This implies that, if the method is applied to identify groups of relatedness in stock for which parents are not available, distinguishing full- from half-sibs may be difficult. However, in these cases, we have shown that a second MDS run, performed only on individuals identified as somehow related in the first run, can provide a finer resolution. In any case, whatever the degree of kinship, their genetic proximity suggests not to use individuals of the same group as reproductive partners. In the present paper, the MDS representation was applied to animals with the same haplotypes, allowing for the detection of full sibs and maternal half-sibs (sharing the same mother). In doing this, however, paternal half-sibs (sharing the same father) would not be detected if the two mothers had different mitochondrial haplotypes. These kinds of full sibs could be identified only if all animals were simultaneously analyzed. This was not our case, as the MDS analyses of all animals did not allow for the identification of multi-haplotype clusters (Figure S1). This limit was possibly due to the high number of families present in the stock and the consequent homogeneous distribution of pairwise Sørensen's distances. In any case, we suggest a preliminary overall analysis that, in the case of stocks descending from a low number of parents, could also allow for the detection of paternal half-sibs.

The availability of the F0 generation, besides facilitating the analyses of relatedness, allowed for validation of the alternative approach on the basis of the presence/absence of the different alleles, which is especially useful in the case of tetraploid species in which it is not possible to infer the exact genotype [6].

Having at our disposal the genotypes of most of the wild founders, the analysis of the relatedness among F1 was much easier than for stocks whose origin is unknown; however, the reliability of the method used here in identifying the groups of relatedness, even in the

case of the not-allocated animals, makes it particularly useful for the analysis of all sturgeon stocks whose composition is unknown. In particular, the approach adopted here is useful for sturgeon species with a high degree of ploidy, for which standard approaches based on the knowledge of the genotype cannot be used. In fact, in polyploid genomes, even markers that in diploids are considered co-dominant, such as microsatellites or single-nucleotide polymorphisms (SNPs), do not allow for a precise inference of the genotype, forcing the use of phenotypic information of the presence/absence of the different alleles. This approach could be very useful for all polyploid species on the brink of extinction in nature but for which significant captive stocks are available, and thus a careful assessment of purity and relatedness is required in order to inform restocking measures. For example, this was the case for the Dabry sturgeon (*Acipenser dabryanus*, Duméril, 1869), recently assessed by IUCN (2022) as extinct in the wild, or of the Chinese sturgeon (*Acipenser sinensis*, Gray, 1835), for which very few wild breeders survive [20].

The different genetic characterizations carried out on this species have shown that the genome of *A. naccarii*, as well as probably of all the other sturgeon species, has a variable level of ploidy. In fact, in addition to the loci here analyzed, which were selected for the presence of a maximum of four alleles for individuals (indicating a probable tetraploidy), loci with a higher number of alleles were observed [13]. This can be explained by the presence of duplicate regions or as a trace of an ancestral genomic duplication [21]. Recently, it was also observed that in *A. naccarii*, some loci follow a disomic segregation modality [22], indicative of a further differentiation of the chromosomes towards a diploid condition [23]. In summary, the coexistence of different ploidy levels in different regions of the genome makes the development of analytical methods for sturgeon species belonging to the group with 240 chromosomes challenging [21]. Relying on “phenotypic” data, therefore, remains the preferred approach for these of animals, as recently demonstrated also on the Chinese sturgeon (*A. sinensis*) [20].

5. Conclusions

Genetic erosion is inevitable in ex situ maintenance of a limited number of animals, but it could be mitigated by including animals coming from other stocks and planning crosses on the basis of accurate pedigree reconstructions [8]. In Europe, mainly in Italy and Spain [9], several stocks of *A. naccarii* exist, both public and private, which, even if all descending from the original F0 stock and generally made up of a small number of family groups, could have accidentally kept rare genetic lines, which should be reproduced with priority. The very recent publication of the new assessment of sturgeons and paddlefish by the IUCN denounces a serious deterioration in the conservation status of most of the populations and indicates the urgency of establishing genetically certified ex situ stocks. Thus, we claim the urgency of characterizing all existing captive stocks of *A. naccarii*, considering them as a unique meta-stock, reared in different plants, on which a breeding program is designed that contemplates the informed exchange of individuals or gametes between the different farms. The approach proposed does not require any preliminary information and can be applied to all the various polyploid sturgeon species for which conservation relies on the maintenance and reproduction of ex situ stocks.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/d14100829/s1>, Figure S1: MDS performed on pairwise Sørensen’s distance matrix among all 505 F1 animals colored according to their mitochondrial haplotype. Table S1: Details about all microsatellite loci analyzed in the present study [24,25]. Table S2: Pedigree information for each individual analyzed in the present study.

Author Contributions: Conceptualization, L.C. and E.B.; methodology, L.C. and E.B.; validation, F.B., S.D.P. and E.B.; formal analysis, F.B. and S.D.P.; investigation, F.B., S.D.P., C.S. and E.B.; software, L.S.; writing—original draft preparation, F.B.; writing—review and editing, E.B. and L.C.; supervision, E.B., G.C. and L.C.; project administration, L.C.; funding acquisition, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Biodiversity Future Centre (NBFC funded by the Italian Ministry of University and Research, PNRR, Missione 4 Componente 2, “Dalla ricerca all’impresa”, Investimento 1.4, Project CN00000033).

Institutional Review Board Statement: The animal study protocol was approved by the Institutional Ethics Committee of the University of Padova (OPBA) (protocol code 421922 of 14 October 2020).

Data Availability Statement: Not applicable.

Acknowledgments: We thank the Storione Ticino farm for providing samples, as well as the Giovannini family for their 50-year-long engagement in the conservation of the Adriatic sturgeon and for having conducted their activity under constant interaction with researchers in different disciplines.

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

References

1. Arlati, G.; Bronzi, P.; Colombo, L.; Giovannini, G. Induzione della riproduzione nello storione italiano (*Acipenser naccarii*) allevato in cattività. *Riv. Ital. Acquacol.* **1988**, *23*, 94–96.
2. Giovannini, G.; Colombo, L.; Bronzi, P.; Arlati, G. Growth of hatchery-produced juveniles of Italian sturgeon, *Acipenser naccarii* Bonaparte, reared intensively in fresh water. In *Acipenser*; Cemagref Publication: Bordeaux, France, 1991; pp. 401–404.
3. Boscari, E.; Pujolar, J.M.; Dupanloup, I.; Corradin, R.; Congiu, L. Captive breeding programs based on family groups in polyploid sturgeons. *PLoS ONE* **2014**, *9*, e110951. [\[CrossRef\]](#)
4. Arlati, G.; Poliakova, L. Restoration of Adriatic Sturgeon (*Acipenser naccarii*) in Italy: Situation and Perspectives. In *Biology, Conservation and Sustainable Development of Sturgeons*; Carmona, R., García-Gallego, M., Ruiz-Rejón, M., Eds.; Springer: Dordrecht, The Netherlands, 2009; Volume 14, pp. 237–238.
5. Congiu, L.; Boscari, E.; Pagani, S.; Gazzola, M.; Bronzi, P. Resumption of natural reproduction of the Adriatic sturgeon in the River Po. *Oryx* **2021**, *55*, 816. [\[CrossRef\]](#)
6. Congiu, L.; Pujolar, J.M.; Forlani, A.; Cenadelli, S.; Dupanloup, I.; Barbisan, F.; Galli, A.; Fontana, F. Managing polyploidy in ex situ conservation genetics: The case of the critically endangered Adriatic sturgeon (*Acipenser naccarii*). *PLoS ONE* **2011**, *6*, e18249. [\[CrossRef\]](#)
7. Boscari, E.; Congiu, L. The need for genetic support in restocking activities and ex situ conservation programmes: The case of the Adriatic sturgeon (*Acipenser naccarii* Bonaparte, 1836) in the Ticino River Park. *J. Appl. Ichthyol.* **2014**, *30*, 1416–1422. [\[CrossRef\]](#)
8. World Sturgeon Conservation Society; WWF. Pan-European Action Plan for Sturgeons. In Proceedings of the Convention on the Conservation of European Wildlife and Natural Habitats Standing Committee: 38th Meeting, Strasbourg, France, 27–30 November 2018.
9. Cabrera-Castro, R.; Zabal, C.; Soriguer, M.C.; Domezain, A.; Hernando, J.A. Morphological development in the first phase of Adriatic sturgeon *Acipenser naccarii* under controlled conditions. *J. Fish Biol.* **2018**, *92*, 1956–1974. [\[CrossRef\]](#)
10. Forlani, A.; Fontana, F.; Congiu, L. Isolation of microsatellite loci from the endemic and endangered Adriatic sturgeon (*Acipenser naccarii*). *Conserv. Genet.* **2007**, *9*, 461–463. [\[CrossRef\]](#)
11. McQuown, E.C.; Sloss, B.L.; Sheehan, R.J.; Rodzen, J.; Tranah, G.J.; May, B. Microsatellite analysis of genetic variation in sturgeon: New primer sequences for *Scaphirhynchus* and *Acipenser*. *Trans. Am. Fish. Soc.* **2000**, *129*, 1380–1388. [\[CrossRef\]](#)
12. Welsh, A.B.; Blumberg, M.; May, B. Identification of microsatellite loci in lake sturgeon, *Acipenser fulvescens*, and their variability in green sturgeon, *A. medirostris*. *Mol. Ecol. Notes* **2003**, *3*, 47–55. [\[CrossRef\]](#)
13. 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. [\[CrossRef\]](#)
14. Henderson-Arzapalo, A.; King, T.L. Novel microsatellite markers for Atlantic sturgeon (*Acipenser oxyrinchus*) population delineation and broodstock management. *Mol. Ecol. Notes* **2022**, *2*, 437–439. [\[CrossRef\]](#)
15. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547. [\[CrossRef\]](#)
16. Sørensen, T. A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons. *Biol. Skr. K. Dan. Vidensk. Selsk.* **1948**, *5*, 1–34.
17. Thioulouse, J.; Dray, S.; Dufour, A.-B.; Siberchicot, A.; Jombart, T.; Pavoine, S. *Multivariate Analysis of Ecological Data with ade4*; Springer: New York, NY, USA, 2018. [\[CrossRef\]](#)
18. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; Available online: <https://ggplot2.tidyverse.org> (accessed on 5 September 2020) ISBN 978-3-319-24277-4.
19. Tortonese, E. *Acipenser naccarii* (Bonaparte, 1836). In *The Freshwater Fishes of Europe*; Holcik, J., Ed.; AULA-Verlag: Wiesbaden, Germany, 1989; pp. 286–293.
20. Boscari, E.; Wu, J.; Jiang, T.; Zhang, S.; Cattelan, S.; Wang, C.; Du, H.; Li, C.; Li, J.; Ruan, R.; et al. The last giants of the Yangtze River: A multidisciplinary picture of what remains of the endemic Chinese sturgeon. *Sci. Total Environ.* **2022**, *843*, 157011. [\[CrossRef\]](#)

21. Fontana, F.; Congiu, L.; Mudrak, V.A.; Quattro, J.M.; Smith, T.I.; Ware, K.; Doroshov, S.I. Evidence of hexaploid karyotype in shortnose sturgeon. *Genome* **2008**, *51*, 113–119. [[CrossRef](#)]
22. Palle, S.D.; Boscarì, E.; Bordignon, S.G.; Muñoz-Mora, V.H.; Bertorelle, G.; Congiu, L. Different chromosome segregation patterns coexist in the tetraploid Adriatic sturgeon *Acipenser naccarii*. *Diversity* **2022**, *14*, 745. [[CrossRef](#)]
23. Stift, M.; Berenos, C.; Kuperus, P.; Van Tienderen, P.H. 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. [[CrossRef](#)] [[PubMed](#)]
24. May, B.; Krueger, C.C.; Kincaid, H.L. Genetic variation at microsatellite loci in sturgeon: Primer sequence homology in *Acipenser* and *Scaphirhynchus*. *Can. J. Fish. Aquat. Sci.* **1997**, *54*, 1542–1547. [[CrossRef](#)]
25. Zane, L.; Patarnello, T.; Ludwig, A.; Fontana, F.; Congiu, L. Isolation and characterization of microsatellites in the Adriatic sturgeon (*Acipenser naccarii*). *Mol. Ecol. Notes* **2022**, *2*, 586–588. [[CrossRef](#)]

ANNEX C



The last giants of the Yangtze River: A multidisciplinary picture of what remains of the endemic Chinese sturgeon

Elisa Boscari^{a,b,1}, Jinming Wu^{a,1}, Tao Jiang^c, Shuhuan Zhang^d, Silvia Cattelan^b, Chengyou Wang^a, Hao Du^a, Chuangju Li^a, Junyi Li^a, Rui Ruan^a, Junmin Xu^e, Yueping Zheng^f, Stefano Dalle Palle^b, Jian Yang^{c,2}, Leonardo Congiu^{b,*,2}, Qiwei Wei^{a,d,2}

^a Key Laboratory of Freshwater Biodiversity Conservation, Ministry of Agriculture and Rural Affairs of China, Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Wuhan 430223, China

^b Ecology Research Unit, Department of Biology, University of Padua, Padova 35131, Italy

^c Key Laboratory of Fishery Ecological Environment Assessment and Resource Conservation in Middle and Lower Reaches of the Yangtze River, Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi City, Jiangsu Province 214081, China

^d Guizhou University of Traditional Chinese Medicine, Guiyang 550025, China

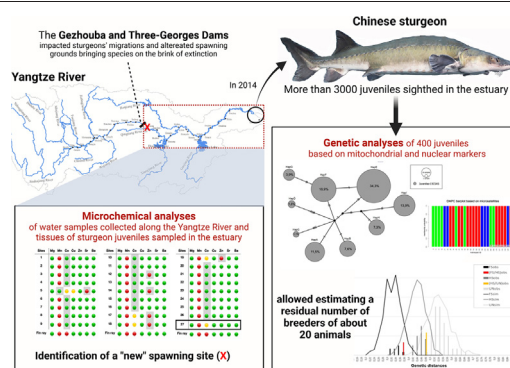
^e Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, BGI, Shenzhen 518083, China

^f Shanghai Aquatic Wildlife Conservation and Research Center, Chongming, Shanghai 200003, China

HIGHLIGHTS

- The Chinese sturgeon, endemic of the Yangtze River, is on the verge of extinction.
- Large dams on the Yangtze River have degraded spawning ground habitats.
- Microchemical analyses indicate a probable shifting of the reproductive sites.
- Genetic analyses made it possible to estimate the number of remaining breeders.
- Environmental restoration and ex-situ species management are urgently needed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Sergi Sabater

Keywords:

Acipenser sinensis residual population
Elemental analyses, genetic analyses
Gezhouba dam
Three-Gorges dam
Yangtze River

ABSTRACT

The Chinese sturgeon, an important endemism of the Yangtze River, belongs to 'the most critically endangered group of species' worldwide, with overfishing and habitat destruction being the main drivers towards extinction. Newly obtained microchemical comparisons between animals and water from different river locations revealed a probable shifting of the spawning ground few kilometers downstream compared to the only previously known site, located under the Gezhouba Dam. This offers a glimmer of hope for an adaptive response to habitat perturbation caused by the recently built Three Gorges dam on the Yangtze River. On the other hand, genetic data provide an estimate of about 20 breeders participating in the only significant breeding event of the past 10 years. This warns of a near species extinction forecast if no in situ and ex situ conservation efforts occur promptly. Given these results we propose a list of priority conservation actions that urgently need to be promoted, supported, and put into practice.

* Corresponding author at: University of Padua, Department of Biology, Via Ugo Bassi 58b, Padova, Italy.

E-mail address: leonardo.congiu@unipd.it (L. Congiu).

¹ Equal contribution

² Equal contribution to the supervision of the work and corresponding authors for elemental analyses (JY), Genetic analyses (LC) and Chinese sturgeon conservation (QW).

1. Introduction

Worldwide, more than 58,000 large dams (ICOLD, 2017) and several millions of smaller ones (Couto and Olden, 2018; Lehner et al., 2011) severely impact river habitats by disrupting sediment and nutrient transport, altering flow regimes and river bottoms, with potential effects on demography, migrations, reproductive patterns, and habitat use of local fish species (Thieme et al., 2021). Drawing an exhaustive picture of impacts of these structures on resident species is often not possible by direct observations, especially in the case of very rare and elusive species for which the size of the residual population and the habitat use patterns are difficult to unravel. Therefore, assessing the impact of damming on local species can benefit from indirect and multidisciplinary approaches. Among these, the evaluation of microchemical composition in fish hard tissues proved to be an informative technique to elucidate habitat uses at different animals ages (Feitosa et al., 2020). In this context, we try to couple genetic and microchemical approaches to evaluate the impact of damming on an imperiled sturgeon species as a case study.

Sturgeons and paddlefish are species among the most heavily impacted by overfishing and anthropogenic habitat destruction, all drivers that push them to the brink of extinction (International Union for Conservation of Nature - IUCN, www.iucnredlist.org; He et al., 2017). Despite worldwide legal protection and their listing under the Convention on International Trade in Endangered Species (CITES) agreement since 1998, the lucrative and illegal business trade of sturgeon meat and caviar has continued to deplete wild stocks (Pikitch et al., 2005). Meanwhile, hydroelectric dams and other development projects in the proximity of rivers have led to the destruction of spawning areas and to the blockade of migration routes (He et al., 2017).

Extremely critical is the situation in the Yangtze River (China), the world's third longest river, which is spanned by the largest hydroelectric dams (the Gezhouba Dam [GD], completed in 1981, and the Three-Georges Dam [TGD], in 2003; (Huang, 2019)). Prior to the dams' construction, the Yangtze was inhabited by the three most important endemic exponents of the Chinese fish fauna: two sturgeon species, the Yangtze or Dabry sturgeon (*Acipenser dabryanus*) and the Chinese sturgeon (*A. sinensis*), and the Chinese Paddlefish (*Psephurus gladius*), and their distribution range extended over 3000 km (Wei et al., 1997) in river length. After damming, the habitat stability was heavily perturbed; the dams cut off almost all known spawning sites for sturgeon species (Fig. 1A; Wu et al., 2017a, 2017b) and influenced the seasonal fluctuations in downstream water temperatures (Huang, 2019). Also, morphological features of the river bottom such as hardness, roughness and embeddedness which are important in shaping suitable spawning grounds and in providing the larvae with adequate hiding place after hatching (Huang, 2019), were significantly modified by the hydrological alteration, possibly causing the loss or the positional changes of breeding sites (Du et al., 2015). The dramatic loss of suitable hydro-ecological conditions for spawning strongly reduced the reproductive activity negatively affecting the recruitment and leading to a worsening of the general conservation status of these species to 'Critically Endangered' (Zhang et al., 2013, 2016; Zhou et al., 2014). Based on the consequent decrease of animal sightings, Huang and Wang (2018) predict, based on in silico simulations, a possible species extinction in less than 20 years. Recently, after the declaration of the extinction of the Chinese Paddlefish (Zhang et al., 2020), global attention has turned to the other two remaining species whose recovery needs urgent interventions.

Data about reproduction events of the Chinese sturgeon after 1981 officially recorded a 'new' spawning ground located just below the Gezhouba Dam, along a river stretch of 5 km (Fig. 1A; Wei, 2020). However, despite this, the number of mature animals that reach this 'new' spawning site dropped off from about 2000 in the early 1980s to less than 100/year in 2013–2017 (Wu et al., 2015). No spawning was recorded between 2013 and 2015, although a few dozen adults were sighted in that period (Wu et al., 2017a, 2017b; Zhang et al., 2016). In this context, providing data on the residual effective number of breeders and on their genetic make-up is crucial.

From April and August 2015, unexpectedly, more than 3000 juveniles were recorded in the Yangtze estuary potentially representing the most relevant signal of spawning in 10 years, which is also the last one recorded so far. Their smaller size (approximately 10 cm measured in May) compared with what has usually been observed in the past in the same area (12–15 cm) of obligate passage during the downstream migration to the sea, led previous authors to hypothesize the presence of a different spawning location, likely closest to the estuary (Wu et al., 2015; Zhang et al., 2016).

Also considering the high degradation of the only remnant 'new' spawning site, to shed light on this hypothesis and to estimate the number of remaining breeders of the 'Critically Endangered' Chinese sturgeon, we analysed a population sample of the above stock of juveniles. Using an integrated approach of genetic and elemental analyses, we hypothesized a possible ongoing process of adaptation in response to habitat destruction and provided a quantitative estimation of the status of the species. The elemental signature of pectoral fins, known to effectively provide life-history information on individual fish (Clarke et al., 2007), provided information useful for the location of a new reproductive site never used previously, while genetic analyses allowed for an accurate estimate of the number of breeders that participated in the last known spawning.

2. Material and methods

2.1. Study site, organism and sampling

The study was conducted from April to August 2015 on the Yangtze River Estuary, where more than 3000 juveniles of Chinese sturgeon have been caught after years of unobserved spawning activity. We focused on the estuary region mainly because it is the only way forward to reach the sea for this anadromous species. Moreover, the estuary represents an important feeding ground for juveniles that, based on data from past spawns, stay there for a few months between April and August before dispersing into the sea. Thus, the annual monitoring of the estuary, performed by using tidal nets, represents the best chance to sight animals generated by spawning the previous winter. Most juveniles were released back to the river immediately after capture. Only for 400 randomly selected specimens we collected a fin clip and performed genetic analyses to infer relatedness and estimate the possible number of adult spawners. Meanwhile, we also performed an elemental analysis on pectoral fin ray samples from a subgroup of 179 individuals randomly selected. The processing of fin ray samples followed Jiang et al. (2016). A laboratory experiment was performed to characterize the relationship between water and sturgeon fin ray (Phelps et al., 2012). Firstly, the elemental compositions of the fish were compared with data obtained from five cultured individuals (16APAS) collected from the Yangtze River Fisheries Research Institute in Jinzhou and with a few juveniles (Zhuang et al., 2016) collected in the Yangtze (15CMAS) to confirm the wild origin of the sighted juveniles (15CMAS June–July–August). Secondly, the elemental data from cultural ponds were also compared with water characteristics obtained from 27 sampling sites from different sections of the Yangtze (Fig. 1A) to investigate the hypothesis that this species alters its reproductive strategy in response to habitat degradation.

2.2. Elemental analysis

We performed elemental and isotopic analysis directly on fin rays by using an LA-ICPMS i.e. an NW213 laser ablation system (New Wave Research, Fremont, USA) coupled with an Agilent 7500ce ICPMS (Agilent Technologies, USA), as fin ray microchemistry is usually preferred to otolith analysis in cartilaginous fishes (Allen et al., 2018; Phelps et al., 2012, 2017; Selheim et al., 2017). The analytical process was based on replicated measurements to determine for each sample the actual limits of detection (LODs) and the relative standard deviations (%RSD) (Jiang et al., 2016).

Five water samples (50 mL each) were collected from each sampling site. Samples were filtered (first through 0.45- then through 0.02- μ m pore-size

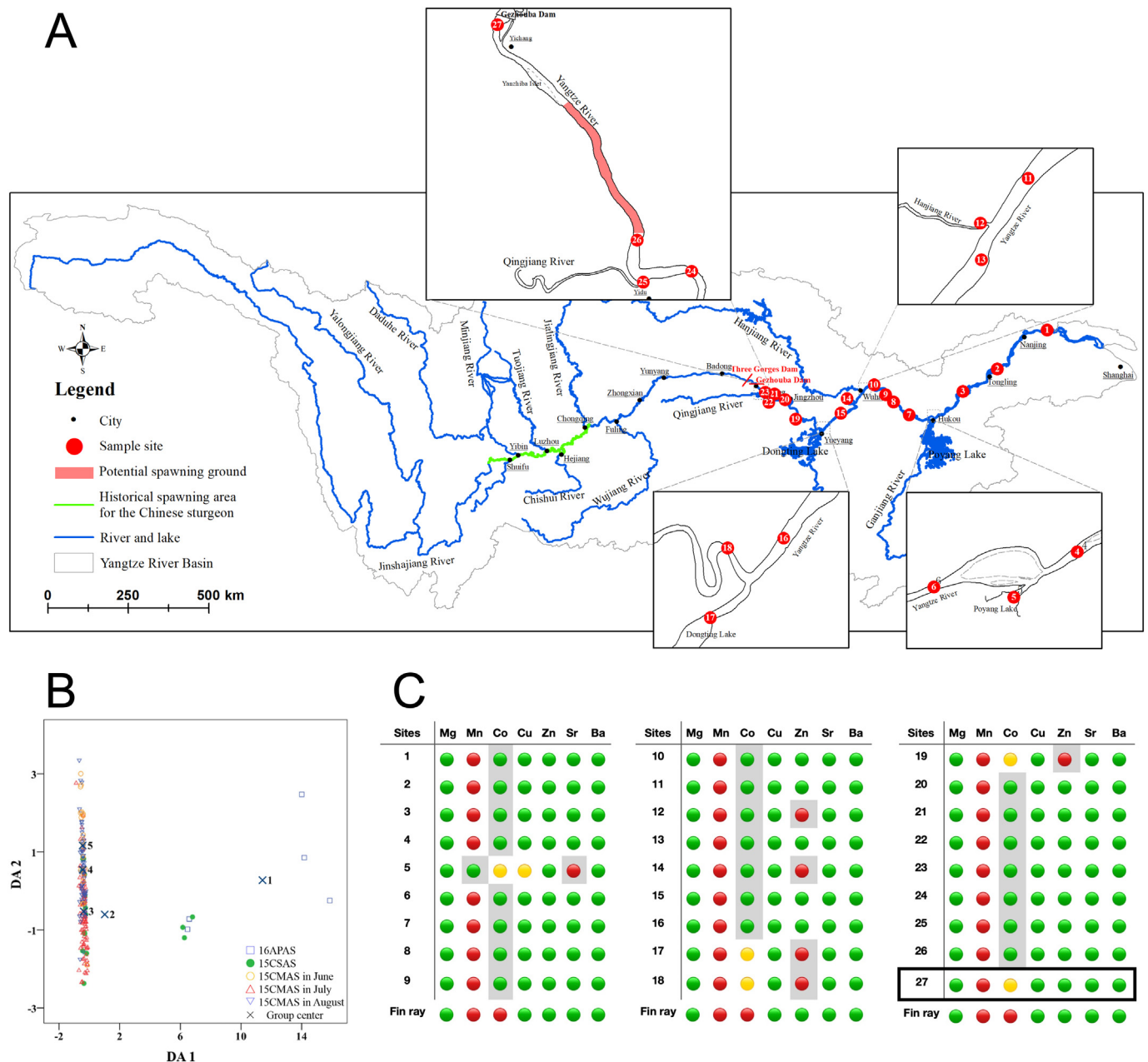


Fig. 1. Study area, sampling sites, and microchemistry results. **A**, Map of the Yangtze River and tributaries upstream and downstream the TGD and GD (red bars). The region where historical spawning grounds for the Chinese sturgeon were known is highlighted in green. The numbers within red circles are referred to the sampling sites interested by the water collection for elemental analysis. 1: Zhengjiang, 2: Wuwei, 3: Anqing section, 4: Poyang Lake Downstream (D), 5: Poyang Lake Mouth (M), 6: Poyang Lake Upstream (U), 7: Wuxue, 8: Huangshi, 9: Ezhou, 10: Tianxingzhou, 11: Hanjiang River D, 12: Hanjiang River Estuary (E), 13: Hanjiang River U, 14: Paizhouwan, 15: Laowan, 16: Dongting Lake D, 17: Dongting Lake M, 18: Dongting Lake U, 19: Shishou, 20: Mayangzhou, 21: Songzi River D, 22: Songzi River E, 23: Songzi River U, 24: Qingjiang River D, 25: Qingjiang River E, 26: Qingjiang River U, 27: Gezhou Dam. Very close sampling sites are displayed by pop-up windows. The only known site below the dam where spawning occurred in the past corresponds to the number 27. Potential current spawning area is highlighted in pink. **B**, Scatterplot of fin ray elemental ratios of Chinese sturgeon samples for wild versus cultured origin discrimination. 16APAS = cultured samples taken from the Yangtze River Fisheries Research Institute in Jinzhou; 15CSAS = juvenile animals taken from the Changshu Section in the Yangtze; 15CMAS in June, 15CMAS in July, 15CMAS in August = juveniles object of the study collected from the Yangtze Estuary, grouped by harvest month. Analytical standard errors were 5.4×10^{-1} for Mg/Ca, 4.3×10^{-4} for Mn/Ca, 8.4×10^{-6} for Co/Ca, 2.4×10^{-4} for Cu/Ca, 1.1×10^{-3} for Zn/Ca, 1.7×10^{-3} for Sr/Ca, and 3.3×10^{-4} for Ba/Ca. **C**, Graphical representation of elemental concentrations to visualize the more likely origin of wild juveniles. The presence of the different elements in each site (1–27) is codified as the ratio between natural water and pond water. The same approach was used to represent the elemental concentrations in the fin rays, expressed as the ratio between values measured in wild juveniles and those measured in cultured animals. Red dot = natural water elemental or juvenile fin ray ratios are lower than the cultured. Green dot = natural water elemental or juvenile fin ray ratios are higher than the cultured. Yellow dot = not detected. Elemental ratios shadowed in grey are discordant with the fin ray values. The black frame identifies the more probable site of origin.

filters) and analysed for elemental concentrations using inductively coupled plasma mass spectrometry. A total of eight elemental concentrations were estimated (Mg, Mn, Co, Cu, Zn, Sr, Ba and Ca) and converted to molar

elements, Ca ratios (mmol/mol). Linear Discriminant Analysis (LDA) performed by SPSS 23.0 (IBM SPSS Statistics Inc., Chicago, IL, USA) was used to analyse and visualize relationships among groups of collected animals.

2.3. Genetic analysis and definition of the number of spawners

We purified total genomic DNA following a salt-extraction method (Salah and Iciar, 1997) and analysed all samples through mitochondrial and nuclear markers to infer relatedness among samples.

The mitochondrial control region was amplified using the primer pair DL-H740 and conditions reported in the original reference (Liu et al., 2018). All successful amplifications were purified (Liu et al., 2018) and sequenced in both directions with the ABI BigDye Terminator mix followed by electrophoresis in an ABI3730 automated sequencer (Sangon Biotech Co., Ltd., Shanghai). Sequences were checked and aligned by MegaX (Kumar et al., 2018), and different mtDNA haplotypes were identified by TCS v 1.13 (Clement et al., 2000) and graphically represented by PopArt 1.7 (Leigh and Bryant, 2015).

All specimens were also genotyped at 10 microsatellite loci (Asi-66,034, Asi-65,194, Asi-75,067, Asi-77,057, Asi-72,040, Asi-67,648, Asi-73,834, Asi-70,421, Asi56700 and Asi-67,123) following recommendations reported in the original reference (Xin et al., 2016). The locus Asi-67,123 was discarded and the final dataset composed by 382 animals, genotyped at nine loci, was analysed through the Discriminant Analysis of Principal Components (DAPC; Jombart et al., 2010) implemented in the R package 'adeqenet' (Jombart, 2008) to estimate the best number of genetically differentiated clusters (K) within each haplogroup (A-L). We named each identified K with a capital letter identifying the haplogroup (A-L) and a number that refers to a differentiated sub-group within such a haplogroup (e.g. K_{A1}, K_{A2}, K_{A3}, K_{B1}, ..., K_{Hn}). To run DAPC, microsatellite data obtained from this polyploid species were turned into dominant data, as the exact number of copies of each allele at each locus cannot be accurately resolved (Boscari et al., 2014a, 2014b).

The effective degree of kinship was estimated by comparing the genetic distances between the identified groups (K_{Hn}) with simulated distributions of expected genetic distances between groups with known degree of relatedness (full-sibs – FS, half-sibs – HS and unrelated – UN). In this regard, we used only the five loci (Asi-75,067, Asi-65,194, Asi-70,421, Asi-66,034 and Asi-56,700) that did not show any sign of possible duplication (more than four alleles per individual genotype) and for which the mode of inheritance (Boscari et al., 2011) had previously been verified by genotyping of real familiar groups (data not shown).

We implemented the generation of multi-locus genotypes of virtual individuals with a known degree of relatedness in 'BreedingSturgeon', a dedicated software program developed for pedigree reconstruction and generation of virtual offspring in tetraploid sturgeon species (Boscari et al., 2014a, 2014b). The software, starting from real multi-locus profiles, generates virtual progenies as follows. Firstly, it infers all possible tetraploid genotypes for each 'parental' individual at each locus and then generates random mating pairs by simulating gametes and randomly drawing alleles according to the tetrasomic mode of inheritance (Boscari et al., 2011). The whole procedure finally results in the generation of offspring multi-locus profiles that are converted into allele profiles to be used in diversity analyses. We generated 198 groups of FS with 112 virtual individuals each, starting from a panel of 66 parental female and 66 parental male multi-locus profiles. These virtual offspring profiles were pairwise compared, resulting in 6216 comparisons between FS per each mate (198), 396 comparisons between groups of HS (virtual individuals that share one of the two parents) and 396 with groups of UN (virtual individuals that did not share either of the parents). All obtained virtual pairwise genetic distances were used to draw reference distributions for FS, HS and UN, based, respectively, on 1.230.778 comparisons for FS, and 4.967.424 for HS and UN.

We compared the above simulated distributions with the average observed band sharing (BS) within and between K_{Hn} groups to estimate kinship in the observed data. Statistical significance was obtained by comparing an equal number of randomly selected comparisons within the observed K_{Hn} group and the three reference distributions. Thus, we counted how many times the expected averages for the three simulated distributions exceeded the observed values out of 200 replicates and expressed it as a

percentage. For instance, a result of 50 % means the observed average is exactly in the middle of the distribution of a certain category, implying that the probability of belonging to that category is 100 %. The probability decreases in both directions, reaching zero at the two extremities of the distribution. We applied this reasoning to each comparison in order to calculate the probability of belonging to the FS, HS and UN categories. For the FS category, to obtain probability values, we multiplied the original percentage by 2 if it was below 50 %, and if it was above 50 % the probability was necessarily fixed at 100 %. For HS, we multiplied the original percentage by 2 if it was below 50 %, and if it was above 50 %, we subtracted the original percentage from 100 and multiplied by 2. Finally, for UN, if the original percentage was above 50 %, we subtracted it from 100 and multiplied by 2, while if it was below 50 % the probability was necessarily fixed at 100 %. If the difference between the assignment probabilities between two categories was lower than 10 %, the assignment was considered unresolved. Results of kinship assignments were represented by a heatmap using the R function heatmap.2 (R package version 2.17.0 – <http://CRAN.R-project.org/package=gplots>).

3. Results

3.1. Elemental data reveal a possible downstream shifting of spawning

For the seven elemental ratios considered, the LODs (mmol/mol) were 3.99×10^{-4} (Mg/Ca), 3.45×10^{-3} (Mn/Ca), 7.14×10^{-5} (Co/Ca), 1.62×10^{-4} (Cu/Ca), 1.28×10^{-3} (Zn/Ca), 6.84×10^{-6} (Sr/Ca) and 1.96×10^{-4} (Ba/Ca), and the %RDS, which reflects the analytical accuracy of the standard across all samples, ranged from 1.55 (Cu) to 9.05 (Mg).

A clear separation between cultured Chinese sturgeon (16APAS) and wild Chinese sturgeon was detectable (Fig. 1B). Among the 15 juveniles (15CSAS) taken from the Changshu Section in the Yangtze, 12 appeared to be mixed with the juveniles collected in the Estuary (15CMAS), while the other three appeared more similar to some cultured ones (16APAS), leading us to hypothesize that they might be released animals.

Our microchemistry results clearly show that there would be little possibility for the hatch/spawning ground to be located downstream from the Qingjiang River (Sites n. 24–26 in Fig. 1A and C). In particular, the Co/Ca ratio shows significant variation among sites and, in combination with Mn, Cu, Zn and Sr/Ca ratios, has been used to identify site no. 27 as the most likely hatching site of our specimens. Accordingly, it is reasonable to hypothesize that the spawning area is located between the Qingjiang River mouth and the Gezhouba Dam. Raw elemental data for fin rays and water samples, respectively, are reported in Supplementary Tables S1 and S2.

3.2. Genetic inferences reveal the actual number of breeders

By analysing 382 mitochondrial sequences, we identified 21 variable sites resulting in only nine haplotypes (A-I) differing from each other by a few mutations (from 1 to 10). The frequencies of the nine haplotypes appear unevenly distributed within the population sample (see the haplotype network in Fig. 2A). Based on microsatellite information, DAPC (Fig. 2A) allowed to identified a total of 31 genetically differentiated groups of animals (K_{Hn}) with a sample size ranging from 2 to 34 individuals. The pairwise comparisons within and between groups have been compared with the reference distributions of simulated degrees of relatedness (Fig. 2B) to infer the most likely kinship category. Comparisons involving groups with at least eight individuals are reported in Fig. 2B (105 comparisons), while the complete histogram showing the distribution of all the 496 comparisons (Supplementary Fig. S1). For each comparison the probability of belonging to each of the three kinship categories was estimated (Fig. 2C).

The mean genetic distance within and between groups and the final kinship assignment for each of the 496 comparisons between K_{Hn} groups are represented in Fig. 3.

These results allow estimating only 9 females with different mitochondrial haplotypes, and consequently unrelated, participated to the spawning.

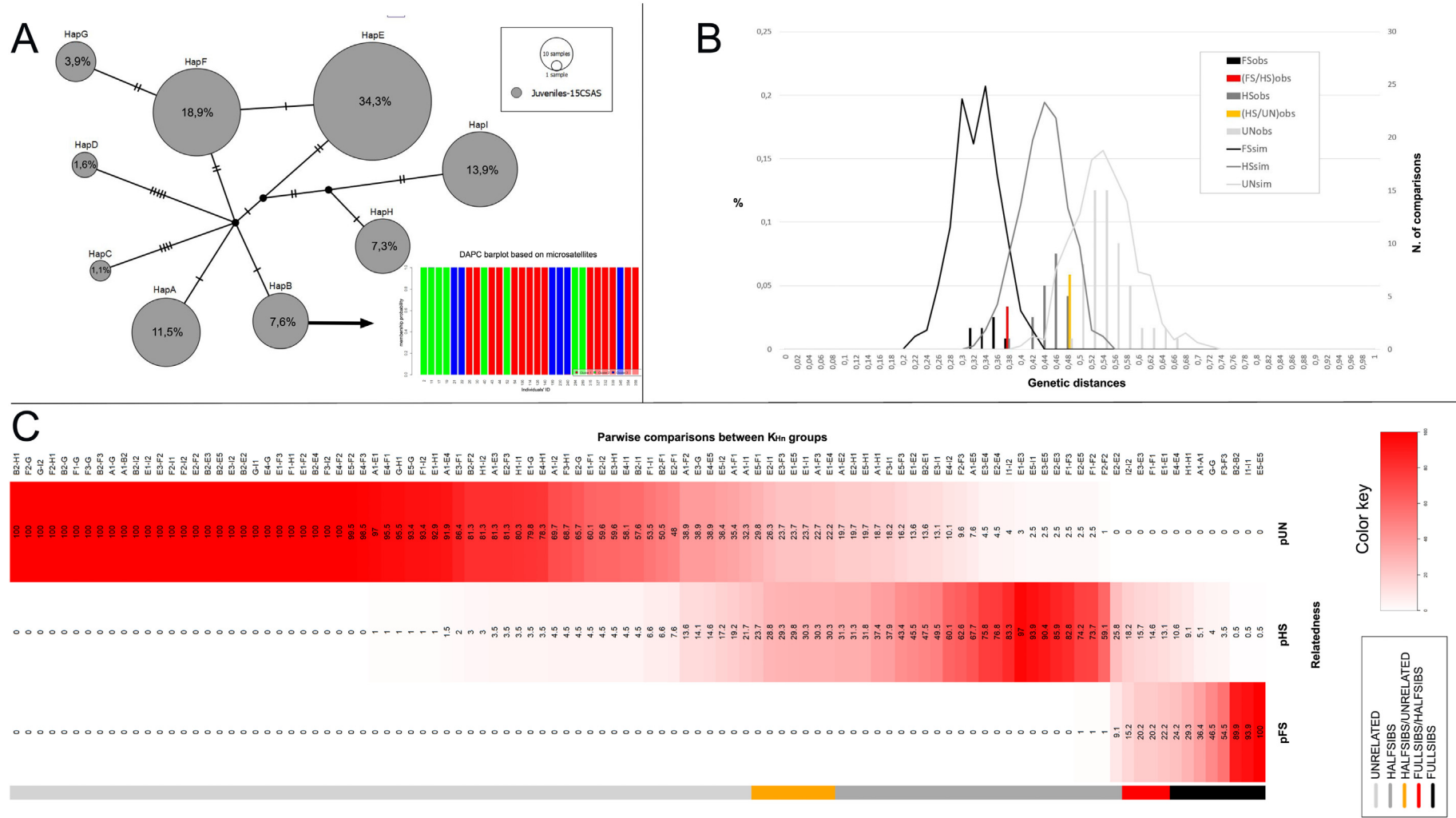


Fig. 2. Kinship estimation between genetically differentiated groups of animals. A, Haplotype network showing relationships among the nine mitochondrial haplotypes detected in the population sample and their relative frequencies expressed in percentages. The inset bar-plot represents a result example of the DAPC analysis based on microsatellite data and performed to detect differentiated K_{Hn} groups within each haplogroup. The example is referred to Haplotype B for which we considered three differentiated cluster of individuals (K_{B1} , K_{B2} , K_{B3} , corresponding to groups reported in Fig. 3). B, The three curves represented by continuous lines show the pairwise genetic distributions of simulated groups of animals with known degree of relatedness: Full Sibs (FS), Half sibs (HS), and Unrelated (UN). The superimposed bar-plot shows the distribution of the mean observed pairwise distances between K_{Hn} groups considering groups with a number of individuals greater than 8 (105 comparisons) and the different colors represent the final kinship assignment: FS = black, HS = dark grey, UN = light grey. Comparisons whose probability of assignments have a difference of less than 10 % for two different kinship categories remains unresolved and are represented in red (FS/HS) and orange (HS/UN). The same curved with a superimposed bar-plot based on all 496 comparisons is reported in Supplementary Fig. S1. C, Heatmap showing the probability of assignment to the three kinship categories for 105 comparisons between K_{Hn} groups with a sample size greater than 8. The leftside-coloured bar shows the final assignment obtained considering as threshold to accept the assignment a difference greater than 10 % between the probabilities in different categories. All comparisons here shown are the same reported as histogram in Fig. 2B. The final category assignment for all 496 comparisons between the 31 K_{Hn} groups is schematically represented in Fig. 3.

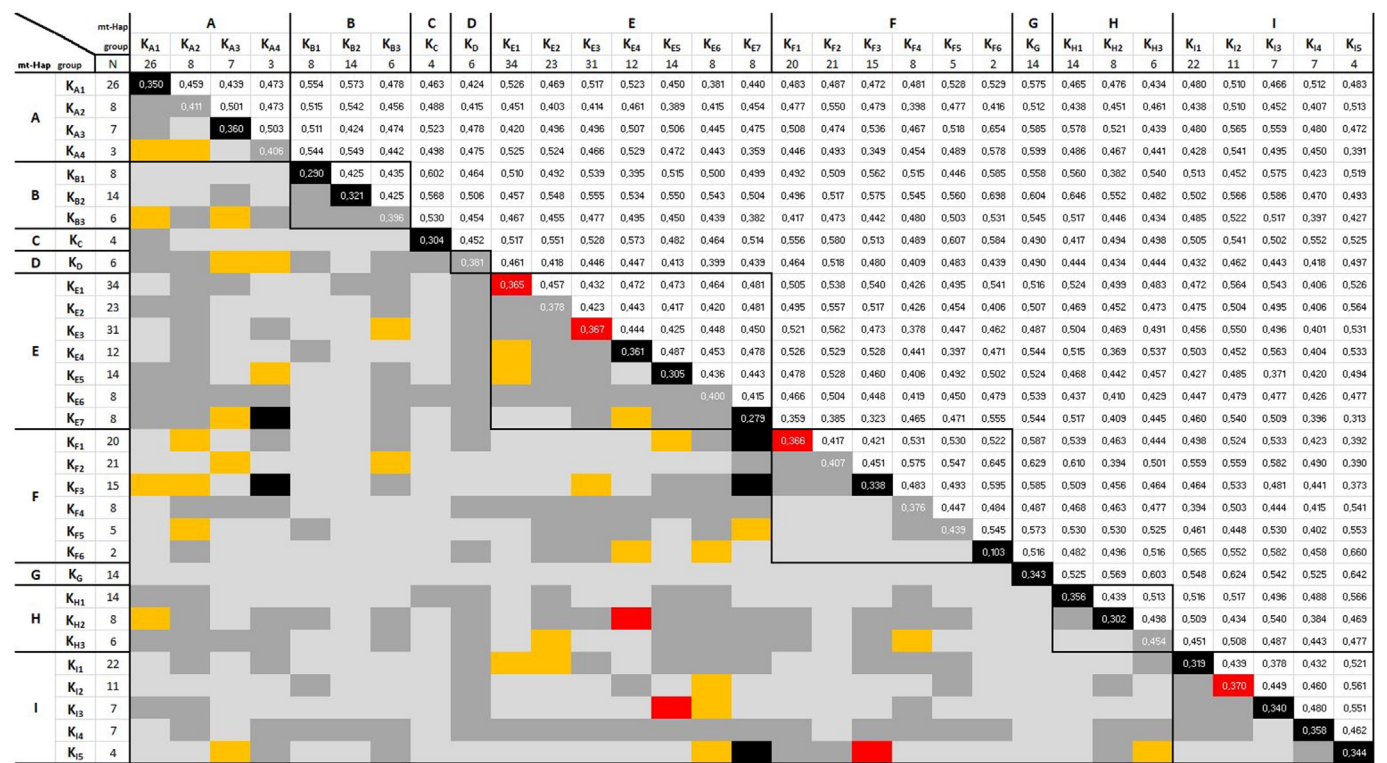


Fig. 3. Matrix showing a synthesis of genetic results. Individuals are grouped by mitochondrial haplotype (mt-DNA) represented by capital letters (A-I). Within each haplogroup, individuals are further grouped in genetically differentiated cluster (K_{Hn}) identified by analysing 9 microsatellite loci with DAPC. The number of animals (N) per K_{Hn} is also reported. The upper side of the triangular matrix reports the average value of genetic distances observed within and between K_{Hn} groups. Genetic distances, here approximated to three decimal places, have been estimated by Band Sharing (BS) (Congiu et al., 2011) based on the five microsatellite loci showing tetrasomic inheritance (Boscarì et al., 2011) and selected to simulate virtual familial groups. The lower side of the triangular matrix shows the more likely kinship category obtained by simulations. Colors Black, Dark Grey, and Light Grey represent the category of Full Sibs, Half Sibs, and Unrelated, respectively. While colors red and yellow identified familial comparisons for which the kinship category were classified as unresolved (red = Full Sibs/Half Sibs, yellow = Half Sibs/Unrelated). Familial comparisons within haplogroup are highlighted by a bold rectangle.

Conversely, the number of males cannot be surely assessed due to the low sample size of some familial groups but based on the overall analysis of progeny relatedness, it is likely similar to the number of females. In conclusion, a total number of about 20 breeders contributed to the spawning in 2014.

4. Discussion

Monitoring spawning grounds and/or detecting new ones are the first important tasks in boosting protection of species in perturbed or fragmented habitats. However, traditional survey methods are extremely challenging in extensive environments and further limited by the complex life history of anadromous species. We applied the more promising method of hard tissue microchemistry (Jiang et al., 2016, 2021; Wu et al., 2021) to gather information on the possible origin of this population of juveniles.

By analysing seven elemental content ratio patterns in a subgroup of animals, we were clearly able to differentiate between wild and cultured animals, thus confirming the wild origin of the sampled juveniles. This evidence confirms that natural propagation of the Chinese sturgeon has occurred in the Yangtze River even though no spawning activity was directly observed in the 20 Km section immediately downstream of the GD in 2014.

As to the possible location of spawning, the area of interest has been restricted to site 27, corresponding to the area below the GD, based on the elemental pattern. Given that no spawning activity had been recorded in 2014 at the usual spawning ground just below the GD and given that there are no other tributaries flowing into the Yangtze between the GD and the Qingjiang, the putative ‘new’ hatching/spawning water should be

located in the section from the GD to 5 km upstream the Qingjiang Estuary (site n. 27 in Fig. 1A and C). Accordingly, adult spawners of the Chinese sturgeon usually prefer pebble bottoms for spawning grounds (Huang, 2019), and careful surveys allowed us to exclude the presence of such a riverbed below the Qingjiang Estuary; only sandy or sludge bottoms were detected instead. This allowed us to posit that the Chinese sturgeon would hardly succeed in reproducing in downstream sections, thus corroborating our findings. Since 2003, the water discharge caused by the completion of the Three Gorges Reservoir water storage degraded the quality of the pebbly riverbed in the original spawning site immediately downstream the GD and resulted in the accumulation of a potentially suitable cobble-gravel riverbed 20-50 km further downstream (Du et al., 2015). These considerations, taken together with the absence of observed spawning activity in the usual spawning ground and with the fin ray microchemistry data, support the hypothesis that the animals probably moved a little downstream to reproduce, perhaps to avoid the highly disturbed dam area where deep modifications of the hydrological regime with important effects on sediment flushing might have compromised the habitat suitability. Such a short shifting of the spawning site does not justify any variation in the average length of the juveniles observed in the river mouth that is probably better ascribable to other ecological or behavioural reasons.

From a genetic viewpoint, we carried out a kinship estimation based on mitochondrial and nuclear analyses. Our statistical approach identified most genetically differentiated groups as composed by FS, suggesting that the number of haplotypes approximates well the real number of breeding females. Distinct groups with the same haplotype were mostly identified as HS, pointing out the female as the shared parent. A minor fraction of within-group comparisons was preferentially assigned to HS rather than

FS, leading us to hypothesize that there could be a further genetic subdivision within those groups, but this is more likely an effect of the small sample size available for these K_{Hn} groups.

If the above counting of nine females can be considered as strongly reliable, estimating the number of males is less straightforward. In many cases, comparisons between groups with different haplotypes (different maternal contribution) were assigned to HS, suggesting that the father might be shared. Considering that estimating the exact number of males remains a speculation, owing to the very limited sample size of many K_{Hn} groups and to the presence of unresolved kinship assignments, the significant fraction of groups that share the male can be taken as a good indication that the overall number of male breeders might be considerably lower than the number of K_{Hn} groups. This allows us to hypothesize almost the same order of magnitude for the number of female and male breeders indicating a total of approximately 20 animals participating to the last significant spawning of the Chinese sturgeon.

On the other hand, since on average each female spawns every 2–7 years (Yue et al., 2015), we can reasonably hypothesize that additional animals exist that did not reproduce in 2014. Very few animals were sighted in the following years in the proximity of the GD spawning site, but no spawning activity was recorded (Wei Qiwei, Pers. Comm.), with a last small exception in 2016 when a few eggs and 22 sturgeon larvae, to which we have not had access for analyses, were observed (Wu et al., 2017a, 2017b).

Furthermore, we cannot exclude the presence of additional animals participating to the spawning without reproductive success or whose progeny was not sampled.

The implications of our results go far beyond the simple decline in recruitment due to the low number of breeders. In fact, a very small population may undergo a further acceleration of decline for different reasons among which the greater exposure to demographic or environmental stochastic events, the Allee effect (Myers et al., 1995) and the erosion of genetic diversity, which represents the adaptive potential of a population especially relevant in instable environments.

To date, even focussing strictly on the spawning stock here estimated, the presence of nine different haplotypes in such a small population suggests a discrete residual genetic diversity, of the same order of magnitude of what observed in other sturgeon species such as the Adriatic sturgeon (Boscari et al., 2014a, 2014b). Given the rapid demographical collapse of the species, this precious heritage should be the target of conservation measures that must be urgently implemented, taking the opportunity of the 10 years fishing ban imposed by the Chinese government. Additionally, in all sturgeon species mortality rate is very high in the first months of life but dramatically decreases with time (Jager et al., 2010; Wang et al., 2017). So, even if not reiterated, a remarkable recruitment like the one observed could, in the long run, significantly contribute to the survival of the species. Individuals that will survive the first year will have concrete probabilities to come back to the Yangtze River in about ten-fifteen years for reproduction and in the meanwhile effective habitat restoration measures should be established. The only way for long term protection of the Chinese sturgeon is the construction of a bypass channel connecting the medium-high stretch of the river with the ocean (Wei, 2020). In the meanwhile, it is extremely urgent to implement short-term conservation strategies that include direct actions both at ecosystem and species level. Firstly, localize and protect the spawning area predicted to be between 20 km below the original spawning ground and the mouth of Qingjiang River, as well as restoring suitable environments for reproduction in line with what already done for other sturgeon species (Český et al., 1985). Secondly, establishing ex situ conservation programs through the informed management of the over 3000 F1 kept in aquaculture and belonging to 12 year-classes (Wei, 2020). This includes the purity assessment and the genetic diversity management according to strategies already developed for other polyploid sturgeons (Congiu et al., 2011; Boscari et al., 2014a, 2014b, 2017). Thirdly, refining breeding and restocking strategies to minimize the mortality rate of released individuals, in harmony with the natural life cycle of the species including its marine phase, with a “land-sea-land” perspective (Wei, 2020).

5. Conclusions

To summarize, our estimates provide compelling evidence that the status of the iconic Chinese sturgeon has deeply worsened compared with the recent past, raising the need for an urgent and revolutionary recovery effort to reverse its destiny. The only 20 breeders that participated to the last significant spawning in 2014 showed a discrete genetic diversity that can be complemented by ex situ conservation programs using animals already present in captivity. The 10-year fishing ban imposed on the Yangtze River is a good scenario on which measures for the ecological restoration of the river and the species can be implemented. Among these, the possibility of building suitable artificial spawning grounds should be considered as a priority, together with a careful management of the existing ex situ animals that can be used to integrate the residual natural genetic diversity. In this context, the long generation time of this species, which exceeds 25–30 years, is a great advantage for minimizing the combined effects of genetic drift and adaptation to captivity on these captive stocks.

If the very low number of counted spawners poses the species on the verge of extinction, the resumption of successful spawning in a new site represents an unexpected encouraging signal of resilience. For all the above reasons, we are confident that there will still be a glimmer of hope for this species, which is one step away from extinction in the wild, if appropriate and scientifically informed actions are carried out as soon as possible.

CRedit authorship contribution statement

Elisa Boscari: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **Jinming Wu:** Conceptualization, Methodology, Writing – review & editing. **Tao Jiang:** Conceptualization, Methodology, Investigation, Writing – original draft. **Shuhuan Zhang:** Resources, Investigation. **Silvia Cattelan:** Formal analysis. **Chengyou Wang:** Resources. **Hao Du:** Resources. **Chuangju Li:** Investigation. **Junyi Li:** Resources. **Rui Ruan:** Resources. **Junmin Xu:** Resources, Investigation. **Yueping Zheng:** Resources. **Stefano Dalle Palle:** Investigation. **Jian Yang:** Supervision, Writing – review & editing. **Leonardo Congiu:** Supervision, Conceptualization, Formal analysis, Writing – review & editing. **Qiwei Wei:** Supervision, Conceptualization, Project administration, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by The Central Public-interest Scientific Institution Basal Research Fund, CAFS (2021GH08) by The Scientific Research Project of China Three Gorges Corporation (No. 201603073) and by Italian PRIN 2017 (201794ZXTL). We thank Dr. Peilin Cheng for their support in different phases of the study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2022.157011>.

References

- Allen, P.J., DeVries, R.J., Fox, D.A., Gabitov, R.I., Anderson, W.G., 2018. Trace element and strontium isotopic analysis of Gulf Sturgeon fin rays to assess habitat use. *Environ. Biol. Fish* 101 (3), 469–488.
- Boscari, E., Barbisan, F., Congiu, L., 2011. Inheritance pattern of microsatellite loci in the Adriatic sturgeon (*Acipenser naccarii*). *Aquaculture* 321, 223–229.

- Boscari, E., Barmintseva, A., Pujolar, J.M., Doukakakis, P., Mugue, N., Congiu, L., 2014a. Species and hybrid identification of sturgeon caviar: a new molecular approach to detect illegal trade. *Mol. Ecol. Resour.* 14 (3), 489–498.
- Boscari, E., Pujolar, J.M., Dupanloup, I., Corradin, R., Congiu, L., 2014b. Captive breeding programs based on family groups in polyploid sturgeons. *PLoS One* 9 (10), e110951.
- Boscari, E., Barmintseva, A., Zhang, S., Yue, H., Li, C., Shedko, S.V., Lieckfeldt, D., Ludwig, A., Wei, Q.W., Mugue, N.S., Congiu, L., 2017. Genetic identification of the caviar-producing Amur and Kaluga sturgeons revealed a high level of concealed hybridization. *Food Control* 82, 243–250.
- Českleba, D.G., AveLallemant, S., Thuemler, T.F., 1985. Artificial spawning and rearing of lake sturgeon, *Acipenser fulvescens*, in Wild Rose State Fish Hatchery, Wisconsin, 1982–1983. *Environ. Biol. Fishes* 14, 79–85.
- Clarke, A.D., Telmer, K.H., Shrimpton, J.M., 2007. Elemental analysis of otoliths, fin rays and scales: a comparison of bony structures to provide population and life-history information for the Arctic grayling (*Thymallus arcticus*). *Ecol. Freshw. Fish* 16, 354–361.
- Clement, M., Posada, D., Candler, K.A., 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657–1659.
- Congiu, L., Pujolar, J.M., Forlani, A., Cenadelli, S., Dupanloup, I., Barbisan, F., et al., 2011. Managing polyploidy in ex situ conservation genetics: the case of the critically endangered Adriatic sturgeon (*Acipenser naccarii*). *PLoS One* 6, e18249.
- Couto, T.B.A., Olden, J.D., 2018. Global proliferation of small hydropower plants – science and policy. *Front. Ecol. Environ.* 16 (2), 91–100. <https://doi.org/10.1002/fee.1746>.
- Du, H., Wei, Q., Zhang, H., Wang, C., Wu, J., Shen, L., 2015. Changes of bottom substrate characteristics in spawning ground of Chinese sturgeon downstream the Gezouba dam from impounding of the three gorge reservoir. *Acta Ecol. Sin.* 35, 3124–3131.
- Feitosa, L.M., Dressler, V., Lessa, R.P., 2020. Habitat use patterns and identification of essential habitat for an endangered coastal shark with vertebrae microchemistry: the case study of *Carcharhinus porosus*. *Front. Mar. Sci.* 7, 125. <https://doi.org/10.3389/fmars.2020.00125>.
- He, F., Zarfl, C., Bremerich, V., Henshaw, A., Darwall, W., Tockner, K., et al., 2017. Disappearing giants: a review of threats to freshwater megafauna. *WIREs Water* 4, e1208.
- Huang, Z., 2019. Drifting with flow versus self-migrating—how do young anadromous fish move to the sea? *iScience* 19, 772–785.
- Huang, Z., Wang, L., 2018. Report Yangtze dams increasingly threaten the survival of the Chinese sturgeon. *Curr. Biol.* 28, 3640–3647.
- ICOLD, 2017. ICOLD world register of dams. Retrieved August 2, 2017, from http://www.icold-cigb.net/GB/world_register/general_synthesis.asp.
- Jager, H.I., Lepla, K.B., Van Winkle, W., James, B.W., McAdam, S.O., 2010. The elusive minimum viable population size for white sturgeon. *Trans. Am. Fish. Soc.* 139, 1551–1565.
- Jiang, T., Liu, H., Lu, M., Chen, T., Yang, J., 2016. A possible connectivity among estuarine tapertail anchovy (*Coilia nasus*) populations in the Yangtze River, Yellow Sea, and Poyang Lake. *Estuarine and Coasts* 39, 1762–1768.
- Jiang, T., Wang, C.Y., Du, H., et al., 2021. A possible connectivity among estuarine tapertail anchovy (*Coilia nasus*) Populations in the Yangtze River, Yellow Sea, and Poyang Lake. *J. Fish. China* 45, 424–432.
- Jombart, T., 2008. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405.
- Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11, 94.
- Kumar, S., Stecher, G., Li, M., Niyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35 (35), 1547–1549.
- Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endean, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J.C., Rödel, R., Sindorf, N., Wisser, D., 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Front. Ecol. Environ.* 9, 494–502.
- Leigh, J.W., Bryant, D., 2015. Popart: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6 (9), 1110–1116.
- Liu, J., You, X., Xu, P., Zhuang, P., Zheng, Y., Zhang, K., et al., 2018. Assessing the genetic diversity of the critically endangered Chinese sturgeon *Acipenser sinensis* using mitochondrial markers and genome-wide single-nucleotide polymorphisms from RAD-seq. *Sci. China Life Sci.* 61, 1090–1098.
- Myers, R.A., Barrowman, N.J., Hutchings, J.A., Rosen-berg, A.A., 1995. Population dynamics of exploited fish stocks at low population levels. *Science* 269, 1106–1108.
- Phelps, Q.E., Whitedge, G.W., Tripp, S.J., Smith, K.T., Garvey, J.E., Herzog, D.P., et al., 2012. Identifying river of origin for age-0 *Scaphirhynchus sturgeons* in the Missouri and Mississippi Rivers using fin ray microchemistry. *Can. J. Fish. Aquat. Sci.* 69, 930–941.
- Phelps, Q.E., Hupfeld, R.N., Whitedge, G.W., 2017. Lake sturgeon *Acipenser fulvescens* and shovelnose sturgeon *Scaphirhynchus platyrhynchus* environmental life history revealed using pectoral fin-ray microchemistry: implications for interjurisdictional conservation through fishery closure zones. *J. Fish Biol.* 90, 626–639.
- Pikitch, E.K., Doukakakis, P., Lauck, L., Chakrabarty, P., Erickson, D.L., 2005. Status, trends and management of sturgeon and paddlefish fisheries. *Fish. Fish.* 6, 233–265.
- Salah, M.A., Iciar, M., 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Resour.* 25, 4692–4693.
- Selheim, K., Willmes, M., Hobbs, J.A., Glessner, J.J.G., Jackson, Z.J., Merz, J.E., 2017. Validating fin ray microchemistry as a tool to reconstruct the migratory history of white sturgeon. *Trans. Am. Fish. Soc.* 146, 844–857.
- Thieme, M., Tickner, D., Grill, G., Carvalho, J., Goichot, M., Hartmann, J., Higgins, J., Lehner, B., Mulligan, M., Nilsson, M., Tockner, K., Zarfl, C., Opperman, J., 2021. Navigating trade-offs between dams and river conservation. *Glob. Sustain.* 4, E17. <https://doi.org/10.1017/sus.2021.15>.
- Wang, T., Gao, X., Jakovlić, I., Liu, H.-Z., 2017. Life tables and elasticity analyses of Yangtze River fish species with implications for conservation and management. *Rev. Fish Biol. Fish.* 27, 255–266. <https://doi.org/10.1007/s11160-016-9464-8>.
- Wei, Q., 2020. Conservation of Chinese sturgeon (*Acipenser sinensis*) based on its life history: dilemma and breakthrough. *J. Lake Sci.* 32 (5), 1297–1319.
- Wei, Q., Ke, F., Zhang, J., Zhuan, P., Luo, J., Zhou, R., et al., 1997. Biology, fisheries, and conservation of sturgeons and paddlefish in China. *Environ. Biol. Fish.* 48, 241–255.
- Wu, J.M., Wang, C.Y., Zhang, H., Du, H., Liu, Z.G., Shen, L., et al. (2015) Drastic decline in spawning activity of Chinese sturgeon *Acipenser sinensis* Gray 1835 in the remaining spawning ground of the Yangtze River since the construction of hydrodams. *J. Appl. Ichthyol.* 31, 839–842.
- Wu, C., Chen, L., Gao, Y., Jiang, W., 2017a. Seaward migration behavior of juvenile second filial generation Chinese sturgeon *Acipenser sinensis* in the Yangtze River, China. *Fish. Sci.* 84, 71–78.
- Wu, J.M., Wang, C.Y., Zhang, S.H., Zhang, H., Du, H., Liu, Z.G., et al., 2017b. From continuous to occasional: small-scale natural reproduction of Chinese sturgeon occurred in the Gezouba spawning ground, Yichang, China. *J. Fish. Sci. China* 24, 425–431.
- Wu, R., Li, J.C., Zhang, C., Pan, X., Jiang, T., Yang, J., et al., 2021. Fine-scale variability in otolith chemistry: application to the life history analysis of Pacific cod (*Gadus macrocephalus*) in the Yellow Sea. *Estuar. Coast. Shelf Sci.* 257, 107392.
- Xin, M.M., Zhang, D.Q., Wang, D.Q., Li, C.J., Yue, H.M., Wei, Q.W., 2016. Twenty-four novel microsatellites for the endangered Chinese sturgeon (*Acipenser sinensis* Gray, 1835). *J. Appl. Ichthyol.* 32, 405–408.
- Yue, H., Li, C., Du, H., Zhang, S., Wei, Q., 2015. Sequencing and De Novo assembly of the gonadal transcriptome of the endangered Chinese sturgeon (*Acipenser sinensis*). *PLoS One* 10 (6), e0127332. <https://doi.org/10.1371/journal.pone.0127332>.
- Zhang, H., Yang, D.G., Wei, Q.W., Du, H., Wang, C.Y., 2013. Spatial distribution and spawning stock estimates for adult Chinese sturgeon (*Acipenser sinensis* Gray, 1835) around the only remaining spawning ground during the trial operation of the newly constructed Three Gorges Project in the Yangtze River, China. *J. Appl. Ichthyol.* 29, 1436–1440.
- Zhang, H., Wu, J.M., Wang, C.Y., Du, H., Liu, Z.G., Shen, L., et al., 2016. River temperature variations and potential effects on fish in a typical Yangtze River reach: implications for management. *Appl. Ecol. Environ. Res.* 14, 553–567.
- Zhang, H., Jaric, I., Roberts, D.L., He, Y., Du, H., Wu, J., et al., 2020. Extinction of one of the world's largest freshwater fishes: lessons for conserving the endangered Yangtze fauna. *Sci. Total Environ.* 710, 136242.
- Zhou, J., Song, L., Bi, S., Zhang, H., 2014. Assessing the effect of the Three Gorges reservoir impoundment on spawning habitat suitability of Chinese sturgeon (*Acipenser sinensis*) in Yangtze River, China. *Ecol. Inform.* 20, 33–46.
- Zhuang, P., Zhao, F., Zhang, T., Chen, Y., Liu, J., Zhang, L., et al., 2016. New evidence may support the persistence and adaptability of the near-extinct Chinese sturgeon. *Biol. Conserv.* 193, 66–69.

ANNEX D

MIPs: Multi-locus Intron Polymorphisms in species identification and population genomics

Running title: **MIPs: Multi-locus Intron Polymorphisms**

Boscari E^{1,*}, Dalle Palle S¹, Vitulo N², Scapolatiello A¹, Schiavon L¹, Cariani A^{3,4,5}, Papetti C^{1,4,6}, Zane L^{1,4,5}, Marino IAM¹ and Congiu L^{1,4,5}

¹ University of Padova, Department of Biology, Via Ugo Bassi 58b, 35121 Padova (Italy)

² University of Verona, Department of Biotechnology, Strada le Grazie, 15, 37134 Verona (Italy)

³ University of Bologna, Department Biological, Geological and Environmental Sciences, Campus of Ravenna, Via Sant'Alberto 163, 48123, Ravenna (Italy)

⁴ Consorzio Nazionale Interuniversitario Per le Scienze del Mare (CoNISMa), Piazzale Flaminio 9, 00196 Roma (Italy)

⁵ National Biodiversity Future Center, Palermo, Italy

⁶ Zoological Station Anton Dohrn, Naples, Villa Comunale, 80121 Napoli (Italy)

*corresponding author

elisa.boscari@unipd.it

ABSTRACT

- 1- Inferring species identification and studying population genomics typically require the analysis of bi-parentally inherited molecular markers shared among individuals. Currently, most used approaches ensure sufficient resolving power by genotyping few dozen highly variable microsatellite loci, with high costs and low transferability of markers across different species, or thousands of bi-allelic Single Nucleotide Polymorphisms (SNPs) through massive sequencing, with a great wasting sequencing data as most loci genotyped are not shared among individuals.
- 2- Here, we propose an alternative approach aimed at complement currently available DNA tools for the analysis of genetic diversity. We implemented an intron-targeted amplicon sequencing method to genotype Multi-locus Intron Polymorphisms (MIPs). We hypothesized that the hypervariable nature of intron regions coupled with the high inter-specific transferability of *ad hoc* selected loci make MIPs powerful multiple-SNP markers potentially suitable at different complexity levels from species and hybrids identification to population comparisons and analyses of relatedness, without the need of a priori information on the studied species.
- 3- Applying our methods to teleosts, we present the first panel of MIPs highly transferable across fish genomes and demonstrate that our approach accurately distinguishes species, even closely related, and identifies different gradient of population structure by using a limited number of loci. We also show that selecting intron loci with potentially highly conserved exon flanking regions reduces computational challenges inherent to post-processing of data.
- 4- MIPs offer hypervariable 'universal' loci to study genetic diversity across species. This method, here applied to teleost as case study, can be easily transferred to any taxa unveiling an unexplored source of genetic variability. Moreover, the availability of sequence information of longer regions compared with what obtained through genotyping-by-sequencing of SNPs ease the detection of single-locus diagnostic markers and the setup of time- and costless species identification tools.

KEYWORDS

Forensic identification, High-throughput DNA sequencing, introns, multiple-SNP haplotypes, non-model organisms, population genomics, species identification, teleosts

1 | INTRODUCTION

Species identification, population genomics and molecular ecology investigations rely on the availability of highly polymorphic independent markers. Lowering isolation and characterisation costs and increasing transferability across species are major challenges to improve efficiency and applicability of markers (Camacho-Sanchez et al., 2020; Toews & Brelsford, 2012). A special attention is devoted to non-model organisms usually involved in molecular ecology studies, for which genetic and genomic information is very scarce or incomplete. Long before the establishment of Next Generation Sequencing (NGS) technology, traditional genetic approaches preferentially involved the use of highly variable microsatellites to study patterns of distribution of genetic diversity, parental relationships, and species and hybrid identification (Camacho-Sanchez et al., 2020). Despite their high power in diversity assessment, microsatellites hold different problems mostly associated to a low transferability across species and reproducibility across laboratories, homoplasmy, as well as still high costs for markers isolation and genotyping. These problems have led more recent studies to shift the attention to panels of Single Nucleotide Polymorphism (SNP) markers (Allendorf, 2017; Camacho-Sanchez et al., 2020). SNPs are mostly bi-allelic markers less variable than microsatellites, but more abundant within genomes and therefore preferable to obtain information at different evolutionary scales (Camacho-Sanchez et al., 2020). Typically, methods relying on

Reduced Representation Sequencing (RRS) techniques, as different variants of the Restriction site-Associated DNA sequencing (RADseq; Baird et al., 2008) and Genotyping by Sequencing (GBS; Elshire et al., 2011), are employed for a quick and affordable discovery of stand-alone SNPs in many individuals simultaneously. These SNP panels are analysed to understand genetic differentiation patterns, intraspecific demographic processes and interspecies relationships with high resolution power (Baetscher et al., 2018; Boscari et al., 2019; Christiansen et al., 2021; Hodel et al., 2016; Puckett, 2017). More recent studies focused on microhaplotypes (multiple, in linkage, SNPs) demonstrating how their use, instead of stand-alone SNPs, can effectively increase the informative potential for diversity analysis. This new category of nuclear markers increases the efficiency of genotyping (lowering the false SNP calling) and provides greater analytical power (Baetscher et al., 2018; Pakstis et al., 2021). We focussed on multiple-SNP alleles targeted on nuclear intronic regions thus proposing the Multi-locus Intron Polymorphisms (MIPs). Introns are non-coding portions of gene structures, carried by all eukaryotic genomes, involved in the phenomenon of splicing into a wide range of fundamental processes related to gene expression (Ressayre et al., 2015). Although introns are associated to a variety of functions, which impose strong functional constraints on at least a part of their sequence, their mutation rate is similar to that of neutral genomic regions (Forcina et al., 2021). Accordingly, even if

intron features cannot be simply associated to a random mutational model, more appropriate for real junk DNAs, overall they show a high sequence variability (Bong-Seok & Choi, 2015; Chorev & Carmel, 2012; Irimia & Roy, 2008; Yeo et al., 2004) which makes them good candidates as source of multi-allele DNA markers and a good proxy to assess genetic diversity effectively like microhaplotypes. So far, introns have already been considered as a potential alternative to commonly used nuclear markers. The pioneer study on introns involved the use of Exon-Primed Intron-Crossing PCR (EPIC-PCR; Lessa, 1992) to amplify these regions with primer pairs designed on the more conserved exon-flanking regions. However, before the advent of NGS, the application of introns as molecular markers was nontrivial and the exploitation of their full potential has been hindered by demanding lab procedures (e.g., the need of cloning amplicons to obtain trustworthy alleles for each genotype per locus). For this reason, studies have focused almost exclusively on loci showing length polymorphisms thus avoiding direct sequencing and cloning procedures (Borsa, 1999; Côrte-Real et al., 1994; Daguin & Daguin et al., 2001; Villablanca et al., 1998). The analysis of intron sequence variability with single-locus approaches has been applied only on a limited number of species (Chow & Hazama, 1998; Palumbi & Baker, 1994), mainly for species identification and forensic purposes (Boscari et al., 2014; Boscari et al., 2017a, 2017b). Only recently, with the concomitant increase of the number of available genomes, the approaches have shifted towards an *in silico* isolation of predicted introns to analyse length polymorphisms or sequence variability on a multi-locus scale (Igea et al., 2010; Li et al., 2010; Ströher et al., 2013).

However, the limited number of species used in the development and definition of intron loci have not promoted the transversal applicability to multiple species that these markers might have.

In this study, we aim at improving the current DNA tools to assess genetic diversity by a targeted amplicon high-throughput sequencing characterization of MIP markers. With this strategy, we demonstrated the informative potential and transferability of a panel of predefined intronic regions selected across several teleost genomes. Their effectiveness in species identification has been investigated on several additional species. Moreover, three species/populations case studies have been re-analysed to compare MIPs' results with those previously obtained with other types of nuclear markers. The first case study considers species identification of the Antarctic fish of the family *Channichthyidae* (icefishes, *Notothenoidei*). Congeneric icefish of the genus *Chionodraco*, among the most common species distributed on the Antarctic continental shelf, are morphologically similar and can be misidentified (Schiavon et al., 2021). Moreover, recent studies have suggested the occurrence of hybridisation among these species (Marino et al., 2013; Schiavon et al., 2021) limiting the potential for precise assessment of fish biodiversity by morphological identification only. For these reasons, we have tested species identification and verified hybridization on *Chionodraco* spp. samples previously analysed in Schiavon et al. (2021) including eleven animals for which either morphological and mitochondrial characterisation were discordant or microsatellites indicated some degrees of interspecific hybridisation. For comparative reasons, we have included

samples of other three phylogenetically very close icefish of the genera *Pagetopsis* and *Neopagetopsis*. In the second case study, we have estimated the genetic differentiation between the Common sole (*Solea solea*) and the sympatric and cryptic Egyptian sole (*S. aegyptiaca*). These two species are valuable fishery resources and their meaningful management would benefit from the availability of diagnostic markers, necessary requirements to realize fast species identification tools (Sabatini et al., 2018). The third case study investigates the population differentiation among geographical samples of *S. solea* from the Mediterranean Sea for which a divergence pattern along the west to east axis of the Mediterranean basin is expected based on a previous analysis of microsatellite markers (Garoia et al., 2007; Rolland et al., 2007).

In this study, we demonstrate how large-scale use of MIPs substantially increases the power in identifying species, even phylogenetically closely related, and how these markers are also suitable to investigate relationships among populations in genetic studies as alternative markers to microsatellites and SNPs. We also discuss limits and potentials of the MIPs panel here proposed for teleosts, and we suggest actions that could promote the implementation of the approach in other taxonomic groups by developing fast and user-friendly protocols and accessing IT support.

2 | METHODS

2.1 | Samples

Tissue samples were collected from a total of 361 individuals: 250 belong to 65 species (20 different taxonomic orders) and 11 samples with no clear classification (*Chionodraco* spp.) (Table 1). One to five specimens per species were generally

considered to evaluate interspecific transferability of the proposed panel of markers. For the species involved in the case studies (*Chionodraco* and *Solea* spp.) aimed at estimating species detection power, larger sample sizes were considered and, for species of the genus *Solea*, different populations across their distribution area were also included (Fig. 1) to evaluate power to detect genetic differences among populations.

Almost all individuals were originally identified to the species level based on morphological characters. For some specimens with ambiguous morphological features which prevented a straightforward species identification, we also verified mitochondrial information at the Control Region (Schiavon et al., 2021). Genomic DNA was purified from fin clips using DNeasy Blood and Tissue kit (Qiagen) with an elution volume of 100 µl; obtained extracts were stored at -20°C until analysis.

2.2 | *In silico* isolation of candidate intron markers and primer design and validation

Putative intron candidate markers, with high potential to be transversally applicable for ecological studies on teleosts, were isolated by aligning 48 annotated fish genomes distributed across phylogeny (see Supplementary Material - Table S1). FASTA files and GTF files, which contain the coordinates of different annotated genes on the reference genomes, were downloaded from Ensembl database and intron positions embedded into well-conserved automatically generated exonic alignment were considered.

Firstly, according to the GTF files, the exon-flanking regions of each intron were extracted and 100 bp before and after the intron junction were retained for further analysis. On average, 135,488 sequences

were retrieved from each genome, for a total of 6,503,466 introns. Secondly, an intra-locus sequencing clustering was performed using CD-HIT program version 4.8.1 (Fu et al., 2012; Limin et al., 2012; Weizhong & Godzik, 2006) and setting i) sequence identity to 0.9, ii) the ‘-g’ option to 1 allowing for a slower but more accurate clustering, and iii) the word length to 5. The clustering step produced a total of 4,846,172 different clusters. The CD-HIT clusters were further processed to retain only those who were composed by sequences belonging to at least 15 different genomes with the aim to identify candidate intron markers whose exon-flanking regions are potentially highly conserved across teleost genomes. Only 2,441 clusters passed the above filter and were considered for further analysis, for a total of 43,744 sequences.

With the final aim to further maximise the species representation within filtered clusters, the last step included a screen for potential non predicted genes. For each cluster, the representative sequence provided by CD-HIT were aligned on reference genomes of all species excluded during the sequencing clustering step. The alignment was implemented in BLAST (Altschul et al., 1990) by using high stringent criteria. Only sequences aligned with a percentage of identity of at least 90% and a sequence coverage of 100% were retained as previously undetected homologous sequences. As a result of this process, a total of 16,593 sequences were retrieved and added to the previously filtered clusters.

The 2,441 filtered clusters were examined for primer construction and introns potentially usable as highly transferable markers were selected to be tested on 65 species based on the following criteria: i) selection of clusters showing an average

intron length between 200 and 800 bp calculated on the genomes successfully aligned for each cluster, ii) the highest possible number of genomes matching cluster, and iii) the more conserved exon-flanking regions. These criteria allowed to concentrate efforts on loci with predicted dimension compatible with Illumina sequencing, easily amplifiable with PCR and potentially more conserved across species for which genome was not yet available.

Based on these criteria, a panel of 121 candidate intron markers has been selected for the present study and locus specific primer pairs were designed on exon-flanking regions, close to intron junctions, based on consensus sequences for each cluster (Table S2).

All primer pairs have been validated by single-locus PCR on a few individuals belonging to 15 different species (of which only two, *Danio rerio* and *Psetta maxima*, had available genome information) thus confirming their cross-applicability.

2.3 | Library preparation and amplicon sequencing

DNA concentration was normalized across individuals, and samples were processed with an optimized method to obtain individual intron libraries to be simultaneously sequenced on Illumina platform. The proposed method includes a two-steps PCR protocol (see Supplementary material - Fig. S1). The first step consisted in multiplexing intron loci using locus-specific primer-pairs opportunely modified by adding a tail at 5' position

(5'CTACACGACGCTCTTCCGATCTTCAGA3' and 5'CAGACGTGTGCTCTTCCGATCT3' respectively for forward and reverse primer) and which will act as binding site for primers used in the second step. Six

multiplex reactions (M1-6) have been performed with the Multiplex PCR Master Mix (Qiagen) following the manufactures protocol. The final PCR volume of 50 µl for each individual has been divided in three wells during amplification to reduce biases introduced by PCR. Cycling parameters were the following: denaturation at 95°C for 15', 35 cycles with denaturation at 94°C for 30'', annealing for 1'30'' at 58 and 60°C respectively for M1-2 and M3-6, and extension at 72°C for 1', followed by a final extension at 60°C for 30'. Detailed on multiplex composition are reported in Supporting Information (Table S2).

After pooling the three reactions for each multiplex and all multiplex PCRs with different loci for each sample, the second step consisted in indexing-PCRs with primer pairs usually used in the 2bRAD protocol (Boscari et al., 2019; Wang et al., 2012;) with the aim of adding individual-specific 7bp barcodes for demultiplexing and adapters compatible with Illumina platform (P5-P7). For all indexing PCRs, the following reagents have been used in a final volume of 50 µl (always divided in three reactions for amplification as before): 1X HF Buffer (5X), 0.31 mM of dNTPs (20mM), 0.5 µM of primer F and Barcoded-Primer R (10 µM), 0.2 µM of primer 2bRAD_amp_F and 2bRAD_amp_R (10 µM), 0.04 U/µl of Taq Phusion (New England Biolabs) and the purified multiplexed PCR products opportunely diluted (1:200). Cycling parameters were the following: denaturation at 95°C for 3', 10 cycles with denaturation at 94°C for 30'', annealing at 60°C for 30'', and extension at 72°C for 45'', and a final extension at 72°C for 5'.

Size selection and purification have been performed for each individual library after the first amplification step, and for the final library after indexing-PCRs and pooling

amplicons from different samples. The size selection involved the use of SPRISelect magnetic beads (Beckman Coulter, 0.5X and 0.56X are the used ratios for the first and second purification steps respectively) and allowed removing primer dimers and amplicons larger than 800 bp which might negatively affect sequencing performances. The final libraries of targeted intron amplicons were visualized on agarose gel and quantified by TapeStation 4150 (Agilent);

A total of 384 samples (351 individuals of which 33 processed as replicates) were successfully amplified and sequenced on two 300 pb paired-end Illumina MiSeq v3 runs performed at the Norwegian Sequencing Centre (Oslo).

2.4 | Bioinformatic data processing for haplotype generation

The raw reads in fastq format were firstly processed with *Cutadapt* 2.8 (Martin, 2011) to demultiplex reads based on the individual-specific barcode and to search for the presence of the intron-specific primer pairs used to amplify each locus. *Cutadapt* was also used to discard low quality bases and remove any possible adapters that remain inside by setting i) quality cutoff '-q' to 15, ii) minimum reads length after trimming to 100 pb, and iii) minimum overlap length with the primer sequence to 15 bp. During the processing, when one of the paired-end sequence is eliminated, the entire for-rev pair is removed from the entire dataset. Secondly, the program FLASH v 2.2.00 (Magoc & Salzberg, 2011) was used to merge paired-ends intra-sample for each locus. Given the heterogeneous dimensional range expected for different intron loci in different species genomes, an effective merging of paired-ends is expected only for shorter introns.

Consequently, whenever the number of merged for-rev reads at a locus are less than 20% of the total, the locus has been discarded at the species level as too long to generate reliable haplotypes. At this step, also loci showing a coverage lower than 30 sequences were removed at the species level.

Finally, the *SeekDeep* pipeline (v 3.0.0) (Hathaway et al., 2018) was implemented to genotype MIP markers by generating a de novo clustering of allelic variants at each locus based on an intra- and inter-individual analyses. The *qluster* package was implemented to predict and estimate frequencies and relative abundance of trailing locus-specific alleles at individual level. *Qluster* was run using default parameters to collapse errors, group reads, and identify chimeric sequences. After that, the *processClusters* package was used to compare alleles of each locus across samples. *ProcessClusters* applies a filter to the final results of *qluster* to remove low frequency alleles, singlets and chimeric alleles getting rid of artefacts. To compare alleles across samples, the option 'noErrors' has been used.

The final output of the *SeekDeep* pipeline has been further filtered by retaining for each locus and each sample only the first n_i alleles (sorting them by decreasing frequency) whose cumulative abundance was greater than 75% of the total count.

2.5 | Selection of loci for the analysis of case studies and data analysis

The detected alleles for different loci have been grouped per species and formatted in FASTA format by using BioPhyton tools (Cock et al., 2009; Van Rossum & Drake, 1995). FASTA files for each locus per species were aligned and analysed with

MEGAX (Kumar et al., 2018) and all sequences were manually checked.

Loci were screened separately per species target of each case study. Each locus was retained in accordance with the following criteria: i) the presence of the genotype in at least the 75% of individuals per species or population, ii) an allele coverage greater than 30 high-quality assembled reads, iii) concordant genotypes between replicates, iv) no more than two called alleles per genotype, and v) the absence of a homopolymer stretch or microsatellites in the merging region of the for-rev reads.

For the final panel of 121 MIPs, the unique name of each allele was converted into *digit* codes and genotype information at each locus per individual and across samples has been concordantly converted into the corresponding digits to generate *genepop* input files for genetic analysis (Rousset, 2008). The software CONVERT (Glaubitz, 2004) was used to convert the *genepop* input files to other type of inputs for populations genetic software (e.g., ARLEQUIN, STRUCTURE).

For the case studies, Allelic Richness (A_R) and private Allelic Richness (pA_R) were calculated with HP-Rare 1.1 (Kalinowski, 2004) per population or per species and standardized to the smallest sample size. Mean number of alleles per locus (N_A), Expected and Observed heterozygosity (H_E and H_O), and inbreeding coefficient (F_{IS}) were calculated by Genetix 4.05.2 (Belkhir et al., 2000). Departure from Hardy-Weinberg Equilibrium (HWE), within populations for each locus and overall loci was estimated using the Markov chain method (dememorizations = 100,000, steps in Markov chain = 1,000,000) in ARLEQUIN 3.5 (Excoffier & Lisher, 2010). The significant threshold was set to 0.05, and then adjusted by Benjamini & Hochberg (1995) correction for multiple tests.

ARLEQUIN 3.5 was also used to calculate pairwise F_{ST} values and to perform non-hierarchical and hierarchical analysis of molecular variance (AMOVA; Excoffier & Lisher, 2010) to estimate population differentiation.

For species differentiation and population structuring, the Principal Component Analysis (PCA) performed in R with packages *adegenet* (Jombart, 2008; Jombart & Ahmed 2011) and *ggplot2* (Wickham, 2016) and the software STRUCTURE ver. 2.3.4 (Falush et al., 2003, 2007; Pritchard et al., 2000) were used to infer, with no a priori information on the species or population origin, the most likely number of differentiated clusters and to estimate admixture signals among populations. For the STRUCTURE analysis, each run was performed with 100,000 length of burnin and 1M of Markov chain Monte Carlo (MCMC) steps. An admixture model and correlated allele frequencies were used. Five replicates per run were performed testing for K from 1 to the number of species/populations expected plus two in relation to the dataset used for each run. CLUMPAK (Kopelman et al., 2015) and STRUCTURE-HARVESTER (Earl et al., 2012) were used to estimate the most likely number of K through the calculation of ΔK values following the method proposed by Evanno et al. (2005) and to show the convergence between runs for each K and graphical interpretation of the results.

3 | RESULTS

3.1 | Sequencing performance

The two MiSeq v3 runs with a total of 384 individuals (including 33 technical replicates) genotyped at 121 intron loci yielded 32,455,511 paired-end raw sequences, with a number per individual ranged from 706 to 199,362 (mean 84,520). Almost all individuals (N=23) with

less than 10,000 raw sequences were old samples of *Atherina* spp. stored for more than 20 years at -20°C indicating that the low quality of DNAs can strongly affect results. After the first filtering phase by *Cutadapt*, the 24.4% (6,370,386) of raw reads did not exceed the established thresholds and were removed from further bioinformatics steps. The 26,085,125 demultiplexed sequences that passed the above step were distributed across samples with a mean of 67,930, while the number of sequences retained per locus ranged from 10 (Locus_76) to 969,783 (Locus_15) with a mean of 215,580 pointing to high heterogeneity in locus amplification performances.

The measure used for decreasing artefacts formation in the library amplification steps successfully limited the formation of chimeric sequences that were detected by bioinformatics pipelines at a very low frequencies and easily removed from further analysis. Out of 46,464 expected single-locus genotyping (384 individuals genotyped at 121 loci), the 24.4% (11,340) were discarded during merging of paired-ends. The more frequent reasons were sequencing problems in one direction or, more rarely the excessive intron length (3,021 cases) that prevent paired-ends from merging reliably. Of the remnant 35,124 single-locus genotyping that successfully passed the merging phase, 32,499 (92.5%) showed for-rev overlapping over the 90% of the total sequences per individual per locus. After the final filtering phase by *SeekDeep*, that allows generating alleles and genotypes, a total of 25,643,808 high-quality filtered and merged sequences were retained for further analysis, with a mean of 67,130 per individuals and 215,494 per locus. The mean coverage associated to each allele per locus per individual was 1,762 reads.

3.2 | Exploring transferability of proposed intron markers

On average, from each species, the 56% of 121 MIP loci was successfully amplified and sequenced. These loci are different among species. Based on the classification reported in Betancur-R et al. (2017) and Nelson et al. (2016), for species included within Eupercaria, Ovalentaria and Carangiaria a greater number of loci were successfully sequenced with a mean of 76 loci per species (62.8%), 72 (59.6%), e 71 (58.7%) respectively for the three clades (Figs 2A and 2B). For species included in other taxa such as the orders of *Clupeiformes*, *Gadiformes*, *Cypriniformes* (with a single species tested, *Danio rerio*) and *Anguilliformes* (with a single species tested, *Anguilla anguilla*), genotyping yielded lower performant results with genotypes obtained on average at 51 loci (42.4%) per species.

Replicates showed a percentage of loci with identical genotypes on average of 87.2% and of 88.1% respectively considering high-quality filtered data and further filtering the output excluding alleles with a coverage lower than 30 sequences. Comparisons include replicates analysed and sequenced within the same MiSeq run or in different runs. Replicates within or between runs are comparable with a mean of genotype identity between replicates of $89.65\% \pm 6.05$ intra-run and $84.18\% \pm 7.42$ inter-run, pointing at a good replicability of the approach (see supplementary material – Fig. S2). Loci showing different genotypes among replicates are suboptimal loci for the species under examination and should be therefore discarded from downstream analysis.

Besides the good observed transferability of MIP markers across teleost species,

these loci also revealed a strong potential for species identification. Pairwise allele sharing between samples (within and among species) has been graphically represented in Fig. 3 including species with at least three individuals genotyped with high coverage at almost all loci. The clustering of individuals reflects perfectly the species they belong to, suggesting the presence of many private alleles per species at different loci.

3.2 | Case study 1: Exploring species identification power in Antarctic notothenioids

Results here reported are referred to 8 icefish species for a total of 81 individuals successfully sequenced. Individuals processed in replicate were considered only once. Of the 121 intron loci tested, 54 (44%) passed all filters (described in Methods, section 2.5) and were selected for notothenioid species identification in this study. Most loci resulted polymorphic with many alleles per species. Most alleles were private and therefore informative for species discrimination. The species *A. mitopterox* showed the highest Allelic richness and private Allelic richness (see supplementary material - Table S3), suggesting a clear differentiation from other tested notothenioid species also according to its phylogenetic position (Near et al., 2012, Near et al., 2018). Additional descriptive information about loci in Antarctic species are reported in Supplementary material (Table S3).

The differentiation among specimens within and among population samples based on the 54 intron loci is represented by PCAs (Fig. 4). Without considering a priori sample classification based on morphology, the first, the second and the third principal components (PC1, PC2, and PC3) explain respectively 16.14%, 10.19%

and 8.75% of the variance in the dataset, and almost all samples cluster according to the morphological identification of species. The same clustering was obtained by the STRUCTURE analysis (see supplementary material – Fig. S3). The only exception is represented by a single individual labelled as *C. hamatus* (Cham_PS82_2711) which is completely differentiated from all other *C. hamatus* specimens and also from all individuals included in the dataset. The sequencing of the mitochondrial gene coding for the Cytochrome Oxidase I (COI) confirmed the above evidence assigning the sample to the species *Trematomus scotti* (suborder Notothenioidei) not included in the tested species. This is a clear case of sample mislabelling since *C. hamatus* and *T. scotti* are morphologically very different and easy to distinguish.

The eleven animals presenting ambiguous classification based on previous analysis (Schiavon et al., 2021) were assigned in purity to different species of the genus *Chionodraco* as shown by PCA-based analysis (Fig. 4) and by Structure-based analysis (see supplementary material – Fig. S3).

3.3 | Case study 2: species and population differentiation within the genus *Solea*

Results below are referred to the species *S. solea* and *S. aegyptiaca* considering a total of 106 individuals (71 from 6 populations and 35 from 5 populations, respectively) (Fig. 1). Individuals processed in replicate were considered only once.

Of the 121 MIP markers tested, 43 (35%) passed all filters and were selected for the analysis. All selected loci were polymorphic in *S. solea* and 93.1% in *S. aegyptiaca*. The species *S. solea* showed higher genetic diversity with an overall Allelic richness (A_r) overall of 10.91 while the *S. aegyptiaca*

ranked lower with an overall A_r of 4.32 (see supplementary material - Table S4). Moreover, many private and therefore diagnostic alleles were identified at almost all loci ($pA_r=10.8$ and $pA_r=4.24$, respectively; Table S4) suggesting a high power in differentiating the two species as confirmed by cluster analysis (Fig. 5A). Out of 43 MIP markers, 39 showed alternative alleles for the two species and each one of these, taken alone, might be used to develop rapid single locus tools to differentiate the two species. No interspecific hybrids were observed suggesting complete reproductive isolation between the two species.

Without considering a priori sample classification based on morphology, the first, the second and the third principal components (PC1, PC2, and PC3) explain 42.5, 4.45 and 3.77% of the variance in the dataset, respectively. Most of the variation (PC1) is explained by species divergence (Fig. 5B) while the PC2 and PC3 explain better subtler subdivisions. In *S. aegyptiaca*, all the analyses are concordant in identifying the best clustering as congruent with population subdivision (Fig. 5B on the left). Concerning the species *S. solea*, the MIPs analysis substantially overlaps what already identified based on previous works (Corti, 2022; Corti et al., 2022), showing a gradient of divergence from the west to east side of the Mediterranean basin, with the Adriatic population characterized by admixed membership probabilities of assignment to the two main groups identified as the West and East Mediterranean clusters (Fig. 5B on the right).

The only exception is represented by three individuals fished in the North Adriatic, sampled at the Chioggia Fish Market, and morphologically classified as *S. solea* which clustered with the Adriatic population of *S.*

aegyptiaca in all the analyses. This discrepancy has been confirmed by sequencing their mitochondrial barcode that assigned all samples to the species *S. aegyptiaca*, thus confirming the original misclassification.

4 | DISCUSSION

In a context largely dominated by the use of microsatellites and SNPs analysed by whole genome approaches, such as GBS (Tibihika et al., 2019) and RAD methods (Baird et al., 2008), the development and application of nuclear intronic markers have struggled so far to take off due to demanding laboratory procedures and the need of single-locus sequencing.

Here, we exploited major advances in next-generation technologies and developed a protocol for the simultaneous analyses of multiple, highly transferable selected introns, using teleosts as test group. The results obtained go beyond our expectations and the approach based on MIPs has shown not only a high transferability but also a flexible resolving power, successfully validated both at the species and population levels.

4.1 | Transferability across teleost species

Considering the estimation of the transferability of loci across species, it must be considered that 30 (46%) and 12 (18%) of the species analysed in this work are represented by one and two individuals respectively, and that the origin of the different samples is very heterogeneous, as well as their state of sample preservation. It is therefore possible that some species are represented by few animals with too low DNA quality. This leads to an underestimation of the degree of transferability because loci for which the

amplification failed due to poor DNA quality may perform better if amplified on more individuals or on better quality DNAs. Despite this, MIP markers have proven to be highly transferable and, in most species, over 50% of loci were successfully genotyped. Given the wide taxonomic range of our samples dataset, we consider this a very good result. Even by reversing the perspective and evaluating the percentage of species on which each locus is amplifiable, the results are very satisfactory and more than 60% of the loci are amplified on more than 50% of the species. Also in this case, these percentages are probably underestimated because of DNA quality as previously described.

The panel of MIP loci proposed here is the result of a pilot experiment conducted on 65 species and 121 loci. Both the number of species analysed and the number of tested loci may be increased in the future, allowing the preparation of a reference dataset that will enable to identify a priori efficient MIP loci to be genotyped for the different species or groups of species.

4.2 | MIP loci at species and population level

The efficiency at the species level, potentially very useful for species and hybrids identification, has been clearly testified in the present study by several lines of evidence. First, in the case studies on Antarctic icefish and *Solea* spp., all species were perfectly distinguished by a panel of 54 and 43 loci respectively. Second, the perfect clustering of the individuals of other 22 taxa occurred according to the original morphological identification of species (only considering presence/absence information of alleles at different loci, Fig. 3). The only two cases of discrepancy between a priori classification

and genetic analysis turned out to be one classification error and one mislabelling: the three *Solea* samples collected in the northern Adriatic Sea, close to Chioggia and classified as *S. solea* consistently showed at all loci diagnostic alleles of sole *S. aegyptiaca*; and the notothenioid labelled as *C. hamatus* that did not cluster in any *Chionodraco* sp. pool in this study and was assigned to the species *T. scotti* based on the mitochondrial COI haplotype.

MIP markers showed an informative degree of differentiation also at population level: the analysis on different populations of the Egyptian sole showed a clear geographical pattern of differentiation. Also in the Common sole, the differentiation gradient detected by MIPs from the East to the West side of the Mediterranean basin well overlaps the results previously obtained with microsatellites (Corti, 2022; Corti et al., 2022).

The availability of multilocus codominant information with a high degree of polymorphism, biparental inheritance and an expected lower incidence of homoplasmy compared to microsatellites, suggests a potential suitability of MIPs as a tool for the analysis of hybridisation and introgression at inter- and intra-specific level.

It is advisable to evaluate the advantages and limitations of the use of MIPs by comparing them with other nuclear markers currently in use. As far as the informative power of a single locus is concerned, the high number of polymorphisms within an intron region can originate a very large number of sequences (haplotypes), each of which can be considered as an allele. Consequently, considering that the size of the analysed intronic regions ranges between 200 and 600 bp, the number of alleles for a single locus can be very high, of the order of

magnitude of microsatellites if not greater (for example the *Solea* sample showed loci with a number of alleles up to 40 per locus among MIPs used for data analyses). This represents an undoubted advantage of MIP markers which, like microsatellites, are promising as a powerful analysis tool at the population level and, due to their Mendelian inheritance, in paternity or relatedness analysis.

As for the comparison with GBS methods such as all RAD methods, the approaches are quite similar as regards the experimental effort as both require the preparation of genomic libraries and subsequent sequencing on the Illumina platform. However, there are some important differences in the type of information generated. The number of loci of GBS approaches is much greater than in MIPs and is therefore more suitable, for example, for association studies. However, the fact that SNPs are generally biallelic requires the use of a very large number of loci. The possibility to genotype multi-allele markers a priori selected to be shared among individuals greatly improves the efficiency of the approach based on MIPs, dramatically reducing the main limit of all GBS methods based on the reduction of genome complexity, which is the massive loss of a posteriori information (He et al., 2014). In fact, the rarefaction of the genome information and the unpredictable bottleneck introduced during restriction and amplification steps, produce thousands of loci but most of them are scarcely shared among individuals, needed condition to perform population genetic analyses. Therefore, this drastically reduces the promised throughput in terms of the number of loci actually usable as of the thousands of loci produced most will be discarded afterwards. In MIPs approach, the a priori selection of informative loci

specific for each case study allows a better genotyping at all loci, thus guaranteeing a good coincidence between expected and observed throughput. As a result, sequencing data wasting is reduced with a consequent costs benefit, and potential increase of the number of multiplexable individuals in a sequencing run.

Moreover, MIP markers are intrinsically highly variable and, besides being analysed considering their allele frequencies as in the case of microsatellites, they can also provide sequence information that is especially suitable for the development of diagnostic markers. This can be done also with GBS methods, but MIPs here developed, being based on MiSeq sequencing, have considerably long sequences with ample margins for the construction of single locus diagnostic primers even without the need for reference genomes. This is an important advantage to set up faster and cheaper diagnostic tools highly useful in forensic species identification.

In summary, the method we proposed shares with microsatellites the high degree of polymorphism and, with RAD methods, the possibility to simultaneously process many individuals at many loci thanks to the setting up of individual tagged libraries proposing an efficient method for genotyping highly informative loci.

4.2 | Possible developments

The potential of MIP markers is still far from being fully explored but we see several opportunities for leveraging this approach and further advance the application of MIPs at large scale.

Firstly, exploring the effectiveness of MIPs in paternity tests, relatedness analysis and genetic tagging could help provide highly informative markers for the management

of genetic diversity in species/populations at risk of extinction. Since the introns have a Mendelian pattern of inheritance, it is likely expected that they will be very efficient. Microsatellites are currently the markers of choice for this purpose but for some species they can be difficult to isolate, not enough polymorphic or difficult to interpret as in the case of polyploid species in which inferring the allelic dosage based on the intensity of the peaks is challenging. Dealing with MIPs, the exact genotype can be deduced from the coverage of the different alleles but dedicated tests to confirm this will be performed.

Secondly, developing and making available an automated and user-friendly bioinformatics pipeline for processing of sequencing outputs will facilitate workflow management and data analysis.

Thirdly, increasing efficiency and cost benefits will be achieved by developing an interactive reference database dedicated to MIP markers where all intron information available will be collected. This would help researchers in consulting, downloading, selecting a priori the best loci for each case study and, in a longer-temporal perspective, integrating data resulting thereof. The advent of a dedicated database, in combination with optimized analytical procedures for the MIPs' analysis, promises to improve the workflow to choose and genotype the most suitable universal MIP markers for each case study, thereby lowering the stochastic process that underlies the restriction-based genotyping-by-sequencing methods. Finally, another possible important development is the transfer of this approach to other systematic groups. For all taxa for which genomic information is available, at least for some species, it will be possible to identify conserved intron

regions across species and develop a panel of MIP markers. Once the panel of loci has been identified for a given systematic group, it will become the reference panel for all species in that group further speeding up the entire process.

ACKNOWLEDGEMENTS

None of the author have any conflict of interest. The present project has been partially funded by a Small Research Grant of the Fishery Society of the British Isles (FISB, n. 256475). We thank “Ittico Sostenibile” (Dr. Andrea Chinellato) which contributed to sample collection or sharing and Dr. Leonardo Girlanda that provided support to laboratory procedures. The authors are grateful to Dr. Magnus Lucassen and Nils Koschnick (Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, AWI, Bremerhaven, Germany) and the ship crew for their help in collecting samples on board the R/V Polarstern. CP acknowledges financial support of the European Marie Curie Project “Polarexpress” Grant No. 622320 and the University of Padova BIRD Grant No.164793. Lorenzo Zane acknowledges support by the Italian National Programme of Antarctic Research (PNRA) Project 2016_00307. We thank EuroFishMarket for future support in dissemination of present MIP markers applicability.

REFERENCES

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: Allozymes to genomes. *Molecular Ecology*, **26**(2), 420–430. <https://doi.org/10.1111/mec.13948>

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Baetscher, D.D., Clemente, A. J., Ng, T.C., Anderson, E.C. & Garza, J.C. (2018). Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources*, **18**, 296–305. <https://doi.org/10.1111/1755-0998.12737>

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. & Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, **3**(10), e3376. <https://doi.org/10.1371/journal.pone.0003376>

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme, F. (2000). GENETIX 4.02, Logiciel Sous Windows TM Pour La Génétique Des Populations. Laboratoire Génome, Populations Interactions, CNRS UMR 5000. Montpellier: Université de Montpellier II.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*, **57**(1), 290–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

Betancur-R, R., Wiley, E.O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G. & Orti, G. (2017). Phylogenetic classification of bony fishes. *BMC Evolutionary Biology*, **17**(1), 162. <https://doi.org/10.1186/s12862-017-0958-3>

Bong-Seok, J. & Sun Shim, C. (2015). Introns: the functional benefits of Introns in Genomes. *Genomics Information*, **13**(4), 112–118. <http://dx.doi.org/10.5808/GI.2015.13.4.112>

Boscari, E., Barmintseva, A., Pujolar, J.M., Doukakis, P., Mugue, N. & Congiu, L. (2014). Species and hybrid identification of sturgeon caviar: a new molecular approach to detect illegal trade. *Molecular Ecology Resources*, **14**(3), 489–498. <https://doi.org/10.1111/1755-0998.12203>

Boscari, E., Vitulo, N., Ludwig, A., Caruso, C., Mugue, N.S., Suci, R., Onara, D.F., Papetti, C., Marino, I.A.M., Zanem L. & Congiu L. (2017a). Fast genetic identification of the Beluga sturgeon and its sought-after caviar to stem illegal trade. *Food Control*, **75**, 145–152. <https://doi.org/10.1016/j.foodcont.2016.11.039>

Boscari, E., Barmintseva, A., Zhang, S., Yue, H., Li, C., Shedko, S.V., Lieckfeldt, D., Ludwig, A., Wei, Q.W., Mugue, N.S. & Congiu, L. (2017b). Genetic identification of the caviar-producing Amur and Kaluga sturgeons revealed a high level of concealed hybridization. *Food Control*, **82**, 243–250. <https://doi.org/10.1016/j.foodcont.2017.07.001>

Boscari, E., Abbiati, M., Badalamenti, F., Bavestrello, G., Benedetti-Cecchi, L., Cannas, R., Cau, A., Cerrano, C., Chiment, G., Costantini, F., Frascchetti, S., Ingrassio, G., Marino, I.A.M., Mastrototaro, F., Papetti, C., Paterno, M., Ponti, M., Zane, L. & Congiu, L. (2019). A population genomics insight by 2b-RAD reveals populations' uniqueness along the Italian coastline in *Leptopsammia pruvoti* (Scleractinia, Dendrophylliidae). *Diversity & Distributions*, **25**, 1101–1117. <https://doi.org/10.1111/ddi.12918>

Camacho-Sanchez, M., Velo-Antón, G., Hanson, J.O., Verissimo, A., Martinez-Solzano, I., Marques, A., Moritz, C. & Carvalho, S.B. (2020). Comparative assessment of range-wide patterns of genetic diversity and structure with SNPs and microsatellites: A case study with Iberian amphibians. *Ecology & Evolution*, **10**, 10353–10363. <https://doi.org/10.1002/ece3.6670>

Chorev, M., & Carmel, L. (2012). The Function of Introns. *Frontier in Genetics*, **3**, 55. <http://dx.doi.org/10.3389/fgene.2012.00055>

Chow, S. & Hazama, K. (1998). Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology*, **7**, 1247–1263.

Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. & de Hoon, M.J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**(11), 1422–3. <https://doi.org/10.1093/bioinformatics/btp163>

Corte-Real, H.B.S., Dizon, D.R., Holland, P.W.H. (1994). Intron-targeted PCR: a new approach to survey neutral DNA polymorphism in bivalve populations. *Marine Biology*, **120**(3), 407–413. <https://doi.org/10.1007/BF00680214>

Christiansen, H., Heindler, F.M., Helleman, B., Jossart, Q., Pasotti, F., Robert, H., Verheye, M., Danis, B., Kozzius, M., Leliaert, F., Moreau, C., Patel, T., Van de Putte, A.P., Vanreusel, A., Volckaert, F.A.M. & Schon, I., (2021). Facilitating population genomics of non-model organisms through optimized experimental design for reduced representation sequencing. *BMC Genomics*, **22**, 625. <https://doi.org/10.1186/s12864-021-07917-3>

Corti, R. Moving toward stock units identification based on spatial population structure of marine species in the Mediterranean Sea and adjacent waters through multidisciplinary and holistic approaches for the sustainability of fisheries resources. 2022 University of Bologna. PhD dissertation.

Corti, R., Piazza, E., Mokhtar-Jamaï, K., Masnadi, F., Armelloni, E.N., Scarcella, G. & Cariani, A. (2022). Seascape genomics approach to describe population structure of two marine species: *Solea solea* and *Merluccius merluccius* case studies.

Proceedings of the 9th Congress of the Italian Society for Evolutionary Biology (SIBE), 4-7 September 2022, Ancona (Italy). https://www.sibe-iseb.it/_files/ugd/744a74_4775ee0936dd424181479d6f63ef0a13.pdf

Daguin, C. & Borsa, P. (1999). Genetic characterisation of *Mytilus galloprovincialis* Lmk. In North West Africa using nuclear DNA markers. *Journal of Experimental Marine Biology and Ecology*, **235**(1), 55-65. [https://doi.org/10.1016/S0022-0981\(98\)00163-4](https://doi.org/10.1016/S0022-0981(98)00163-4)

Daguin, C., Bonhomme, F. & Borsa, P. (2001). The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism at locus mac-1. *Heredity*, **86**(3), 342-354. <https://doi.org/10.1046/j.1365-2540.2001.00832.x>

Earl, D.A. & vonHoldt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources*, **4**, 359–361. <https://doi.org/10.1007/s12686-011-9548-7>

Eastman, J.T. & McCune, A.R. (2000). Fishes on the Antarctic continental shelf: evolution of a marine species flock? *Journal of fish biology*, **57**(Suppl.A), 84-102.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., & Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**(5), e19379. <https://doi.org/10.1371/journal.pone.0019379>

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, **14**, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>

Excoffier, L. & Lisher, H.E.L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**(3), 564-567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587. <https://doi.org/10.1093/genetics/164.4.1567>

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578. <https://doi.org/10.1111/j.1471-8286.2007.01758.x>

Forcina, G., Camacho-Sanchez, M., Tuh, F.Y.Y., Moreno, S. & Leonard, J.A. (2021). Markers for genetic change. *Helyion*, **7**(1), e05583. <https://doi.org/10.1016/j.helyion.2020.e05583>

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next generation sequencing data. *Bioinformatics*, **28**, 3150-3152. <https://doi.org/10.1093/bioinformatics/bts565>

Garoia, F., Guarniero, I., Grifoni, D., Marzola, S., & Tinti, F. (2007). Comparative analysis of AFLPs and SSRs efficiency in resolving population genetic structure of Mediterranean *Solea vulgaris*. *Molecular Ecology*, **16**(7), 1377-1387. <https://doi.org/10.1111/j.1365-294X.2007.03247.x>

Glaubitz, J.C. (2004). Convert: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes*, **4**(2), 309-310. <https://doi.org/10.1111/j.1471-8286.2004.00597.x>

Hathaway, N.J., Parobek, C.M., Juliano, J.J. & Bailey, J.A. (2018). SeekDeep: single-base resolution de novo clustering for amplicon deep sequencing. *Nucleic Acids Research*, **46**(4), e21. <https://doi.org/10.1093/nar/gkx1201>

He, J., Zhao, X., Laroche, A., Lu, Z.X., Liu, H. & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, **5**. <https://doi.org/10.3389/fpls.2014.00484>

Hodel, R. G. J., Segovia-Salcedo, M. C., Landis, J. B., Crowl, A. A., Sun, M., Liu, X., ... Soltis, P. S. (2016). The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. *Applications in Plant Sciences*, **4**(6), 1600025. <https://doi.org/10.3732/apps.1600025>

Igea, J., Juste, J. & Castresana, J. (2010). Novel intron markers to study the phylogeny of closely related mammalian species. *BMC Evolutionary Biology*, **10**(1), 369. <https://doi.org/10.1186/1471-2148-10-369>

Irimia, M. & Roy, S.W. (2008). Spliceosomal introns as tools for genomic and evolutionary analysis. *Nucleic Acids Research*, **36**(5), 1703-1712. <https://doi.org/10.1093/nar/gkn012>

Jombart, T. (2008). Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403-1405. <https://doi.org/10.1093/bioinformatics/btn129>

Jombart, T. & Ahmed, I. (2011). Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btr521>

Kalinowski, S.T. (2004). HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**(1), 187-189. <https://doi.org/10.1111/j.1471-8286.2004.00845.x>

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. & Mayrose, I. (2015). CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**(5), 1179-1191. <https://doi.org/10.1111/1755-0998.12387>

Kumar, S., Stecher, G., Li, M., Knyazm C. & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology & Evolution*, **35**(6), 1547-1549. <https://doi.org/10.1093/molbev/msy096>

Lessa, E.P. (1992). Rapid surveying of DNA sequence variation in natural populations. *Molecular Biology & Evolution*, **9**, 323–330.

Li, C., Riethoven, J.J.M. & Ma, L. (2010). Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. *BMC Evolutionary Biology*, **10**, 90. <https://doi.org/10.1186/1471-2148-10-90>

Limin, F., Beifang, N., Zhengwei, Z., Sitao, W. & Weizhong, L. (2012). CD-HIT: accelerated for clustering the next generation sequencing data. *Bioinformatics*, **28**(23), 3150-3152. <https://doi.org/10.1093/bioinformatics/bts565>

Magoc, T. & Salzberg, S. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**(21), 2957-63.

Marino, I.A.M., Benazzo, A., Agostini, C., Mezzavilla, M., Hoban, S.M., Patarnello, T., Zane, L., & Bertorelle, G. (2013). Evidence for past and present hybridization in three Antarctic icefish species provides new perspectives on an evolutionary radiation. *Molecular Ecology*, **22**(20), 5148-5161. <https://doi.org/10.1111/mec.12458>

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, **17**, 10–12. ISSN 2226-6089. Available at: <<https://journal.embnet.org/index.php/embnetjournal/article/view/200>>. <https://doi.org/10.14806/ej.17.1.200>

Near, T.J., Dornburg, A., Kuhn, K.L., Eastman, J.T., Pennington, J.N., Patarnello, T., Zane, L., Fernandez, D.A. & Jones, C.D. (2012). Ancient climate change, antifreeze, and evolutionary diversification of Antarctic fishes. *PNAS*, **109**(9), 3434-3439. <https://doi.org/10.1073/pnas.1115169109>

Near, T.J., MacGuigan, D.J., Parker, E., Struthers, C.D., Jones, C.D. & Dornburg, A. (2018). Phylogenetic analysis of Antarctic notothenioids illuminates the utility of RADseq for resolving Cenozoic adaptive radiations. *Molecular Phylogenetic & Evolution*, **129**, 268-279. <https://doi.org/10.1016/j.ympev.2018.09.001>

Nelson, J.S., Grande, T.C. & Wilson, M.V. (2016). Fishes of the world. John Wiley and Sons.

- Nielsen, E., Cariani, A., Aoidh, E. et al. (2012). Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communication*, **3**, 851 (2012). <https://doi.org/10.1038/ncomms1845>
- Pakstis, A.J., Gandotra, N., Speed, W.C., Murtha, M., Scharfe, C. & Kidd, K. (2021). The population genetics characteristics of a 90 locus panel of microhaplotypes. *Human Genetics*, **140**, 1753–1773. <https://doi.org/10.1007/s00439-021-02382-0>
- Pritchard, J. K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, **155**(2), 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Puckett, E.E. (2017). Variability in total project and per sample genotyping costs under varying study designs including with microsatellites or SNPs to answer conservation genetic questions. *Conservation Genetics Resources*, **9**(2), 289–304. <https://doi.org/10.1007/s12686-016-0643-7>
- Ressayre, A., Glémin, S., Montalent, P., Serre-Giardi, L., Dillmann, C. & Joets, J. (2015). Introns Structure Patterns of Variation in Nucleotide Composition in Arabidopsis thaliana and Rice Protein-Coding Genes. *Genome Biology & Evolution*, **7**(10), 2913–28. <https://doi.org/10.1093/gbe/evl189>
- Rolland, J.L., Bonhomme, F., Lagardere, F., Hassan, M. & Guinand, B. (2007). Population structure of the common sole (*Solea solea*) in the northeastern atlantic and the Mediterranean Sea: revisiting the divide with epic markers. *Marine Biology*, **151**, 327–341. <https://doi.org/10.1007/s00227-006-0484-0>
- Rousset, F. (2008). Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**(1), 103–6. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Sabatini, L., Bullo, M., Cariani, A., Celic, I., Ferrari A., Guarniero, I., Leoni, S., Marceta, B., Marcone, A., Polidori, P., Raicevich, S., Tinti, F., Vrgoc, N. & Scarcella, G. (2018). Good practices for common sole assessment in the Adriatic Sea: Genetic and morphological differentiation of *Solea solea* (Linnaeus, 1758) from *S. aegyptiaca* (Chabanaud, 1927) and stock identification. *Journal of Sea Research*, **137**, 57–64. <https://doi.org/10.1016/j.seares.2018.04.004>
- Schiavon, L., Dulière, V., La Mesa, M., Marino, I.A.M., Codogno, G., Boscari, E., Riginella, E., Battistotti, A., Lucassen, M., Zane, L., & Papetti, C. (2021). Species distribution, hybridization and connectivity in the genus *Chionodraco*: Unveiling unknown icefish diversity in Antarctica. *Diversity & Distributions*, **27**(5), 766–789. <https://doi.org/10.1111/ddi.13249>
- Ströher, P.R., Li, C. & Pie, M.R. (2013). Exon-primed intron-crossing (EPIC) markers as a tool for ant phylogeography. *Revista Brasileira de Entomologia*, **57**(4). <https://doi.org/10.1590/S0085-56262013005000039>
- Tibihika, P.D., Curto, M., Dornstaeder-Schrammel, E. et al. (2019). Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. *Conservation Genetics*, **20**, 357–372. <https://doi.org/10.1007/s10592-018-1136-x>
- Toews, D.P.L. & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**(16), 3907–30. <https://doi.org/10.1111/j.1365-294X.2012.05664.x>
- Van Rossum, G., & Drake Jr, F. L. (1995). Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam.
- Villablanca, F.X., Roderick, G.K. & Palumbi, S.R. (1998). Invasion genetics of the Mediterranean fruit fly: variation in multiple nuclear introns. *Molecular Ecology*, **7**(5), 547–560. <https://doi.org/10.1046/j.1365.294x.1998.00351.x>
- Wang, S., Meyer, E., McKay, J., McKay, J.K. & Matz, M.V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, **9**, 808–810. <https://doi.org/10.1038/nmeth.2023>
- Weizhong, L. & Godzik, A. (2006). CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>
- Yeo, G., Hoon, S., Venkatesh, B. & Burge, C.B. (2004). Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *PNAS*, **101**(44), 15700–5. <https://doi.org/10.1073/pnas.0404901101>

DATA ACCESSIBILITY STATEMENT

Results of the *in silico* clustering across fish genomes: 2441 conserved intron loci are available on DataDryad (XXXXX)
Raw fastq sequence reads are deposited in the SRA (BioProject XXX).
Individual genotype data per locus are available on DataDryad (XXXXX)
Unique haplotype data are deposited to NCBI Nucleotide Database (XXXXX)

AUTHOR CONTRIBUTIONS

Elisa Boscari and Leonardo Congiu conceptualized the idea and designed research; Elisa Boscari, Stefano Dalle Palle and Annalisa Scapolatiello performed research; Nicola Vitulo and Annalisa Scapolatiello developed the bioinformatics pipeline and processed informatics data; Elisa Boscari, Leonardo Congiu, and Lorenzo Zane contributed new reagents or analytical tools; Elisa Boscari, Annalisa Scapolatiello and Luca Schiavon analysed genetic data; Elisa Boscari and Leonardo Congiu led the drafting of the manuscript; Alessia Cariani and Chiara Papetti provided samples; all authors contributed critically to the draft and gave final approval for publication.

TABLE 1 Species and relative sample sizes used for library preparation and sequencing. The species further involved in the proposed case studies are in bold.

Order	Species	N.	Order	Species	N.
Acanthuriformes	<i>Sciaena umbra</i>	1	Perciformes	<i>Chionodraco spp. †‡</i>	11
	<i>Umbrina cirrosa</i>	1		<i>Dicentrarchus labrax</i>	2
Anguilliformes	<i>Anguilla anguilla</i>	3		<i>Neopagetopsis ionah†</i>	6
Atheriniformes	<i>Atherina boyeri</i>	60		<i>Pagetopsis macropterus†</i>	6
	<i>Atherina hepsetus</i>	2		<i>Pagetopsis maculatus†</i>	6
	<i>Atherina presbiter</i>	1		<i>Platichthys flesus</i>	4
Beloniformes	<i>Belone belone</i>	1		<i>Scorpaena porcus</i>	1
Carangiformes	<i>Arnoglossus laterna</i>	3		<i>Scorpaena scrofa</i>	2
Clupeiformes	<i>Alosa fallax</i>	1		<i>Serranus hepatus</i>	2
	<i>Engraulis encrasicolus</i>	5		<i>Trachinus draco</i>	3
	<i>Sardina pilchardus</i>	4		<i>Trachinus radiates</i>	1
	<i>Sprattus sprattus</i>	2		<i>Trigloporus lastoviza</i>	1
	<i>Danio rerio</i>	1	Pleuronectiformes	<i>Psetta maxima</i>	1
Cypriniformes	<i>Sardinella aurita</i>	1		<i>Solea aegyptiaca</i>	35
Cyprinodontiformes	<i>Poecilia reticulata</i>	6		<i>Synapturichthys kleinii</i>	1
Gadiformes	<i>Merlangius merlangus</i>	2		<i>Scophthalmus rhombus</i>	1
	<i>Phycis blennoides</i>	1		<i>Solea solea</i>	71
	<i>Merluccius merluccius</i>	1		<i>Trachurus mediterraneus</i>	6
	<i>Deltentosteus quadrimaculatus</i>	2		<i>Trachurus trachurus</i>	1
	<i>Gobius cobitis</i>	1	Priacanthiformes	<i>Cepola macrophthalmia</i>	1
Gobiiformes	<i>Zosterisessor ophiocephalus</i>	4		<i>Cepola rubescens</i>	1
Mugiliformes	<i>Chelon auratus</i>	1	Scombriformes	<i>Pomatomus saltatrix</i>	1
	<i>Chelon labrosus</i>	2		<i>Scomber colias</i>	3
	<i>Chelon ramada</i>	1	Spariformes	<i>Diplodus annularis</i>	1
	<i>Mugil cephalus</i>	1		<i>Lophius budegassa</i>	2
Ophidiiformes	<i>Ophidion barbatum</i>	1		<i>Lithognathus mormyrus</i>	5
Perciformes	<i>Auxis thazard</i>	1		<i>Pagellus erythrinus</i>	2
	<i>Aethotaxis mitopteryx†</i>	6		<i>Sparus aurata</i>	5
	<i>Chaenodraco wilsoni†</i>	8		<i>Sarpa salpa</i>	1
	<i>Chelidonichthys lucerna</i>	1	Syngnathiformes	<i>Mullus barbatus</i>	2
	<i>Chionodraco hamatus†</i>	17		<i>Mullus surmuletus</i>	1
	<i>Chionodraco myersi†</i>	16	Uranoscopiformes	<i>Uranoscopus scaber</i>	1
	<i>Chionodraco rastropinosus†</i>	14	Zeiformes	<i>Zeus faber</i>	2

† Samples obtained during multiple *R/V Polarstern* cruises. Sampling was approved by the competent national authority for Antarctic research (Umweltbundesamt, UBA, Germany).

‡ *Chionodraco* individuals with ambiguous morphological features which prevented a clear species classification based on morphology or putative hybrids based on previous studies (Schiavon et al., 2021).

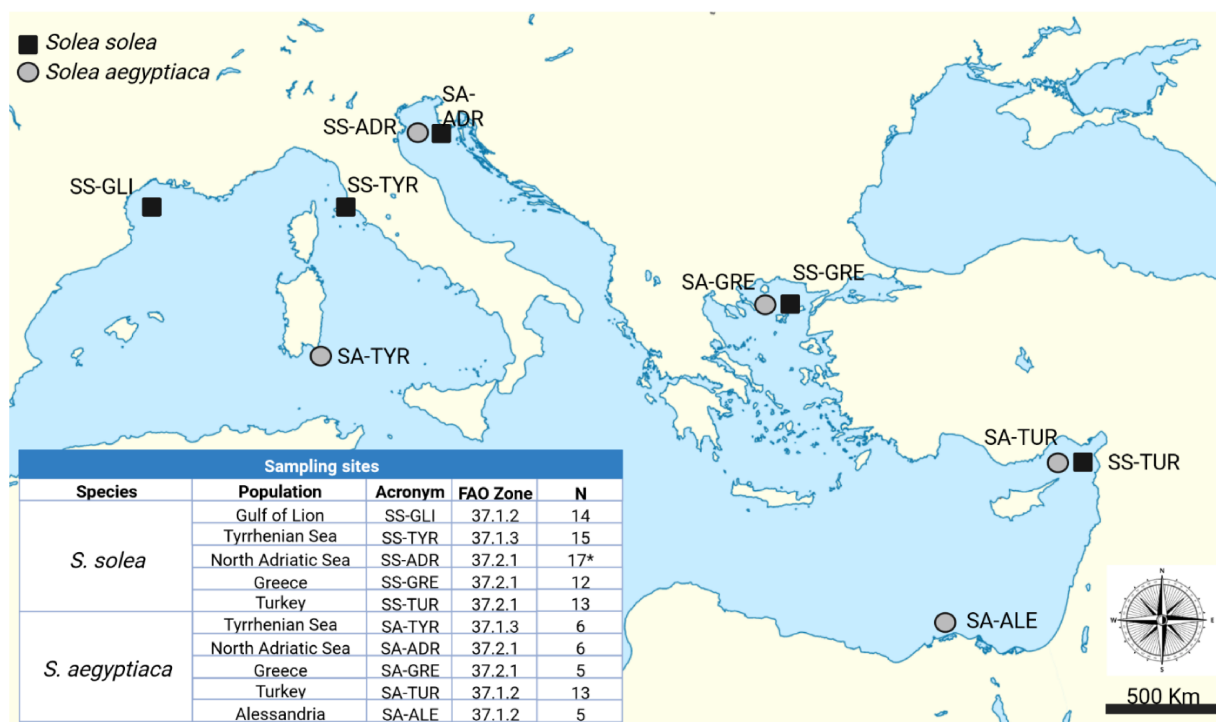


FIGURE 1

Sampling sites and population details of the species *Solea solea* and *S. aegyptiaca* involved in the case studies 2 and 3. Almost all samples are legacy FishPopTrace Consortium (Nielsen et al., 2012). The only exception is represented by three individuals of the North Adriatic population of *S. solea* (marked with * in the table inset) which were sampled at the Chioggia Fish Market and morphologically classified as *S. solea*. Created with Biorender.

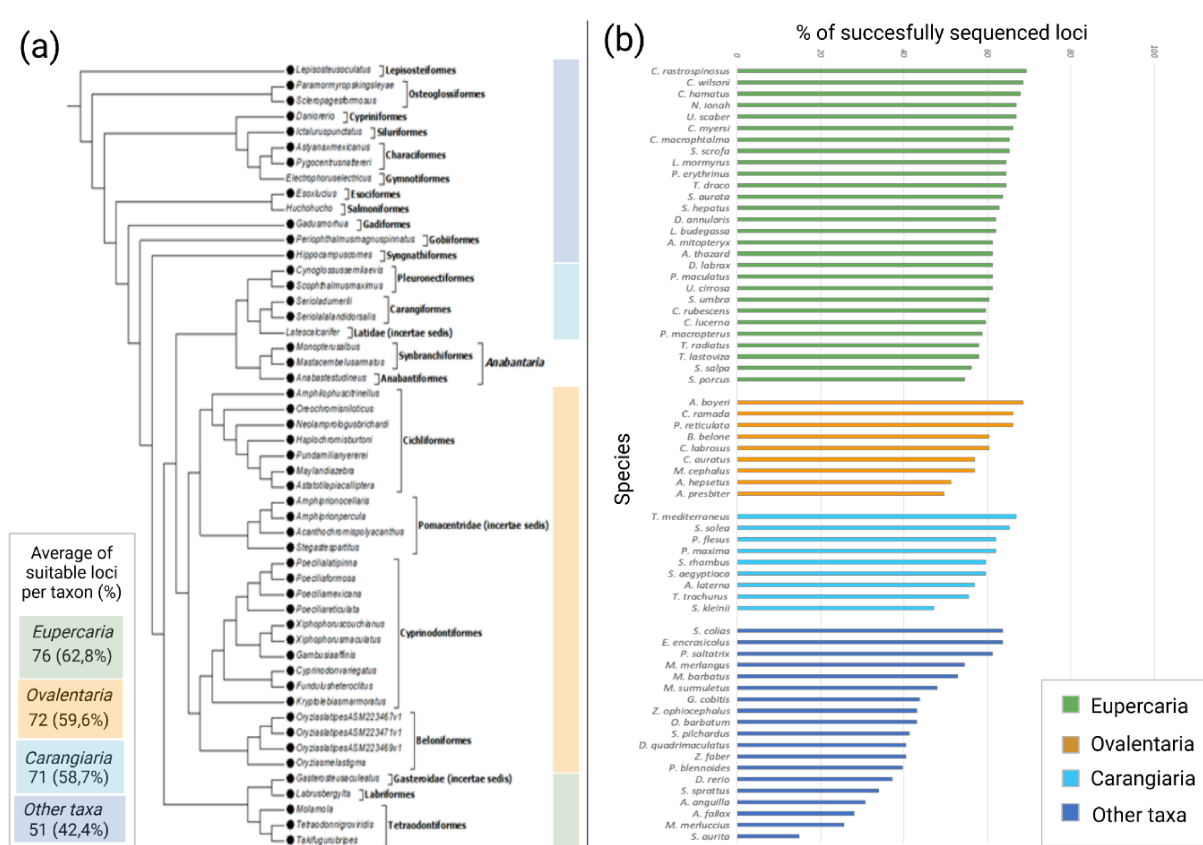


FIGURE 2

Transferability of loci across species tested and reproducibility of the approach. (a) Modified version of Ensembl phylogeny and taxon nomenclature reported by Betancur-R et al. (2017) and Nelson et al. (2016) with the mean number of loci successfully genotyped per taxon. We reported in the phylogeny only the species for which we used the genome in the first part of the isolation of intron loci. (b) Distribution of the percentage of loci successfully amplified and sequenced across 65 fish species tested. Species are grouped based on taxa reported in Fig. 2A. Created with Biorender.

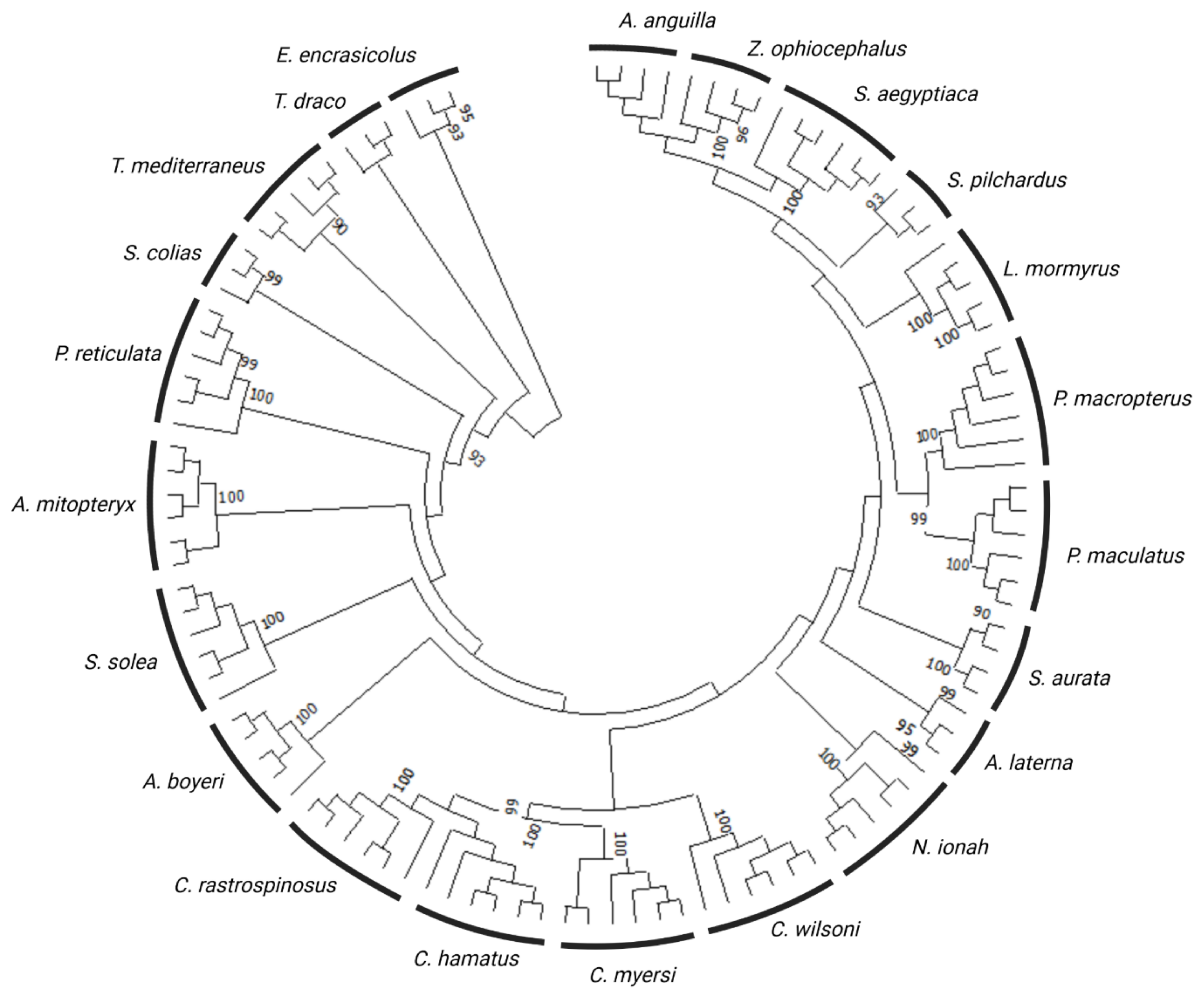


FIGURE 3

Graphical representation by UPGMA of the allele sharing across a subsample of genotyped individuals. Only species with at least three individuals were considered and, for highly represented species, a maximum of six individuals with the highest coverage were included. Species represented by a single individual and animals showing a very low coverage were discarded owing to missing data at different loci. Created with Biorender.

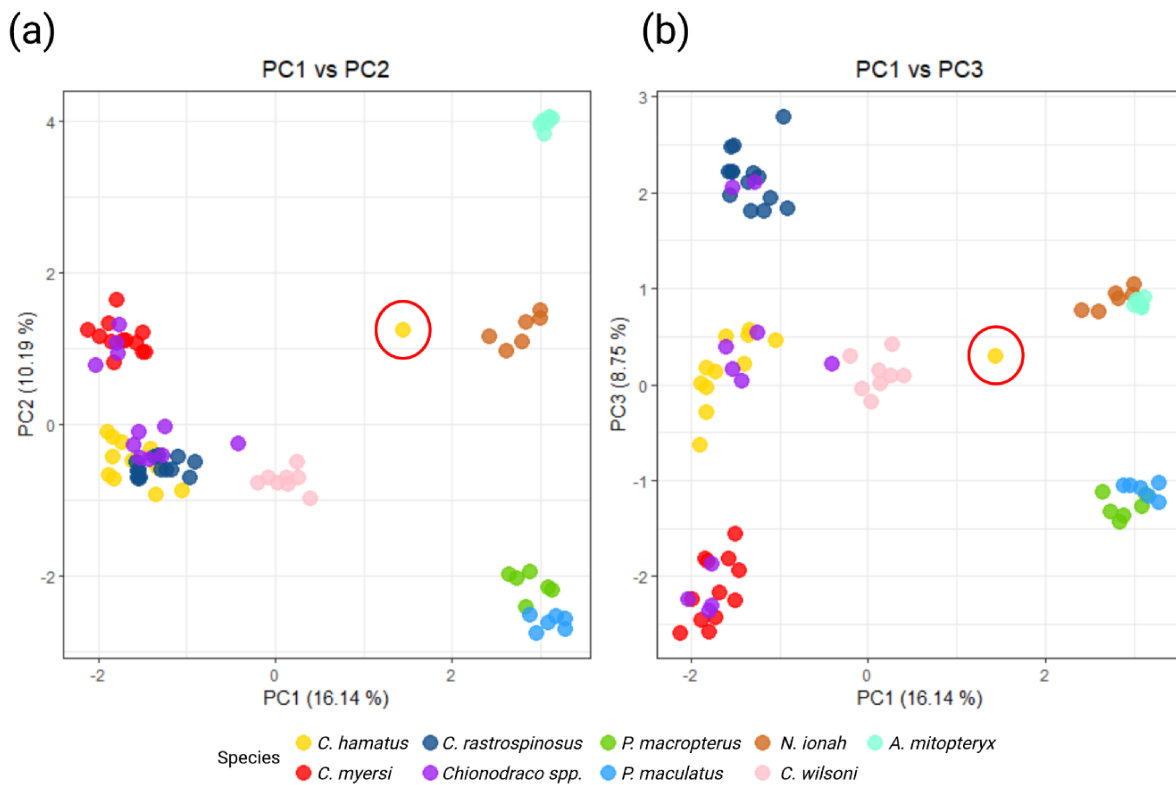


FIGURE 4

Plot of the principal component analysis of variation based on 54 intron loci genotyped at a total of 81 icefish individuals (including 8 species defined based on morphological features and 11 individuals with unknown species origin marked as *Chionodraco spp.*). Colours indicate species. Eigenvalues of the first three axes explain 16.14, 10.19 and 8.75% of the variation. The individual genetically assigned to the species *Trematomus scotti* was originally mis-assigned to *C. hamatus* due to sample mislabelling and is highlighted by a red circle. (a) PC1 vs PC2. (b) PC1 vs PC3. Created with Biorender.

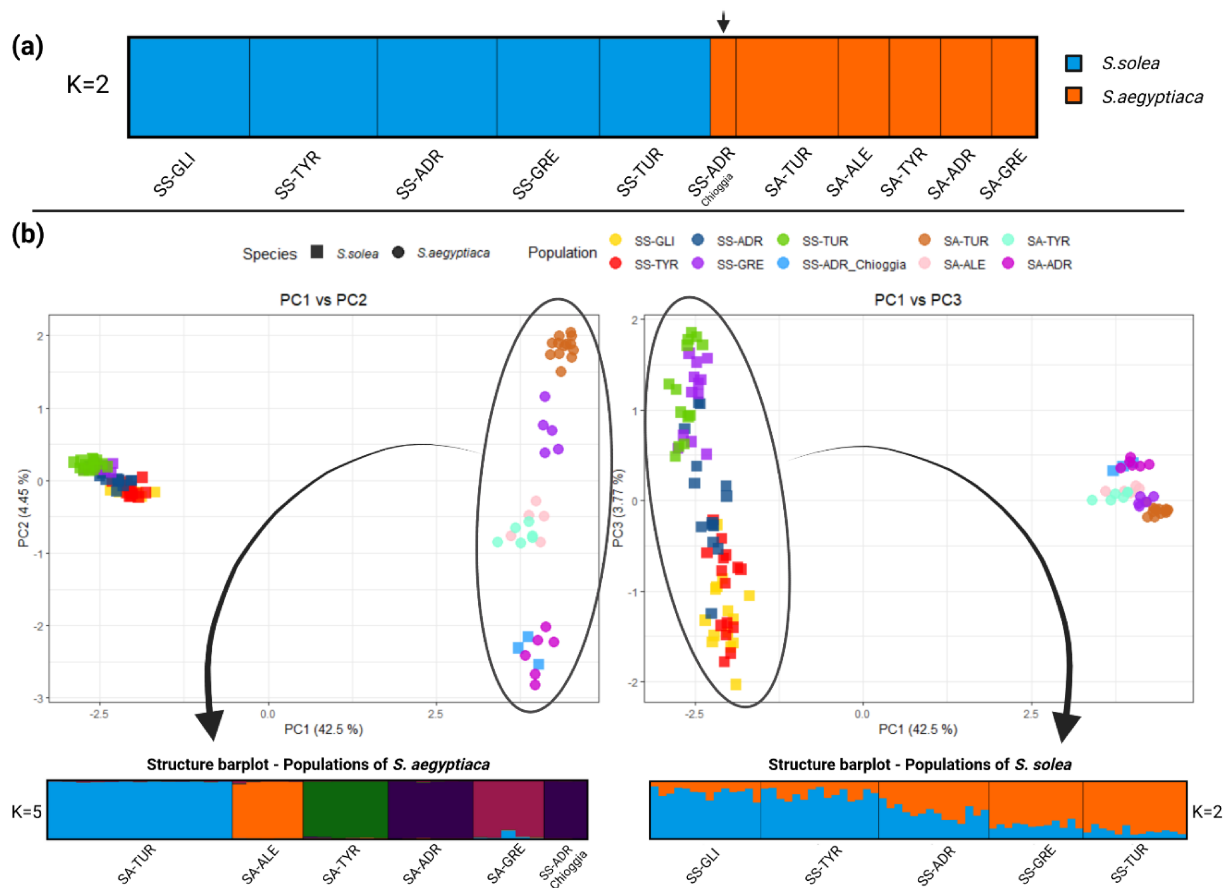


FIGURE 5

Distribution pattern of genetic differentiation between species and populations of the genus *Solea* based on the genotype information at 43 intron loci. (a) Structure bar-plot showing the membership probabilities of cluster assignment for each of the 106 individuals of the two species (71 *S. solea* and 35 *S. aegyptiaca*). The best number of cluster ($K=2$) that overlaps the species subdivision is reported. The black arrow indicates the three individuals morphologically classified as *S. solea*, which cluster with *S. aegyptiaca* according to their mitochondrial barcode information. (b) Plot of the Principal Component Analysis of variation (PC1 vs PC2 on the left, PC1 vs PC3 on the right). Eigenvalues of the first three axes explain 42.5, 4.45 and 3.77% of the variation. Shapes indicate species, while colours indicate populations. Below each PCA, the best cluster subdivision obtained by structure analysis performed within species is also reported as bar-plot showing individual membership probabilities. Created with Biorender.

Supporting Information:

Table S1. List of the 48 genomes considered in the intron *in silico* isolation phase. The name of the species, the assembly, date and version of the genome release are specified.

Species	Assembly	Data	Ensembl release
<i>Lepisosteus oculatus</i>	LepOcu1	Dec 2011	94.1
<i>Scleropages formosus</i>	ASM162426v1	Apr 2016	94.1
<i>Paramormyrops kingsleyae</i>	PKINGS_0.1	Jan 2018	94.1
<i>Danio rerio</i>	GRCz11	May 2017	94.11
<i>Ictalurus punctatus</i>	IpCoco_1.2	Jun 2016	94.12
<i>Astyanax mexicanus</i>	Astyanax_mexicanus-2.0.	Sep 2017	94.2
<i>Pygocentrus nattereri</i>	Pygocentrus_nattereri-1.0.2	Jul 2016	94.102
<i>Esox lucius</i>	Eluc_V3	Jan 2017	94.3
<i>Gadus morhua</i>	gadMor1	Jan 2010	94.1
<i>Hippocampus comes</i>	H_comes_QL1_v1	Dec 2016	94.1
<i>Periophthalmus magnuspinnatus</i>	PM.fa	Dec 2014	94.1
<i>Monopterus albus</i>	M_albus_1.0	Jan 2017	94.1
<i>Mastacembelus armatus</i>	fMasArm1.1	Mar 2018	94.11
<i>Anabas testudineus</i>	fAnaTes1.1	Mar 2018	94.11
<i>Seriola dumerili</i>	Sdu_1.0	Aug 2017	94.1
<i>Seriola lalandi dorsalis</i>	Sedor1	Dec 2017	94.1
<i>Scophthalmus maximus</i>	ASM318616v1	Jun 2018	94.1
<i>Cynoglossus semilaevis</i>	Cse_v1.0	Jan 2014	94.1
<i>Amphilophus citrinellus</i>	Midas_v5	Sep 2014	94.5
<i>Oreochromis niloticus</i>	Orenil1.0	Jan 2011	94.1
<i>Neolamprologus brichardi</i>	NeoBri1.0	Dec 2011	94.1
<i>Haplochromis burtoni</i>	AstBur1.0	Dec 2011	94.1
<i>Astatotilapia calliptera</i>	fAstCal1.2	May 2018	94.12
<i>Maylandia zebra</i>	M_zebra_UMD2a	Apr 2018	94.2
<i>Pundamilia nyererei</i>	PunNye1.0	Dec 2011	94.1
<i>Acanthochromis polyacanthus</i>	ASM210954v1	Apr 2017	94.1
<i>Amphiprion ocellaris</i>	AmpOce1.0	Nov 2017	94.1
<i>Amphiprion percula</i>	Nemo_v1	Apr 2018	94.1
<i>Stegastes partitus</i>	Stegastes_partitus-1.0.2	May 2014	94.1
<i>Oryzias latipes</i>	ASM223471v1	Jul 2017	94.1

<i>Oryzias latipes</i>	ASM223469v1	Jul 2017	94.1
<i>Oryzias latipes</i>	ASM223467v1	Jul 2017	94.1
<i>Oryzias melastigma</i>	Om_v0.7.RACA	Feb. 2018	94.07
<i>Fundulus heteroclitus</i>	Fundulus_heteroclitus-3.0.2	Jan 2015	94.302
<i>Cyprinodon variegatus</i>	C_variegatus-1.0	Oct 2015	94.1
<i>Poecilia reticulata</i>	Guppy_female_1.0_MT	Apr 2014	94.1
<i>Poecilia formosa</i>	Poecilia_formosa-5.1.2	Oct 2013	94.512
<i>Poecilia latipinna</i>	P_latipinna-1.0	Nov 2015	94.1
<i>Poecilia mexicana</i>	P_mexicana-1.0	Nov 2015	94.1
<i>Xiphophorus maculatus</i>	X_maculatus-5.0-male	Dec 2017	94.5
<i>Xiphophorus couchianus</i>	Xiphophorus_couchianus-4.0.1	Nov 2015	94.401
<i>Gambusia affinis</i>	ASM309773v1	May 2018	94.1
<i>Kryptolebias marmoratus</i>	ASM164957v1	May 2016	94.1
<i>Takifugu rubripes</i>	FUGU5	Oct 2011	94.5
<i>Tetraodon nigroviridis</i>	TETRAODON 8.0	Mar 2007	94.8
<i>Mola mola</i>	ASM169857v1	Aug 2016	94.1
<i>Labrus bergylta</i>	BallGen_V1	May 2016	94.1
<i>Gasterosteus aculeatus</i>	BROAD S1	Feb 2006	94.1

Table S2. List of the 121 primer pair used to prepare intron libraries. The primer sequences are tiled with Illumina adapters. Gene, Primer ID, Primer sequences, Melting Temperature and Multiplex to which they belong are reported. The Locus Number is referred to a progressive number associated to each locus during the demultiplexing step.

Gene	Primer ID	Primer sequence	T _m °C	Multiplex	Locus N.
pou2f1b	cl_101301_pou2f1b_1f	5'-CTACACGACGCTCTCCGATCTTCAGATTCGAGTGGCCTTAGAAAAGAG-3'	54,7	1	41
	cl_101301_pou2f1b_1r	5'-CAGACGTGTGCTCTCCGATCTGGTCAGCGATCATGGTGATC-3'	55,9		
nr4a2a	cl_102952_nr4a2a_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGGACTCCAATCCCTCCATGTC-3'	57,6	4	42
	cl_102952_nr4a2a_1r	5'-CAGACGTGTGCTCTCCGATCTTGGTAGTCAGGGTTAGCCTG-3'	56,3		
rpl4	cl_10332_rpl4_1f	5'-CTACACGACGCTCTCCGATCTTCAGACTGGCCAGGGTGCTTTTG-3'	57,7	4	4
	cl_10332_rpl4_1r	5'-CAGACGTGTGCTCTCCGATCTAGGTTTTAGTGGGGCAAAC-3'	56		
spopla	cl_103786_spopla_1f	5'-CTACACGACGCTCTCCGATCTTCAGATCTGGCCCTAATGACAAAATGA-3'	54,3	1	43
	cl_103786_spopla_1r	5'-CAGACGTGTGCTCTCCGATCTTTCATCATCAAGTCCCTTTGGA-3'	54		
spopla	cl_103792_spopla_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGCTGGGCTGTAAAGATGGAA-3'	55,4	1	44
	cl_103792_spopla_1r	5'-CAGACGTGTGCTCTCCGATCTCAGCCTGCAGTCTCCATTAT-3'	54,8		
trio	cl_110155_trio_1f	5'-CTACACGACGCTCTCCGATCTTCAGATCCACACGCACGTCAAAGAG-3'	57,8	6	45
	cl_110155_trio_1r	5'-CAGACGTGTGCTCTCCGATCTCGTCCAGCAGCTCCTTCTG-3'	58,4		
med13a	cl_110983_med13a_1f	5'-CTACACGACGCTCTCCGATCTTCAGACCTCGACTCCAATCAGACC-3'	57,4	6	46
	cl_110983_med13a_1r	5'-CAGACGTGTGCTCTCCGATCTGGGCGTTGTACTGCTCCA-3'	57,8		
abcg4a	cl_112901_abcg4a_1f	5'-CTACACGACGCTCTCCGATCTTCAGACAGTGCCAAAGCTCTTTGAG-3'	56,7	5	47
	cl_112901_abcg4a_1r	5'-CAGACGTGTGCTCTCCGATCTCACTGACCCTGACTGAGGATG-3'	57,2		
sec24a	cl_115327_sec24a_2f	5'-CTACACGACGCTCTCCGATCTTCAGACTGTCATGAGGATACGCTGC-3'	55,8	3	48
	cl_115327_sec24a_2r	5'-CAGACGTGTGCTCTCCGATCTCGCAAAAGAAGTTTCCGTG-3'	55,3		
rpl19	cl_118300_rpl19_1f	5'-CTACACGACGCTCTCCGATCTTCAGACTGCGTCGCTACAGGGAG-3'	58,1	5	49
	cl_118300_rpl19_1r	5'-CAGACGTGTGCTCTCCGATCTCTCAGGTAGAGGCTGTGGTAC-3'	56,7		
unkl	cl_118929_unkl_1f	5'-CTACACGACGCTCTCCGATCTTCAGAACCGAGAAACCTACCCACTA-3'	54,8	2	50
	cl_118929_unkl_1r	5'-CAGACGTGTGCTCTCCGATCTGAACAACGGGCACTGCTC-3'	57,3		
adgrl1a	cl_119443_adgrl1a_1f	5'-CTACACGACGCTCTCCGATCTTCAGACAGCCATAGACTACAGGAGCT-3'	55,8	3	51
	cl_119443_adgrl1a_1r	5'-CAGACGTGTGCTCTCCGATCTTATCACTCGCAGCCAGC-3'	57,6		
gys1	cl_121169_gys1_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGATTGAGAGTTTGTGAGGGG-3'	55,7	2	52
	cl_121169_gys1_1r	5'-CAGACGTGTGCTCTCCGATCTACACTGTTCAGGTTGAAGTC-3'	55,1		
zmiz1a	cl_126635_zmiz1a_1f	5'-CTACACGACGCTCTCCGATCTTCAGACCTTCTGAGAACTGC-3'	56,8	5	53
	cl_126635_zmiz1a_1r	5'-CAGACGTGTGCTCTCCGATCTACGCTGCTGAAGTTCTCTT-3'	56,6		
kcma1a	cl_126690_kcma1a_1f	5'-CTACACGACGCTCTCCGATCTTCAGATTCCTCAAGGTCTCCATCTT-3'	55,6	1	54
	cl_126690_kcma1a_1r	5'-CAGACGTGTGCTCTCCGATCTGTTGACAGCCCTTAATCTGC-3'	54,7		
ebf3a	cl_126959_ebf3a_1f	5'-CTACACGACGCTCTCCGATCTTCAGACCTCAACGGATCTCTGCC-3'	58,1	6	55
	cl_126959_ebf3a_1r	5'-CAGACGTGTGCTCTCCGATCTGGCGAAGGCGCTCTTTTG-3'	57,9		
ank3a	cl_127174_ank3a_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGGGAAGGGAAAAACGCCAC-3'	57,6	2	56
	cl_127174_ank3a_1r	5'-CAGACGTGTGCTCTCCGATCTTCATCCGATTGACTTTGGTC-3'	55		
arid5b	cl_127539_arid5b_1f	5'-CTACACGACGCTCTCCGATCTTCAGATAACGTCGAGATGGAGCCG-3'	56,6	4	57

	cl_127539_arid5b_1r	5'-CAGACGTGTGCTCTTCCGATCTATGGTCCATGTAAACCGCAAG-3'	55,9		
kif5aa	cl_133622_kif5aa_1f	5'-CTACACGACGCTCTTCCGATCTTCAGACCTGCCAGCTCCTCATCTC-3'	57,9	6	58
	cl_133622_kif5aa_1r	5'-CAGACGTGTGCTCTTCCGATCTTGACTCCGTCAGAGAGCGG-3'	58,2		
kif5aa	cl_133627_kif5aa_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGTTCAAGACCTCACGTCGC-3'	56,4	5	59
	cl_133627_kif5aa_1r	5'-CAGACGTGTGCTCTTCCGATCTCCCCCGCTATCATCAGG-3'	58		
scn8a	cl_134627_scn8a_1f	5'-CTACACGACGCTCTTCCGATCTTCAGATCATGAGCGTCATTACCAACA-3'	54,2	1	60
	cl_134627_scn8a_1r	5'-CAGACGTGTGCTCTTCCGATCTTGGCGTGAGACTCCTCC-3'	57,1		
atp6ap1lb	cl_135436_atp6ap1lb_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAACACATCTGCTAACTGGCACAT-3'	56,7	5	61
	cl_135436_atp6ap1lb_1r	5'-CAGACGTGTGCTCTTCCGATCTACTGCACATTGAAGGCCTGG-3'	58,4		
pacs2	cl_137553_pacs2_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGTGGTGACCAAGGAGAAGAA-3'	54,7	1	62
	cl_137553_pacs2_1r	5'-CAGACGTGTGCTCTTCCGATCTCACCTCCTTGCTTTGCTCT-3'	54,9		
ube2d3	cl_137986_ube2d3_1f	5'-CTACACGACGCTCTTCCGATCTTCAGATTCTCAGATCACAGTGGTCTCC-3'	55,9	3	63
	cl_137986_ube2d3_1r	5'-CAGACGTGTGCTCTTCCGATCTGGAGTGAGCAAATGGAGAGAAG-3'	55,7		
gabrr1	cl_139850_gabrr1_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAACACCACCACCAAGCTGG-3'	58,4	4	64
	cl_139850_gabrr1_1r	5'-CAGACGTGTGCTCTTCCGATCTGGGTGAAGTGGATGTACAAACG-3'	56,3		
syncrpl	cl_139882_syncrpl_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAATTCAAGGACAGCGACCTCT-3'	55,9	4	65
	cl_139882_syncrpl_1r	5'-CAGACGTGTGCTCTTCCGATCTTCATCAGCCACAAAGAAAGG-3'	56		
syncrpl	cl_139885_syncrpl_1f	5'-CTACACGACGCTCTTCCGATCTTCAGATTTCTGCACTAAAGAAGCTGC-3'	53,8	1	66
	cl_139885_syncrpl_1r	5'-CAGACGTGTGCTCTTCCGATCTCGGTGCAATTTATTGTTGTTG-3'	53,4		
id2a	cl_139978_id2a_1f	5'-CTACACGACGCTCTTCCGATCTTCAGACACCTCAACACAGATATCAGC-3'	55,7	3	67
	cl_139978_id2a_1r	5'-CAGACGTGTGCTCTTCCGATCTCAGCTCTGATGGCAACTCC-3'	56,3		
tmem179	cl_140453_tmem179_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGCTGTTCTTTCTGTGCAAAGG-3'	55,3	3	68
	cl_140453_tmem179_1r	5'-CAGACGTGTGCTCTTCCGATCTTCAGCAGGTTGAGGAAGGC-3'	57,6		
zfyve1	cl_140944_zfyve1_1f	5'-CTACACGACGCTCTTCCGATCTTCAGATGGAACCCAACTCCCTCATC-3'	57,1	3	69
	cl_140944_zfyve1_1r	5'-CAGACGTGTGCTCTTCCGATCTCAGTGATGCTTGGTGTCGTT-3'	55,8		
dlx3b	cl_144842_dlx3b_1f	5'-CTACACGACGCTCTTCCGATCTTCAGACAGCTCGGCCCTCACTCAGAC-3'	60,1	6	70
	cl_144842_dlx3b_1r	5'-CAGACGTGTGCTCTTCCGATCTTGGATCTCCGGTTCTGGAACC-3'	58,9		
loxhd1b	cl_153319_loxhd1b_1f	5'-CTACACGACGCTCTTCCGATCTTCAGACGTAGGAGAGATCAAGAGGGT-3'	55,2	2	71
	cl_153319_loxhd1b_1r	5'-CAGACGTGTGCTCTTCCGATCTCAACCAGCAGCTCTCTGG-3'	56,3		
atp5fa1	cl_153330_atp5fa1_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGGCAGAAGAGATGGTGGAATT-3'	55,2	2	72
	cl_153330_atp5fa1_1r	5'-CAGACGTGTGCTCTTCCGATCTAGGCTCCAAGTTCAGAGACA-3'	55,6		
gli1	cl_156199_gli1_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGGAGAGAAGCCACACAAGTG-3'	56,1	4	73
	cl_156199_gli1_1r	5'-CAGACGTGTGCTCTTCCGATCTCCAGCGCAGAGTAGGC-3'	57,5		
tal1	cl_157901_tal1_1f	5'-CTACACGACGCTCTTCCGATCTTCAGACCTCTCGCCCATCAAC-3'	59,2	6	74
	cl_157901_tal1_1r	5'-CAGACGTGTGCTCTTCCGATCTTGCTGGGGTACATGCTGTAC-3'	57,4		
brinp3a2	cl_159363_brinp3a2_1f	5'-CTACACGACGCTCTTCCGATCTTCAGATGCTGGTTTCAGAGCCCTG-3'	57,7	5	75
	cl_159363_brinp3a2_1r	5'-CAGACGTGTGCTCTTCCGATCTCTGGAAGGATCACTTGCAAGC-3'	56,8		
meis2	cl_172049_meis2_1f	5'-CTACACGACGCTCTTCCGATCTTCAGATTATTGGAGCGCAGCCTGATG-3'	57,9	6	76
	cl_172049_meis2_1r	5'-CAGACGTGTGCTCTTCCGATCTGGACTCCGTCCATCCCACC-3'	60,2		
pabpc1b	cl_175437_pabpc1b_1f	5'-CTACACGACGCTCTTCCGATCTTCAGATGAATGGCATGCTGCTCAAT-3'	55,6	2	77
	cl_175437_pabpc1b_1r	5'-CAGACGTGTGCTCTTCCGATCTCTCTTTCGAGATTTGAAGCG-3'	54,9		
rfx3	cl_178477_rfx3_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGAATGCCATTCCACAGAGAATG-3'	54,6	1	78

	cl_178477_rfx3_1r	5'-CAGACGTGTGCTCTCCGATCTGTTTGCGCAAAGGCACTAA-3'	54,8		
cnot6b	cl_180314_cnot6b_1f	5'-CTACACGACGCTCTCCGATCTTCAGACAGAGCTCGGCAACATGGT-3'	58,1	5	79
	cl_180314_cnot6b_1r	5'-CAGACGTGTGCTCTCCGATCTAGAACCCGCAACTGGTTGTTA-3'	56,9		
rack1	cl_180389_rack1_1f	5'-CTACACGACGCTCTCCGATCTTCAGATGATTCTGTCGGCGTCCC-3'	57,4	5	80
	cl_180389_rack1_1r	5'-CAGACGTGTGCTCTCCGATCTAGTTGGTTTCATCACGGGTCA-3'	56,6		
msi2a	cl_180859_msi2a_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGAGACCCACGACAAAACG-3'	56,6	3	81
	cl_180859_msi2a_1r	5'-CAGACGTGTGCTCTCCGATCTTGGCAGCATCTGTAAACGTAAC-3'	55,6		
trpm5	cl_185831_trpm5_1f	5'-CTACACGACGCTCTCCGATCTTCAGAAGGAGGCCAAGTATGAGCAG-3'	56,8	4	82
	cl_185831_trpm5_1r	5'-CAGACGTGTGCTCTCCGATCTTCACTGTTGGAGTAACACTCTCC-3'	56,1		
ptpra	cl_188538_ptpra_1f	5'-CTACACGACGCTCTCCGATCTTCAGACATGGTGACCAATCTGAAGGAG-3'	55,8	3	83
	cl_188538_ptpra_1r	5'-CAGACGTGTGCTCTCCGATCTCTGGTCCGGCCAGTACTG-3'	58,1		
slc1a2	cl_202028_slc1a2_1f	5'-CTACACGACGCTCTCCGATCTTCAGAAGTTCAAAGTGGCATGAATGT-3'	53,5	1	84
	cl_202028_slc1a2_1r	5'-CAGACGTGTGCTCTCCGATCTCCATGCAGATGCCAAATGC-3'	56		
tle2a	cl_20720_tle2a_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGAAGCAAGAGCCCCATGG-3'	56,9	3	5
	cl_20720_tle2a_1r	5'-CAGACGTGTGCTCTCCGATCTAGGAGCGGATGTAGTTATCCC-3'	55,9		
stk11	cl_211520_stk11_1f	5'-CTACACGACGCTCTCCGATCTTCAGATTTCCAGTATTTCAAGCTCACG-3'	53,3	1	85
	cl_211520_stk11_1r	5'-CAGACGTGTGCTCTCCGATCTCTGTGCAAAATATTCAAGGCCA-3'	54,1		
rps5	cl_218030_rps5_1f	5'-CTACACGACGCTCTCCGATCTTCAGACCCCTCCGCAGAGTCAAC-3'	58,4	5	86
	cl_218030_rps5_1r	5'-CAGACGTGTGCTCTCCGATCTGAAAGCAGCTTCTCTGCTCC-3'	56,7		
rimbp2	cl_226487_rimbp2_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGAGGCTCTGAGTTGGGGAGG-3'	59,8	6	87
	cl_226487_rimbp2_1r	5'-CAGACGTGTGCTCTCCGATCTCGGCTGCTTTCGCTGTGATA-3'	58		
dla	cl_230665_dla_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGCAACGGAGGAAGCTGCAC-3'	59,9	6	88
	cl_230665_dla_1r	5'-CAGACGTGTGCTCTCCGATCTAAACCGTGAGGGCAAGTGCA-3'	60		
brsk1x	cl_232234_brsk1x_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGATTTCTGCCAGTCAATTCC-3'	54,4	1	89
	cl_232234_brsk1x_1r	5'-CAGACGTGTGCTCTCCGATCTTTCATCCAGGAGCAAGTTCTC-3'	54,7		
bzw1a	cl_24625_bzw1a_2f	5'-CTACACGACGCTCTCCGATCTTCAGATTTCAGAAGATCGTGGTGCT-3'	54,1	1	6
	cl_24625_bzw1a_2r	5'-CAGACGTGTGCTCTCCGATCTGTACCACTTCAGTATGGCCTC-3'	55,1		
kpnb3	cl_26173_kpnb3_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGACTGCCTCCATCAAGCC-3'	56,4	2	7
	cl_26173_kpnb3_1r	5'-CAGACGTGTGCTCTCCGATCTGATATGTTCTCCATGTCCTGGG-3'	55,1		
smap1	cl_28478_smap1_1f	5'-CTACACGACGCTCTCCGATCTTCAGACAAGTACTGCGCCGACTG-3'	56,5	4	8
	cl_28478_smap1_1r	5'-CAGACGTGTGCTCTCCGATCTCCAGATTCCAGGATGCC-3'	56,1		
olfm1b	cl_3192_olfm1b_1f	5'-CTACACGACGCTCTCCGATCTTCAGACCCGCACAAACAAGTGAAG-3'	57,3	5	1
	cl_3192_olfm1b_1r	5'-CAGACGTGTGCTCTCCGATCTTGGACTGGGTGATGTTCTGG-3'	57		
tmem178b	cl_36988_tm178b_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGCTGGGCTACTCTTCTCATGG-3'	57,8	6	9
	cl_36988_tm178b_1r	5'-CAGACGTGTGCTCTCCGATCTTCTGCCACACATGTGCAC-3'	58,6		
dhx38	cl_37567_dhx38_1f	5'-CTACACGACGCTCTCCGATCTTCAGACATCCACCAAGGCCATG-3'	57,1	3	10
	cl_37567_dhx38_1r	5'-CAGACGTGTGCTCTCCGATCTGCTGCACCATGATGTCCTT-3'	55,9		
dhx38	cl_37568_dhx38_1f	5'-CTACACGACGCTCTCCGATCTTCAGACAGGCTGCCAAGCTCAAG-3'	57,3	6	11
	cl_37568_dhx38_1r	5'-CAGACGTGTGCTCTCCGATCTATGGCATGCCTGTCTCAC-3'	58		
hoxa4	cl_381762_hoxa4_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGGAGCCGGTGGTCTATCC-3'	57,7	3	90
	cl_381762_hoxa4_1r	5'-CAGACGTGTGCTCTCCGATCTCCTCATTTGTAAGTGGCATTGA-3'	55,4		
azin1b	cl_385818_azin1b_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGGTCCATCTCTGACTTGTC-3'	55,1	2	91

	cl_385818_azin1b_1r	5'-CAGACGTGTGCTCTCCGATCTCACGACTGAGCCAATGGG-3'	56,6		
ankib1b	cl_387480_ankib1b_1f	5'-CTACACGACGCTCTCCGATCTTCAGACAAGAACCATGAGCACAGCT-3'	55,8	3	92
	cl_387480_ankib1b_1r	5'-CAGACGTGTGCTCTCCGATCTTGCTCCATCTTCTTTGGC-3'	55,3		
ablim1a	cl_389624_ablim1a_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGGATGTGGACAGAACCAGAC-3'	55,6	1	93
	cl_389624_ablim1a_1r	5'-CAGACGTGTGCTCTCCGATCTCCTGGATCTCCATTCCAAGA-3'	54,4		
lhx9	cl_395776_lhx9_1f	5'-CTACACGACGCTCTCCGATCTTCAGACAAGGACTTAAAGCAGCTGGC-3'	57,1	5	94
	cl_395776_lhx9_1r	5'-CAGACGTGTGCTCTCCGATCTGCTCTTGCGTTTTGGAACCA-3'	56,7		
iqsec1	cl_400572_iqsec1_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGAAGGTGGAGCGGCTCAT-3'	57,7	6	95
	cl_400572_iqsec1_1r	5'-CAGACGTGTGCTCTCCGATCTCGTGGGGTTACAGATGCAGT-3'	57,4		
iqsec1	cl_400576_iqsec1_1f	5'-CTACACGACGCTCTCCGATCTTCAGATCTCTGTATGGGATGCAAGT-3'	55,2	1	96
	cl_400576_iqsec1_1r	5'-CAGACGTGTGCTCTCCGATCTCGCTGATGTAAGTCTGATTCCA-3'	54,7		
brms1lb	cl_40881_brms1lb_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGTCTGAAGATGGAGACAGCTC-3'	55	2	12
	cl_40881_brms1lb_1r	5'-CAGACGTGTGCTCTCCGATCTCACTCCATCCTCCTCTC-3'	56,4		
mef2cb	cl_412427_mef2cb_1f	5'-CTACACGACGCTCTCCGATCTTCAGACAGAACATGCAGCATTACGC-3'	55,4	2	97
	cl_412427_mef2cb_1r	5'-CAGACGTGTGCTCTCCGATCTACCTGACATAAGTGAGAGCT-3'	55,1		
abcc4	cl_413653_abcc4_1f	5'-CTACACGACGCTCTCCGATCTTCAGACAGCACACAGATGAGGACCT-3'	56,9	5	98
	cl_413653_abcc4_1r	5'-CAGACGTGTGCTCTCCGATCTACCACGCGCTTCATCTGC-3'	58,3		
pde10a	cl_416134_pde10a_1f	5'-CTACACGACGCTCTCCGATCTTCAGATGCCTATGCAGACCCACGG-3'	59,9	6	99
	cl_416134_pde10a_1r	5'-CAGACGTGTGCTCTCCGATCTGTTCCGCGTGGTGTAGCC-3'	59,3		
pde10a	cl_416139_pde10a_1f	5'-CTACACGACGCTCTCCGATCTTCAGACCAACGACATCTATGCTGAGTT-3'	54,9	2	100
	cl_416139_pde10a_1r	5'-CAGACGTGTGCTCTCCGATCTTCATCATAGGGATCGGCTG-3'	56,3		
ccn1	cl_440349_ccn1_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGCTACTACTCGTGGCATCTG-3'	55,2	2	101
	cl_440349_ccn1_1r	5'-CAGACGTGTGCTCTCCGATCTGGTAGATCTGCTGTTGTACTC-3'	54,8		
plxnd1	cl_451609_plxnd1_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGCCATTCTCCAGTGGCC-3'	59,4	6	102
	cl_451609_plxnd1_1r	5'-CAGACGTGTGCTCTCCGATCTTGCTGCCWAGTGCAAACCAC-3'	58,1		
dlg3	cl_452435_dlg3_1f	5'-CTACACGACGCTCTCCGATCTTCAGACGAGCATCTTGCTGTGCAGAAC-3'	55,9	4	103
	cl_452435_dlg3_1r	5'-CAGACGTGTGCTCTCCGATCTCAGACACATCCAGAATGCAGTG-3'	56,4		
myo3b	cl_456719_myo3b_1f	5'-CTACACGACGCTCTCCGATCTTCAGAAGACAGCCTGTACTACAACCAG-3'	56,2	4	104
	cl_456719_myo3b_1r	5'-CAGACGTGTGCTCTCCGATCTTTCTCGGCTTCTCTTGCTC-3'	57		
uba52	cl_46164_uba52_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGCGACAAGATGATCTGCCG-3'	56,6	5	13
	cl_46164_uba52_1r	5'-CAGACGTGTGCTCTCCGATCTCTGCGGCAGTTGACAGC-3'	57,8		
ccz1	cl_479548_ccz1_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGGCCTTTGTGAAGCCATTGT-3'	56,8	2	105
	cl_479548_ccz1_1r	5'-CAGACGTGTGCTCTCCGATCTGGGATTAGCTGTTTTGTTGG-3'	55		
slit2	cl_480476_slit2_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGAGGTGCAGAAGAAGGAGTTT-3'	55	2	106
	cl_480476_slit2_1r	5'-CAGACGTGTGCTCTCCGATCTACTGCAGGAGGAAGATGATGA-3'	55,5		
etv4	cl_486174_etv4_2f	5'-CTACACGACGCTCTCCGATCTTCAGATGGATATTTGGACCAGCAAGTG-3'	55,4	3	107
	cl_486174_etv4_2r	5'-CAGACGTGTGCTCTCCGATCTGCCATTTCTTGCGACTT-3'	56,4		
cntn5	cl_48726_cntn5_1f	5'-CTACACGACGCTCTCCGATCTTCAGAAGAGATGACTGCCATACTGTGG-3'	56,3	4	14
	cl_48726_cntn5_1r	5'-CAGACGTGTGCTCTCCGATCTCACTCCACTCTGCTCGG-3'	58		
dyrk1ab	cl_49831_dyrk1ab_1f	5'-CTACACGACGCTCTCCGATCTTCAGACCGTACAGCGACCAGACA-3'	56,9	5	15
	cl_49831_dyrk1ab_1r	5'-CAGACGTGTGCTCTCCGATCTGAAAGCACTGGGGCATGTG-3'	57,6		
dyrk1ab	cl_49838_dyrk1ab_1f	5'-CTACACGACGCTCTCCGATCTTCAGACACCTCCAGCACCTCCTC-3'	57,9	5	16

	cl_49838_dyrk1ab_1r	5'-CAGACGTGTGCTCTCCGATCTATCTTGCTGTGCACTGGT-3'	56,6		
camk2n1a	cl_502173_camk2n1a_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGGACAGATCGGCAGAGCC-3'	58,4	5	108
	cl_502173_camk2n1a_1r	5'-CAGACGTGTGCTCTCCGATCTCCTCGTCTATTCGGTCGTCC-3'	57		
sh2b2	cl_50523_sh2b2_2f	5'-CTACACGACGCTCTCCGATCTTCAGATCCAGGACTGCATAGACCC-3'	56,2	4	17
	cl_50523_sh2b2_2r	5'-CAGACGTGTGCTCTCCGATCTCCATGAGGACATGACGCC-3'	56,3		
hnrpkl	cl_505778_hnrpkl_1f	5'-CTACACGACGCTCTCCGATCTTCAGAAGGTGCCAAGATAAAGGAGCT-3'	56,2	4	109
	cl_505778_hnrpkl_1r	5'-CAGACGTGTGCTCTCCGATCTCACTCTGGAACAACCTTGATGG-3'	56,2		
blmh	cl_52569_blmh_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGGGACATGCTGGTCAATCTC-3'	56,1	3	18
	cl_52569_blmh_1r	5'-CAGACGTGTGCTCTCCGATCTGCATTTCTTGGAATGACGCC-3'	55,7		
mctp1a	cl_532352_mctp1a_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGATGAAGTGGCCTCCTATGG-3'	55,4	1	110
	cl_532352_mctp1a_1r	5'-CAGACGTGTGCTCTCCGATCTAAAGGCACAGTCCAGTTAAATG-3'	53,8		
cul3	cl_548779_cul3_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGTCTAGAAAGAAGATGCAGCAC-3'	53,8	1	111
	cl_548779_cul3_1r	5'-CAGACGTGTGCTCTCCGATCTTCGCAACTGCTGTGTGAC-3'	56		
hat1	cl_561446_hat1_2f	5'-CTACACGACGCTCTCCGATCTTCAGAGACAAACCCGACACGTG-3'	57,1	5	112
	cl_561446_hat1_2r	5'-CAGACGTGTGCTCTCCGATCTGAGGCAGGATCAGCATTTGG-3'	56,7		
ppp3ca	cl_570602_ppp3ca_1f	5'-CTACACGACGCTCTCCGATCTTCAGATGGCCAAAATGTTCTCAGTTCT-3'	54,9	2	113
	cl_570602_ppp3ca_1r	5'-CAGACGTGTGCTCTCCGATCTGCCCTTAAGCGTCAACACAT-3'	56		
sppl3	cl_573257_sppl3_1f	5'-CTACACGACGCTCTCCGATCTTCAGATGCACTCTCATCGATACTTTG-3'	54,6	1	114
	cl_573257_sppl3_1r	5'-CAGACGTGTGCTCTCCGATCTGAGCAGCACGATGGATCC-3'	56,3		
kcnn2	cl_573276_kcnn2_1f	5'-CTACACGACGCTCTCCGATCTTCAGAACTGAACCTCATGGGGAGC-3'	56,7	3	115
	cl_573276_kcnn2_1r	5'-CAGACGTGTGCTCTCCGATCTGGCTTATCAGGCATTTTAGAGCT-3'	55,2		
kcnn2	cl_573278_kcnn2_1f	5'-CTACACGACGCTCTCCGATCTTCAGACTGCATGGACAGTGAGAGC-3'	56,1	3	116
	cl_573278_kcnn2_1r	5'-CAGACGTGTGCTCTCCGATCTAAGTTGCTGGTTATGTCCTGG-3'	55,3		
prrc2b	cl_573834_prrc2b_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGTATGGGCCCGACCAAG-3'	58,8	6	117
	cl_573834_prrc2b_1r	5'-CAGACGTGTGCTCTCCGATCTCTGCCACCACTTCCCTC-3'	58,6		
pcbp3	cl_588389_pcbp3_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGCAAAGGAGGCTCAAAGATCAA-3'	56	4	118
	cl_588389_pcbp3_1r	5'-CAGACGTGTGCTCTCCGATCTCTGAACCTGAGCGCCTGT-3'	57,8		
ano8b	cl_59965_ano8b_2f	5'-CTACACGACGCTCTCCGATCTTCAGACTGCTTTGTCTCTCTGTCATG-3'	56,4	4	19
	cl_59965_ano8b_2r	5'-CAGACGTGTGCTCTCCGATCTCCTTGATCCCCATCACAAATC-3'	56,1		
ano8b	cl_59970_ano8b_1f	5'-CTACACGACGCTCTCCGATCTTCAGAGAGCATCGGCCAGTGG-3'	58,7	6	20
	cl_59970_ano8b_1r	5'-CAGACGTGTGCTCTCCGATCTACTATGATGGCAATCAGGCC-3'	57,5		
grin1a	cl_6155_grin1a_1f	5'-CTACACGACGCTCTCCGATCTTCAGAAGAAGCACAACTATGAGAGTGC-3'	54,9	2	2
	cl_6155_grin1a_1r	5'-CAGACGTGTGCTCTCCGATCTCCAGATGAAAGCATGCAG-3'	55,7		
grin1a	cl_6157_grin1a_1f	5'-CTACACGACGCTCTCCGATCTTCAGAACACTCACCTTTGAAAACATGG-3'	54	1	3
	cl_6157_grin1a_1r	5'-CAGACGTGTGCTCTCCGATCTGAGGAAGATCCCTGCTGC-3'	55,9		
slc6a6	cl_616733_slc6a6_1f	5'-CTACACGACGCTCTCCGATCTTCAGATGTCACTCTTTACTGTGCATG-3'	55,1	2	119
	cl_616733_slc6a6_1r	5'-CAGACGTGTGCTCTCCGATCTCGTCAATGGCTTGATTTGACC-3'	54,9		
pax7a	cl_63278_pax7a_1f	5'-CTACACGACGCTCTCCGATCTTCAGACGACAGAAGCAGATTCCTT-3'	55,2	2	21
	cl_63278_pax7a_1r	5'-CAGACGTGTGCTCTCCGATCTCAAAACGCGGCTGATGGA-3'	56,6		
park7	cl_63289_park7_2f	5'-CTACACGACGCTCTCCGATCTTCAGATGAAGGAGAAGATGATGGCTG-3'	54,6	1	22
	cl_63289_park7_2r	5'-CAGACGTGTGCTCTCCGATCTCCATCCTTCTGTACTCGAGC-3'	55		
tenm2	cl_680052_tenm2_1f	5'-CTACACGACGCTCTCCGATCTTCAGAAACCATCATACCGGAGACA-3'	55,5	3	120

	cl_680052_tenm2_1r	5'-CAGACGTGTGCTCTTCCGATCTAGGCACCTCGATCTGCTC-3'	57		
tenm2	cl_680057_tenm2_1f	5'-CTACACGACGCTCTTCCGATCTTCAGACACCATGGAGCTCAG-3'	56,4	3	121
	cl_680057_tenm2_1r	5'-CAGACGTGTGCTCTTCCGATCTCTAAGTCTGTGGGCCACTC-3'	55,4		
tbx1	cl_68656_tbx1_1f	5'-CTACACGACGCTCTTCCGATCTTCAGATGATGGACTTCCTGCCTGT-3'	56	4	23
	cl_68656_tbx1_1r	5'-CAGACGTGTGCTCTTCCGATCTCGACGAGCTGTGGAAGC-3'	56,8		
ank1a	cl_70737_ank1a_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGGCCAATGTCAACGCTCA-3'	56,4	4	24
	cl_70737_ank1a_1r	5'-CAGACGTGTGCTCTTCCGATCTTGACCCATGTAGAGTGGA-3'	55,9		
rorb	cl_72477_rorb_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAAGAGGTCAAGATGTACCAGAGC-3'	56	4	25
	cl_72477_rorb_1r	5'-CAGACGTGTGCTCTTCCGATCTTGGGTGATCTGGATGGCAC-3'	57,5		
cpeb4a	cl_77594_cpeb4a_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGCCGATTTGTTCAACTGCA-3'	56,8	3	26
	cl_77594_cpeb4a_1r	5'-CAGACGTGTGCTCTTCCGATCTAGGACATATGGCTTCACTTCCA-3'	55,8		
nr3c1	cl_78021_nr3c1_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGCATGAAGTCTCTGCTGC-3'	57,3	6	27
	cl_78021_nr3c1_1r	5'-CAGACGTGTGCTCTTCCGATCTGCTCTTCAGGCCGTCCTT-3'	57,6		
gabrb2a	cl_78077_gabrb2a_1f	5'-CTACACGACGCTCTTCCGATCTTCAGACTGAGACCCGACCCCAAC-3'	58,2	3	28
	cl_78077_gabrb2a_1r	5'-CAGACGTGTGCTCTTCCGATCTTGGCATAAGAGCATCATCTCT-3'	55,5		
tlx3	cl_78713_tlx3_1f	5'-CTACACGACGCTCTTCCGATCTTCAGARGTCAAAACCTGGTTTCAGAAC-3'	54,2	1	29
	cl_78713_tlx3_1r	5'-CAGACGTGTGCTCTTCCGATCTCCTCTCCTCGGCTGTCTG-3'	57,4		
cryba1l2	cl_78754_cryba1l2_1f	5'-CTACACGACGCTCTTCCGATCTTCAGACATCAGGGTGGAGAGCGG-3'	58,2	6	30
	cl_78754_cryba1l2_1r	5'-CAGACGTGTGCTCTTCCGATCTCTGCTGGCCCTGGAAGTC-3'	58,9		
megf8	cl_84965_megf8_2f	5'-CTACACGACGCTCTTCCGATCTTCAGAATTTGGCTGGTGTTCAGAA-3'	55,1	2	31
	cl_84965_megf8_2r	5'-CAGACGTGTGCTCTTCCGATCTTGAGATCTGGTCCACACGG-3'	56,5		
eya1	cl_89045_eya1_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAAGTCTCTCCACAGATTTACCC-3'	54	1	32
	cl_89045_eya1_1r	5'-CAGACGTGTGCTCTTCCGATCTGGGGTAGGAAGAATATGCGG-3'	55,4		
eya1	cl_89049_eya1_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAAACAACCCATCACGCCG-3'	58,8	5	33
	cl_89049_eya1_1r	5'-CAGACGTGTGCTCTTCCGATCTGTTTCATCCAGGTCCCAGATGA-3'	56,7		
eya1	cl_89053_eya1_1f	5'-CTACACGACGCTCTTCCGATCTTCAGATGGCACTGAAAGCACTAACAT-3'	54,9	2	34
	cl_89053_eya1_1r	5'-CAGACGTGTGCTCTTCCGATCTGTGGTCACAGGATATTCACAC-3'	55,7		
smarcd3b	cl_90274_smarcd3b_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAAGGACTGGCTCAAATCCCA-3'	56,2	4	35
	cl_90274_smarcd3b_1r	5'-CAGACGTGTGCTCTTCCGATCTCCACACGCTGTGCAT-3'	57,3		
zfyve9a	cl_93181_zfyve9a_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGTACGGGAGGCCACATCA-3'	57	5	36
	cl_93181_zfyve9a_1r	5'-CAGACGTGTGCTCTTCCGATCTGAGGACGATCTCCGACGG-3'	57,2		
acvr2bb	cl_93416_acvr2bb_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGTGTGTGGGAGCTGGTG-3'	59,3	6	37
	cl_93416_acvr2bb_1r	5'-CAGACGTGTGCTCTTCCGATCTCAGCATGTACTGCCAACTGT-3'	57,7		
c1ql3b	cl_94257_c1ql3b_1f	5'-CTACACGACGCTCTTCCGATCTTCAGAGGGCTGACCTCTGTAAAAACA-3'	55,3	3	38
	cl_94257_c1ql3b_1r	5'-CAGACGTGTGCTCTTCCGATCTGTAGTTCTGGTCGGCGTC-3'	55,6		
cacna1eb	cl_94647_cacna1eb_1f	5'-CTACACGACGCTCTTCCGATCTTCAGATTTCCAGCTGCCATCATGAC-3'	55,9	4	39
	cl_94647_cacna1eb_1r	5'-CAGACGTGTGCTCTTCCGATCTATCACCTCGTTCCAGTCTCT-3'	56,3		
eif4a2	cl_96869_eif4a2_1f	5'-CTACACGACGCTCTTCCGATCTTCAGACACAGAGAGCTGGCTCAG-3'	57,9	6	40
	cl_96869_eif4a2_1r	5'-CAGACGTGTGCTCTTCCGATCTCATGTAGTACCCAGAGCCAG-3'	57,5		

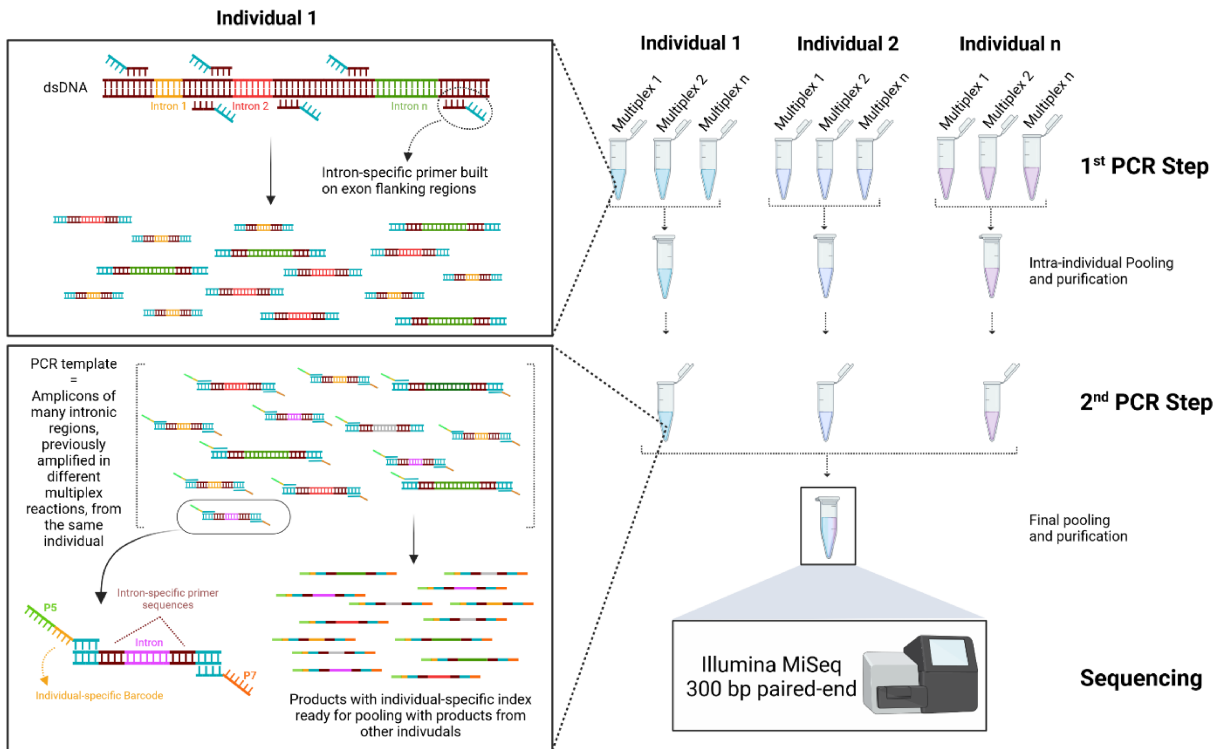


FIGURE S1
Schematic representation of intron library preparation

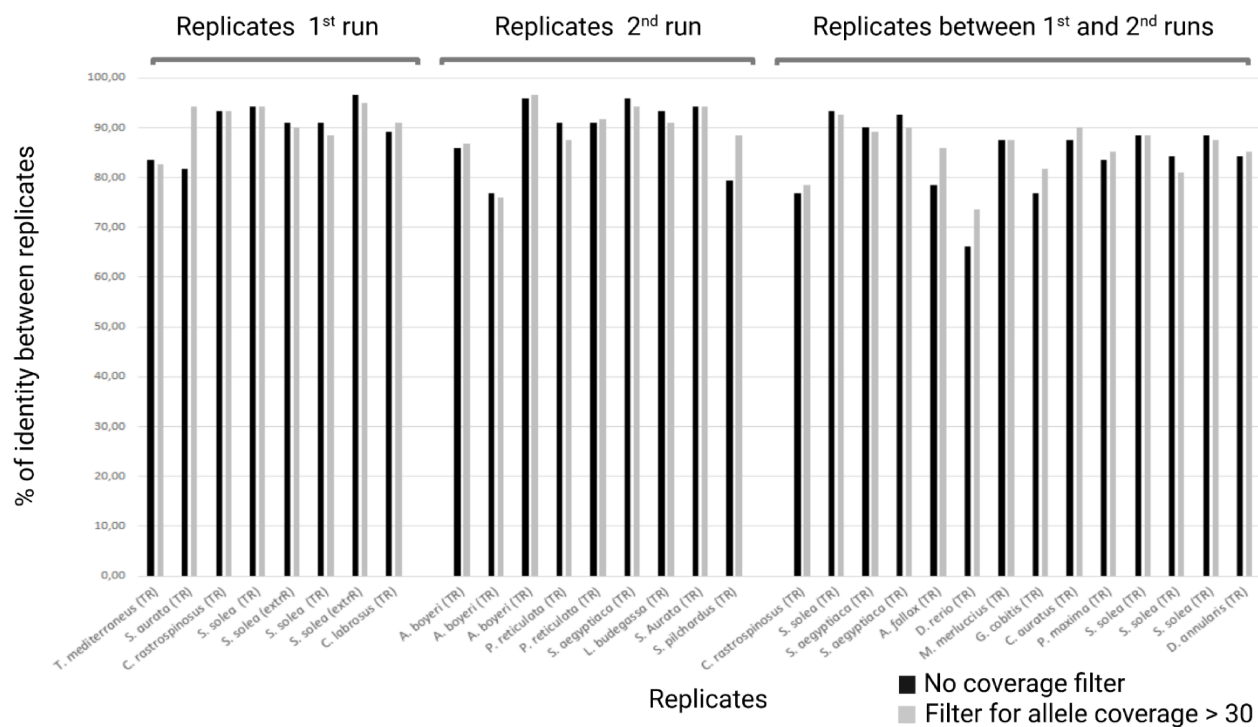


FIGURE S2

Comparisons among replicates to verify the replicability of the approach. Percentages of loci showing identical genotypes among individual replicates within and between MiSeq runs (1st and 2nd). TR = Technical Replicates, extrR = extraction Replicates. The black bars show the percentage of identity between replicates considering the allele count obtained by the informatics procedure, while the grey bars represent the percentage of identity further filter the dataset for an allele coverage over 30 assembled reads.

Table S3. Descriptive statistics calculated at 54 loci within icefish species and on the total dataset without the 11 individuals with dubious identification (N=69). The individual originally classified as *C. hamatus* but resulted as a mislabelling of the sample tube has been removed from this calculation. For each species, samples size (N), proportion of polymorphic loci (P_N), missing data percentage across individuals per species (MVs), unbiased expected heterozygosity (H_E), observed heterozygosity (H_O), inbreeding coefficient (F_{IS}), mean number of alleles per locus (N_a), Allelic richness (A_r) and private Allelic richness (pA_r) are reported. A_r and pA_r values are calculated on a minimum sample size of six diploid individuals. Some inbreeding coefficients were negative reflecting an apparent excess of heterozygotes intraspecies (Table 2).

Out of 424 probability tests for HWE, 41 showed a significant departure from HWE, with a nominal threshold of 0.05. However, it is noteworthy that 90 tests failed because loci were monomorphic within the species sample. Loci showing HW disequilibrium are different across species. After correction for multiple tests, only 5 probability tests for HWE are statistically significant (loci 7 and 71 in *C. hamatus*, locus 24 in *C. myersi* and loci 7 and 31 in *C. rastrispinosus*).

Species	N	P_N (%)	MVs (%)	H_E	H_O	F_{IS}	N_a	A_r	pA_r
<i>C. hamatus</i>	11	88.68	11.49	0.425	0.353	0.168	4.189	3.320	1.140
<i>C. myersi</i>	12	86.79	9.74	0.425	0.429	- 0.011	4.359	3.380	1.020
<i>C. rastrispinosus</i>	14	88.68	5.12	0.428	0.409	0.043	4.566	3.450	1.140
<i>P. macropterus</i>	6	69.81	11.32	0.316	0.337	- 0.067	2.674	5.490	1.460
<i>P. maculatus</i>	6	50.94	9.11	0.201	0.213	- 0.058	1.755	4.640	0.870
<i>N. ionah</i>	6	79.25	9.11	0.376	0.390	- 0.038	2.830	2.830	1.980
<i>C. wilsoni</i>	8	84.91	2.12	0.441	0.424	0.039	4.509	3.880	2.210
<i>A. mitopteryx</i>	6	81.13	8.80	0.448	0.500	- 0.116	4.000	9.060	3.490

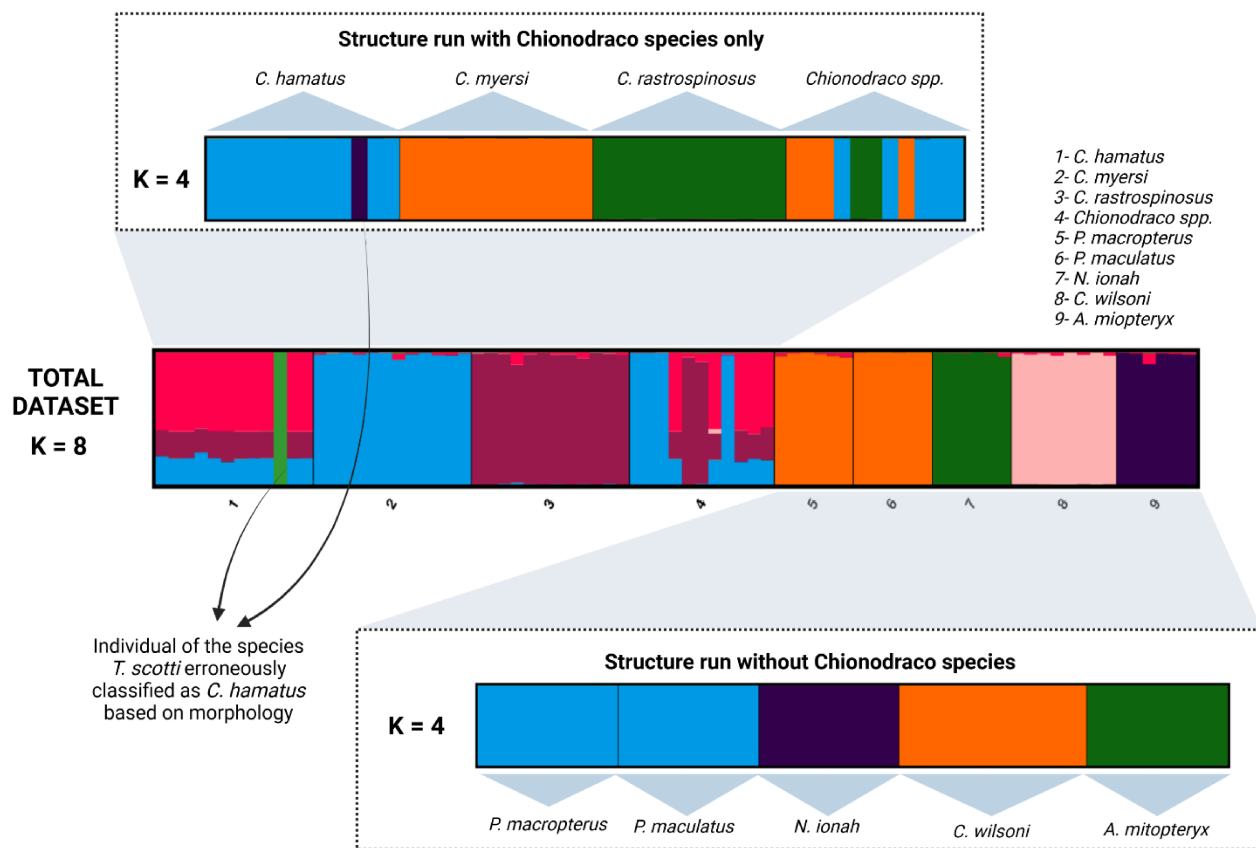


FIGURE S3

Structure bar plots based on the total dataset of icefish and on two datasets including the species of the genus *Chionodraco* and the remaining species, respectively. For each run, the best K identified by the Evanno method is shown.

Table S4. Descriptive statistics calculated at 43 loci within the genus *Solea*. Statistics were calculated per species based on the total dataset (of 68 and 38 individuals for *Solea solea* and *Solea aegyptiaca*, respectively) and per populations. Individuals processed in replicate were considered only once. The three outlier individuals, morphologically misclassified as *S. solea* and resulted instead *S. aegyptiaca* based on genetic identification, were considered as *S. aegyptiaca* for the total dataset but were a priori removed for the analysis population-based. For each samples, size (N), proportion of polymorphic loci (P_N), missing data percentage across individuals per species/populations (MV_s), unbiased expected heterozygosity (H_E), observed heterozygosity (H_o), inbreeding coefficient (F_{IS}), mean number of alleles per locus (N_a), Allelic richness (A_r) and private Allelic richness (pA_r) are reported. A_r and pA_r values are calculated on a minimum sample size of 12 and 5 diploid individuals respectively for *S. solea* and *S. aegyptiaca*.

Species	N	P_N (%)	MV_s (%)	H_E	H_o	F_{IS}	N_a	A_r	pA_r
<i>Solea solea</i>	68	100	2.599	0.592	0.493	0.167	14.14	10.91	10.82
<i>Solea aegyptiaca</i>	38	93.02	1.101	0.364	0.260	0.286	4.32	4.32	4.24
Populations of <i>Solea solea</i>	N	P_N (%)	MV_s (%)	H_E	H_o	F_{IS}	N_a	A_r	pA_r
SS-GLI	14	95.3	2.66	0.604	0.551	0.089	6.19	6.75	2.21
SS-TYR	15	97.7	5.12	0.613	0.567	0.074	7.40	6.78	2.38
SS-ADR	14	97.7	1.83	0.565	0.512	0.094	5.72	5.41	1.06
SS-GRE	12	95.4	1.36	0.493	0.449	0.089	4.51	4.51	1.05
SS-TUR	13	93.0	1.61	0.436	0.371	0.149	3.86	3.76	0.79
Populations of <i>Solea aegyptiaca</i>	N	P_N (%)	MV_s (%)	H_E	H_o	F_{IS}	N_a	A_r	pA_r
SA-TYR	13	48.8	1.97	0.168	0.153	0.090	1.72	1.54	0.15
SA-ADR	5	62.8	0.93	0.287	0.321	- 0.118	2.09	2.09	0.42
SA-GRE	6	72.1	0	0.364	0.407	- 0.118	2.67	2.54	0.78
SA-TUR	6	53.5	0	0.220	0.248	- 0.126	1.76	1.72	0.21
SA-ALE	5	58.1	0.47	0.258	0.287	- 0.115	2.00	2.00	0.24

ANNEX E

Identification and tracking of sturgeons and paddlefish products in trade: implications for trade control and biodiversity management

Leonardo Congiu^{a,b}, Beate Striebel-Greiter^c, Joern Gessner^d, Elisa Boscari^a, Markus Boner^e, Jutta Jahrl^f, Stefano Dalle Palle^a, Arne Ludwig^{g,h}

Corresponding author: Leonardo Congiu

e-mail: leonardo.congiu@unipd.it

Affiliations:

a University of Padova, Department of Biology, Via Ugo Bassi 58b, Padova, Italy

b National Biodiversity Future Center, Palermo, Italy.

c WWF Central and Eastern Europe, Vienna, Austria.

d Leibniz-Institute for Freshwater Ecology and Inland Fisheries, Berlin, Germany

e Agroisolab GmbH, Jülich, Germany.

f WWF Austria, Vienna, Austria.

g Leibniz Inst Zoo & Wildlife Research, Department Evolutionary Genetics, D-10315 Berlin, Germany.

h Humboldt University, Faculty of Life Science, Albrecht Daniel Thaer-Institute, D-10099 Berlin, Germany.

ABSTRACT

Despite strict protection of natural populations and CITES restrictions on international trade of caviar, sturgeons are increasingly at risk of extinction. The latest assessment released in July 2022 by the IUCN confirms sturgeons and paddlefish as the most endangered group of species on the planet. In this context, poaching and the illegal trade of caviar and meat have a significant weight and it is, therefore, necessary to support control activities by identifying diagnostic and reliable methods for the identification of species, interspecific hybrids, aquaculture or natural origin and geographical provenance of commercial samples. The technical approaches used for this purpose now include biochemical, isotopic, and genetic marker analyses, and are constantly evolving with remarkable developments proposed in recent years. Among these, particularly relevant are the numerous nuclear genetic markers, which allow the identification not only of pure sturgeon species but also of interspecific hybrids and the approaches based on the analysis of fatty acids which allow to distinguish feed-fed animals from fed with natural food. Additionally, the use of multi-isotopic fingerprint analyzes to determine the geographical area of origin is rapidly developing and is proving to be extremely useful for commercial controls and for the protection of natural sturgeon populations. This synthesis arises from the need of providing an exhaustive picture of the available technical approaches and to evaluate their potential and limits, focusing mainly on the innovations proposed in the last 10 years which have significantly improved our diagnostic power.

Keywords: sturgeon, caviar, species identification, traceability, poaching, smuggling

1 INTRODUCTION

Sturgeons and paddlefish (Acipenseriformes) are world's most endangered group of species - based upon IUCN's RED List Assessment (IUCN, 2022; Congiu et al., 2023) – (<https://www.iucn.org/content/sturgeon-more-critically-endangered-any-other-group-species>). Besides habitat degradation through damming, channelization, gravel extraction etc. and pollution, a major factor contributing to this status is a long-lasting overharvest for production of one of the most prestigious and valuable food of wildlife origin, the caviar, obtained by processing their eggs.

In response to the dramatic decline of wild populations, listing under the Convention on International Trade in Endangered Species (CITES) was used to increase transparency and sustainability of trade in sturgeon products. This process started with two species (*Acipenser brevirostrum* and *A. sturio*) being listed on Appendix I in 1989, followed by the remaining 25 species in 1998 (<https://cites.org/eng/app/appendices.php>) listed under Appendix II. As such the control and supervision of international trade was mandatory for all sturgeon species.

In order to support national and international trade controls, in 2000 CITES introduced a labeling system for both wild and aquaculture caviar, which uses universal codes and non-removable labels (Resolution CITES Conf 12.7, Rev CoP17). Although recommended also for domestic markets several states with extensive domestic markets, such as China, Japan, Russia, and the United States, have not adopted a mandatory labelling system. This renders the tracking of caviar in trade impossible and exposes domestic markets to an uncontrolled distribution of caviar from illegal sources (whitewashing).

In the same resolution, CITES also recommends range States with shared stocks to establish annual quotas for exporting caviar and meat from wild caught sturgeons. However, since 2011, no quota has been communicated, making any trade in meat or caviar taken from shared stocks illegal. This

holds true for the Caspian Sea, the Danube basin, the Black and Azov seas, the Amur/Heilongjiang River. In addition, many states have independently enacted national or regional bans on fishing and trading in wild sturgeons.

In parallel with the dramatic decline of wild stocks, and driven by the significant market demand for caviar, a rapid development of commercial sturgeon aquaculture producing caviar took place profoundly changing global caviar production and trade dynamics, shifting the import ratio of farmed *versus* wild caviar internationally from 5% aquaculture source, in 2000, to 95% in 2015 (Bronzi et al., 2019). However, the 5% amount of caviar from wild fish seems to be significantly underestimated because the volume of illegal trade is unknown.

According to a report released by TRAFFIC in 2018 (30th meeting of the CITES Animal Committee, AC30 Inf. 33), there are different types of illegal trade. One comprises the "black" market in which wild products are sold "under the counter", through individual contacts or online offers, and the official market in which illegal caviar might be commercialized under 'false labelling' taking advantage of the absence of conformity checks between the actually sold product and what is declared on the label. Hence, wild caviar can be labeled as aquaculture product and thus enter the legal trade illegally. In some cases, paradoxically, farmed caviar is deliberately sold as "wild origin" on the black market to intentionally overprice the product. It is also possible that forged or counterfeit CITES documents, or genuine CITES documents obtained through corruption, are used for export.

Despite CITES restrictions and the high availability of aquaculture products, poaching and illegal trade in wild meat and caviar remain serious threats to the survival of wild populations, as documented in many regions (Birstein et al., 1998; Cohen, 1997; Daea, 2019; Doukakis et al., 2012; Fain et al., 2013; Hruby, 2019; Jahrl et al., 2021; Knight, 2017;

Luca et al., 2020; Ludwig et al., 2015; Pappalardo et al., 2019; Savić et al., 2018; Van Uhm & Siegel, 2016; Zabyelina, 2014). A recent innovative study based on the combination of both genetic and isotopic analyses on 145 commercial products sampled in Bulgaria, Romania, Serbia and Ukraine, demonstrated that a substantial fraction (19% of the total) was of wild and thus illegal origin in these four countries (Jahrl et al., 2021). In addition, 17% of the total samples was sold without the mandatory CITES labels, with false codes or without the required CITES permits. This highlights the urgent need for a greater global effort to control whether sturgeon products in trade comply with all legality criteria.

Currently, most control efforts by law enforcement officials are limited to verifying the presence of proper CITES documentation and the integrity of the label. Conversely, less importance is given to compliance checks between the commercialized products and the declarations made on their labels. These checks should verify the species and the origin of the commercial product, making use of the numerous traceability markers already available and increasing constantly.

Since the publication of the most complete summary of sturgeon forensic markers (Ludwig 2008), novel methods and approaches in genetic analysis have been developed which allow to improve the efficacy in species identification as well as in detection of hybrids. Additionally, the assessments of origin and provenance of samples through analysis of the isotope or chemical compositions have recently undergone rapid development.

In line with the requirements of CITES Resolution Conf. 12.7 (Rev.CoP17), this review examines available techniques for identification of Acipenseriformes species and their products in trade considering their advantages and disadvantages.

2 IDENTIFICATION OF STURGEON SPECIES, HYBRIDS, AND THEIR PRODUCTS

2.1 Morphological determination

Morphological and morphometric approaches, supported by dichotomous identification keys, are usually the first steps to identify intact specimens. One of the most comprehensive guides for the morphological identification of pure species is the CITES identification guide “Sturgeons and Paddlefish 2001” (ECC, Canada 2001). However, correct identification based on dichotomous keys is often complicated by the fact that the geographical distribution area of the species is one of the criteria that must be taken into account. The various sturgeon species raised in aquaculture for production purposes have been globally redistributed and the information of geographical origin no longer makes any sense.

Another problem of morphological and meristic characteristics is the overlap of the ranges of values between species. Often the differences are valid only on a statistical basis, making the identifications meaningful only if carried out on many animals. Moreover, while this approach can be successfully used on live or dead animals if they are intact, this is not the case with carcasses where significant parts are often damaged or missing, or with sturgeons offered for sale in parts, often even without skin. In these cases, it is necessary to proceed with identification through a genetic approach.

Another source of ambiguity in using a morphological approach is the allometric growth. During ontogenetic development, which includes the larval, fry, juvenile and sub-adult phases, the proportions of the body undergo significant changes which are not systematically described in any sturgeon species identification key. In some species, also morphological characters can be modified, as for example in Beluga (*Huso huso*) which reduces its scutes during the sexual maturation. After attaining sexual

maturity, which can be reached after 8-20 years depending on the species and environmental conditions, the morphological and meristic characteristics tend to be consistent.

The effectiveness of the morphological approach collapses when the animals to be identified are hybrids of unknown origin. Many interspecific hybrids between sturgeon species are in fact produced in aquaculture and possess intermediate morphological characteristics compared to the parental species. Furthermore, sturgeon hybrids are often fertile and can be crossed with third species, with other hybrids or backcrossed with parental species for multiple generations, potentially generating a multitude of morphologically indistinguishable intermediate forms.

Regarding both eggs and caviar (consisting of salted – traditionally unfertilized – eggs), the precise identification of species on a morphological basis is not reliable due to the similarity between the eggs of different species and the unpredictable intraspecific variability of size and color, which are affected by the degree of maturation, diet and individual genetic traits. The possibility to use the number of micropyles for identification (Debus et al., 2002) have not been followed up after the initial attempts. Again, the increasing number of species and interspecific hybrids used to produce caviar is decreasing the quality of diagnostic characters. If historically the eggs of the main three caviar producing species (*H. huso*, *A. gueldenstaedtii* and *A. stellatus*) were quite recognizable by experts, today, the contribution of genetic analyses is necessary.

All the by-products of sturgeon processing such as oils, isinglass, swim bladders, and other derived tissues or body parts are not morphologically identifiable, the only exception being the skin which, in some cases, can retain diagnostic characteristics. The morphometric approach can sometimes allow to distinguish animals of wild origin from animals raised in aquaculture thanks to typical

malformations that are observed in reared individuals such as deformed pectoral or anal fins, shortened opercula, missing nostril (Gessner & Friedrich, 2017) or eye abnormalities. However, these malformations do not constitute decisive evidence of the origin of the animals as they may not exist in farmed animals, or they may be observed in animals released within restocking programs after a period of intensive farming.

In any case, any type of morphological sturgeon species identification must be carried out by skilled experts. For a layman, the risk of misinterpreting the indications of the identification guides is high and can lead to incorrect results. This is very relevant for example in the case of animals accidentally caught by sport fishermen who should be able to distinguish native species of a given basin - to release them immediately - from alien species escaped from aquaculture facilities or released illegally - which should be removed. In these cases, it would be important to have the support of experienced personnel.

2.2 Genetic identification

Most limits concerning morphological and morphometric identification discussed above are overcome by DNA-based methods which, in recent decades, have become the approach of choice in the identification of animal and plant species of commercial interest. The analyses of different types of molecular markers as instruments to detect intra- and interspecific genetic variability became powerful approaches to elucidate species evolutionary and demographic histories, population differentiation as well as to search for diagnostic characters which allow to identify species, interspecific hybrids, populations, genetic lines, paternities or even individuals. In this chapter, the current status of the availability and application of DNA markers for sturgeon species is presented, highlighting their potential and limits.

2.2.1 Mitochondrial markers

Mitochondrial markers are among the most frequently used genetic tools and often the prime standard to identify species (Ahmed et al., 2018; Cooper et al., 2009; Dalton & Kotze, 2011; Dawnay et al., 2007; Khedkar et al., 2014) and this is true also for sturgeons. This is mainly due to mitochondrial DNA (mtDNA) properties such as: i) number of copies hundreds or thousands of times greater than the nuclear genome, ii) small size and consequent high resistance to degradation, iii) presence of regions with different mutation rates and therefore with different levels of sequence heterogeneity, vi) ease of analyses, reliability and comparability with a rich reference database (e.g. Genbank, BOLD). Last but not least, this method is generally accepted as evidence during court cases. It is generally used by sequence analysis although several indirect approaches have been proposed in the past to limit costs and processing time (sequencing). For a review of these methods such as RFLP, SSCP nested or semi-nested PCR, see Ludwig (2008).

More recently, after sequencing costs dropped substantially, the mtDNA sequencing analysis became the most common technique for species identification through the comparison of results with sequences deposited in publicly available data banks. However, the use of the reference sequences present in the databases should be treated with caution as some individuals have been identified incorrectly and consequently their deposited sequences can create ambiguities. This is the case of a mitogenome (EU719645) deposited as *A. sinensis* but belonging to *A. gueldenstaedtii* (Dillman et al., 2014), highlighting the importance of vouchers and certified reference samples.

Mitochondrial DNA, for sturgeon in particular the sequencing of the control region, represents an excellent tool for the identification of pure species. Exceptions are sturgeon species that are not clearly differentiated or have encountered introgressions in the past. This is the case for

the Persian sturgeon (*A. persicus*), morphologically very similar to the Russian sturgeon (*A. gueldenstaedtii*) and with overlapping distribution. The two species are not distinguishable based on mitochondrial DNA and in general no diagnostic genetic markers are available, raising doubts about the status of *A. persicus* as a valid species. Although separated geographically, the mitochondrial haplotypes of the Adriatic sturgeon (*A. naccarii*) do not form a separate cluster from haplotypes of *A. gueldenstaedtii* (Birstein et al., 2000). However, the species is distinguishable genetically because all existing individuals descend from a group of animals bred in captivity that are characterized genetically. For these species seven mtDNA control region haplotypes are known and these haplotypes are not shared with other species (Boscari et al., 2014a). Another typical ambiguity that often creates disputes in the case of analysis of commercial products is given by the introgression of the Siberian sturgeon (*A. baerii*) into Russian sturgeon (*A. gueldenstaedtii*) gene pool, followed by backcrosses with the Russian sturgeon (Jennekens et al., 2000; Mugue et al., 2008). Due to the non-recombinant inheritance of mtDNA along the matrilineal genealogy, over 30% of Russian sturgeons still has a *A. baerii* mitochondrial haplotypes that are defined as "*baerii-like*". These *baerii-like* haplotypes, based on a simple query of the databases are attributed to the Siberian sturgeon even if the animal, after several generations of backcrosses, has completely lost any morphological trace of the hybridization event. The only suitable marker for the identification of *baerii-like* haplotypes is a SNP proposed by Mugue et al. (2008).

Due to its maternal inheritance, the main limit of mtDNA is the inability to track the paternal contribution, making it scarcely informative for relatedness analyses and detection of hybrids which is a critical limit for sturgeons. In fact, with the considerable increase in the production of interspecific hybrids for caviar or meat, the use of mitochondrial markers for

commercial controls opens the way to a series of possible commercial frauds, the most common of which is to sell the hybrid caviar under the false label of the maternal species increasing its market value. In this case, mitochondrial markers would erroneously confirm the correctness of the label. Alternatively, pure-bred species wild caviar (illegal) is sold as being of captive hybrid origin (whitewashing). For this reason, the development of biparentally inherited nuclear markers has been particularly pursued in recent years.

2.2.2 Nuclear markers

With the exception of the pioneering studies that had involved multilocus nuclear markers in the past (extensively reviewed in Ludwig et al. 2008 but scarcely used nowadays), the use of nuclear markers to detect species and hybrids has caught on only lately. In the last 10 years, the focus of the development has been upon two classes of nuclear markers, the microsatellites and the Single Nucleotide Polymorphisms (SNPs), used individually or in combination through different approaches (summarized in Table 1). For both classes of markers, loci can be informative if they have species-specific alleles, either fixed or with a frequency high enough to guarantee a good probability of detection. This condition is less common in microsatellites than in SNPs, probably because of their higher rate of variability and the high number of alleles resulting thereof. Furthermore, the published loci have generally been selected for high variability. Conversely, SNPs are usually biallelic and, especially those isolated with genotyping by sequencing approaches, can map in regions with different variability rates increasing the likelihood of having loci with fixed or high frequency species-specific alleles.

Nuclear molecular markers, being equally inherited by the parental individuals, are informative for all applications in which the genetic contribution of both parents is of relevance, such as relatedness analyses or

identification of intra- and interspecific hybrids. The latter case is especially important for sturgeons for which artificial hybridization is recurrent in aquaculture and for the confirmed occurrence of various events of natural introgression (Beridze et al., 2022; Ludwig et al., 2009). Since the fertility of many hybrids can generate genomes with various degrees of admixed contributions through backcross with parental species or further hybridizations with other species, the availability of one diagnostic marker for each species may not be sufficient. In fact, in first generation hybrids the entire chromosomal set of the two parental species is fully represented, and even a single locus marker can successfully detect their target species. From the second generation onward, only a subset of these chromosomes (or part of them, being everything further complicated by meiotic recombination) will pass the random bottleneck of segregation (Figure 1). For this reason, the availability of several loci increases the possibility that at least one of them can be still detectable in the sample. For some species, the occurrence of past events of natural introgression renders the identification of loci with 100% diagnostic power difficult if not impossible. For example, due to the past introgression documented for *A. baerii* and *A. gueldenstaedtii*, as mentioned above, no nuclear loci with alternative alleles in the two species are yet available. This is also the case of *A. schrenckii* and *Huso dauricus* hybrids (sympatric in the Amur River) (Boscari et al., 2017a).

2.2.2.1 Microsatellites

Microsatellites, given their high variability, the widespread occurrence and the easy analyses by PCR, have for decades been the most common markers for the study of genetic differentiation at different levels of diversity (relatedness and population analyses) in many eukaryotes including sturgeons and paddlefish (Ludwig et al. 2008).

The presence of a high number of alleles per locus (for example compared to SNPs) makes

them very informative for the study of genetic diversity, potentially including species identification. However, these markers also have some limitations for species identification, such as the limited comparability of results among laboratories due to different analytical conditions. Small allele shifting may lead to discordant genotyping of the same individual analyzed in different laboratories. To have a reliable allele size, all laboratories applying this approach should include some reference samples with known genotypes in their sample set for the opportune inter-calibration. Other problems such as misidentification due to allelic dropout, null alleles or wrong allele dosage estimation can also occur but they are locus-dependent and require preliminary validation on large sample sets.

For the application of microsatellites in species identification, the presence of alleles fixed in a given species and absent from all the others (private) is a very rare condition, also due to the high intrinsic variability rate of microsatellites. However, in some cases, it was possible to identify efficient markers, such as for the stellate sturgeon (*A. stellatus*) in which at the locus Afu-39 (formerly known as LS -39) (Jenneckens et al., 2001) the allele with 111 base pairs is private and monomorphic for this species. This locus can be used in combination with mitochondrial DNA for the identification of both the pure species and stellate sturgeon hybrids. In this case, the discordance with the mitochondrial haplotype will allow to detect the hybrid.

Loci in which no fully diagnostic alleles are present can also be used if some alleles are present at a high frequency in a given species. In these cases, the identification is performed by estimating the probability of belonging to a given species, following allocation procedures. However, the use of these markers for forensic purposes is much more critical as they cannot provide a certainty, but only an allocation probability. It is worth noting that the probability of observing a given allele in a genome depends not only on

the allele frequency in the species but also on its ploidy level (See paragraph 2.2.5 for details). Based on these criteria, several microsatellite loci have been proposed as diagnostic markers for the detection of different sturgeon species (Barmintseva & Mugue, 2013; Boscari et al., 2017a; Chassaing et al., 2011).

2.2.2.2 Single Nucleotide Polymorphisms (SNPs)

Single Nucleotide Polymorphisms (SNPs) are substitutions of a single nucleotide and represent the most frequent case of genetic polymorphism. Such substitutions can be localized both on coding and non-coding regions and their characterization can be carried out through different approaches. Species identification based on SNP analysis has become increasingly important in recent years, especially thanks to the advent of massive genotyping by sequencing methods (Elshire et al., 2011). The advantage of these approaches is that they allow the simultaneous genotyping of several SNPs in hundreds of individuals without the need to preliminarily identify a panel of polymorphic loci. However, these approaches, being time and lab consuming, are useful for the detection of polymorphism but are not convenient for applications on a limited number of samples. In these cases, it is necessary to use diagnostic SNPs (or combinations of SNPs) that can be genotyped quickly and reliably. Once the diagnostic SNPs have been identified using the above-mentioned whole genome approaches (Havelka et al., 2017; Ogden et al., 2013) or focusing on highly polymorphic regions such as intronic sequences (Boscari et al. 2014b, 2017a, 2017b), a locus-specific test can be developed, for example by developing PCR-based tests based on the presence/absence of diagnostic alleles. The good transferability and replicability of this approach among laboratories represent important advantages of SNPs compared to microsatellites.

The application of SNPs to sturgeon identification was first proposed as a single locus approach by Boscari et al. (2014a). The authors set up a PCR-based identification tool based on a panel of diagnostic SNPs within an intronic region of the RPS7 gene. The approach, combined with mtDNA, allows the unambiguous detection of five species (*A. naccarii*, *A. fulvescens*, *A. stellatus*, *A. sinensis* and *A. transmontanus*) and two hybrids, the Bester (*H. huso* ♀ x *A. ruthenus* ♂) and the AL (*A. naccarii* ♀ x *A. baerii* ♂), with an 80% and 100% success, respectively. The use of intronic SNPs has been later extended to another intronic region, of the gene RPS6, again with a single locus approach, allowing for the identification of *H. huso* (Boscari et al., 2017a; Figure 2A). Introns of RPL4 and RPL5 genes, combined with a microsatellites panel, allowed to setup a multi-locus approach for the differentiation between *A. schrenckii* and *H. dauricus* with discrete overall diagnostic results (Boscari et al., 2017a) even if no fixed markers were found that differentially segregate between the two species (see chapter 2.2 on the importance of reliable reference samples). Havelka et al. (2017) have identified other diagnostic SNPs for *A. ruthenus* and *H. huso* after a massive screening of loci obtained through Next Generation Sequencing. More recently, through a ddRAD approach, Havelka et al. (2019) isolated new SNPs able to differentiate *A. baerii* and *A. gueldenstaedtii* from all the other species, thus allowing to detect the presence of one of them as paternal species in hybrids. A similar whole genome rarefaction approach (2bRAD) enabled Boscari et al (2021) to differentiate geographical origins in the Beluga, but this aspect can be relevant mainly in a conservation perspective, for a reliable identification of conservation units. Multiple, in linkage, SNPs (microhaplotypes) are the focus of more recent studies that demonstrated how their use, instead of stand-alone SNPs, can effectively increase the informative potential for diversity analysis (Baetscher et al., 2018; Pakstis et al., 2021).

The University of Padova is currently developing the multiple SNPs approach focusing on intronic regions, which showed interesting potentials for sturgeon specie identification (Boscari et al., 2014a). This new approach is based on the simultaneous amplification and sequencing of several intronic loci through illumina sequencing with the aim of identifying species-specific haplotypes to be used for the subsequent development of diagnostic primers. In a preliminary application, individuals of fourteen species were simultaneously genotyped at 18 loci and the results, based on allele sharing among individuals, show a promising detection power as all species are clearly differentiated (Figure 2B). The only exceptions are *A. schrenckii* and *H. dauricus*, supporting the hypothesis of hybridization events between the two species, at least in the individuals used in this analysis, consistently with what observed by Boscari et al (2017a).

2.2.3 Effects of ploidy on the efficiency of nuclear markers

Whatever nuclear marker used, the probability of identifying a given species depends on the frequency of the diagnostic allele and the degree of ploidy of the organism.

The 27 species of sturgeon and paddlefish can be separated in different groups based on their ploidy level. The first group with about 120 chromosomes, the second with some 240 and the third, represented only by *A. brevirostrum* with 360 chromosomes. The ploidy level to be ascribed to each group depends on the criteria of classification (evolutionary or functional) as synthesized by Fontana et al. (2008). Focussing on the evolutionary perspective, the first group, having experienced two rounds of whole genome duplications, can be considered as tetraploid, with consequent status of octaploid and dodecaploids for the other two groups. On contrary, when focussing on functional aspects, due to the significant

functional genome re-diploidization, the three groups can be considered as di-, tetra- and hexaploidy (Fontana et al., 2008). The functional reduplication, however, is a dynamic and still ongoing process and the ploidy level at different loci cannot be taken for granted and should be viewed with caution (Ludwig et al., 2001). These aspects should be considered in developing species identification markers, especially in the case of Unfixed Diagnostic Alleles (UDA). These are alleles which are present only in one target species with a frequency lower than one. The target species can be unambiguously identified by the corresponding UDA but cannot be excluded if the UDA is not detected. Being an UDA not fixed by definition, a certain percentage of genomes in the population have alternative alleles; consequently, a genetic test (for example an allele-specific PCR) developed to specifically detect the UDA will yield negative results if, for chance, all genome copies of an individual carry these alternative alleles. The probability that a given individual does not have any copy of the UDA but carries only alternative alleles is counter proportional to his ploidy as schematized in the Figure 3. To better explain this concept, let consider a hypothetical UDA with frequency of 0.33 on which an allele-specific diagnostic PCR is set up. The probability that a diploid animal has only alternative alleles being negative to the PCR, is $0.66^2 = 0.44$, that means that only 54% of the individuals are detectable. If an UDA with the same frequency of 0.33 is observed in a tetraploid species, the probability to carry only alternative alleles is $0.66^4 = 0.19$, and as a result 81% of individuals will be positive to the diagnostic PCR.

For this reason, even relatively low frequency diagnostic alleles can provide valuable detection power, especially in polyploids.

2.3 The importance of Reference Samples (RS)

Irrespective of the methods or the genetic markers used, the selection of Reference Samples (RS) in the developmental phase of forensic tools is crucial. RS must be certified with regard to their purity and should be representative of the genetic variability of the species or population of origin to avoid the following biases.

Firstly, difficulty in identifying diagnostic markers due to not pure individuals used as RS. Often, for the development of diagnostic markers, sturgeons from aquaculture facilities or from captive stocks are used as RS. The history of these animals is sometime unclear. In many cases, following translocations among plants or mixing of individuals of different origins, traces of their individual histories are lost. This generally does not represent a big problem since the species are morphologically recognizable. However, in aquaculture, pure individuals often coexist with different degrees of hybrids showing intermediate features which render the morphological-based identification a problem. This is especially true when the hybrid does not represent a first-generation specimen, but the hybridization event occurred in their earlier genealogy, followed by backcrosses. In this case, the contribution of the introgressed species is retained in the population or in the captive stock for several generations being progressively diluted. Consequently, some animals that phenotypically are classifiable as pure species might have a certain fraction of genome inherited from other species. Their inclusion in the reference dataset dramatically decreases the possibility to identify diagnostic markers. This situation was hypothesized in the case of the previously mentioned study conducted by Boscari et al. (2017a) on *H. dauricus* and *A. schrenckii* (sympatric in the Amur River). The authors have failed in identifying diagnostic markers to distinguish the two species, despite belonging to different genera and the fact that they are morphologically very dissimilar. The production of hybrids between the two species is in fact very common in

aquaculture in China but the two sympatric species are also known to occasionally hybridize in nature. In this case, a potential alternative is developing markers that identify the two species as a whole (Boscari, et al., 2014a) and limiting any testing to the compatibility of a sample with these two species, without attributing them with certainty to one of them. This approach, however, does not meet the identification criteria required by CITES regulations.

Secondly, incomplete representation of a species in the RS. Due to a deficit in sampling effort, it is possible that the individuals used as a reference for a species are not completely representative of its genetic diversity. Often, animals originated from one or few crosses are reared in aquaculture resulting in a high degree of relatedness. In these cases, a genetic variant can be wrongly identified as always present or always absent from a given species and its application to forensic analysis could lead to false interpretations. For this reason, it is important to verify that the validation process of the identified markers has been carried out on the largest number of animals possible, with different origins, possibly analysed by different laboratories. An example of this type of bias is given by Zhang et al. (2013) in which the authors used microsatellites to differentiate several reciprocal hybrids generated by switching the sire's and dam's species (*A. baerii*♀ × *A. gueldenstaedtii*♂ versus *A. gueldenstaedtii*♀ × *A. baerii*♂; *A. baerii*♀ × *A. schrenckii*♂ versus *A. schrenckii*♀ × *A. baerii*♂ and *H. dauricus*♀ × *A. schrenckii*♂ versus *A. schrenckii*♀ × *H. dauricus*♂) as genetically distinct. This is not genetically expected because first-generation hybrids between two species should group together regardless of the maternal or paternal species. It is conceivable that the group of hybrids analysed in this study comprised related animals and were not representative of the potential variability. Thirdly, limited number of reference species. A suitable species identification marker is

expected to show one state in the target species and alternative states in all other ones. This condition can be verified only if all other species are included in the validation test, but this is unfortunately often not the case. An example is given by a *A. sinensis* marker that was proposed by Boscari et al. (2014b) which, following further validation tests, proved to be partially shared with a part of the individuals of *A. dabryanus*, a species that was not included in the original dataset.

3 DETERMINING AND TRACKING THE SOURCE AND GEOGRAPHICAL ORIGINS OF STURGEONS AND THEIR PRODUCTS

Along with the problem of setting up reliable methods for species and hybrid identification, a second challenging issue is providing tools able to identify the origin of sturgeons and to trace the derived products. Particularly relevant to defeat the illegal trade and the whitewashing of wild sturgeon sold as captive-bred is the possibility of distinguishing wild from aquaculture products. Different approaches for these purposes were proposed on fish (Arechavala et al., 2013), among which two have proven to be effective in addressing this question: fatty acid profiling and stable isotope analysis.

3.1 Specific chemical compound profiling (fatty acid)

Fatty acids are both important dietary sources and important structural components of cells in all biological organisms. Their composition is largely determined by the animal's diet (Zhang et al., 2020).

The analyses of fatty acid composition can be used to distinguish wild from farmed origin, as shown in salmonids (Molkentin, 2015) as well as in sturgeon caviar (DePeters et al., 2013; Gessner et al. 2002). While Czesny et al. (2000) pointed out that focusing on stearic and oleic acids would provide suitable tools for this purpose, basing the investigation on

the entire fatty acid profile instead of relying on only one or few discriminating acids provides more reliable results (Vasconi et al., 2019). In fact, since the possibility to distinguish wild from farmed caviar depends on the food ingredients which are variable across farms, the more informative fatty acid for discrimination might change as well.

Unfortunately, fatty acid profiling is a labour-intensive and time-consuming procedure (approximately 5 to 10 days). The proteins need to be hydrolyzed, and the fatty acids are derivatized breaking them into the original esters (triglycerides, waxes, phospholipids etc.) before converting them to methyl esters that can be finally analyzed by gas chromatography. New, upcoming technologies, like Direct Analysis in Real Time (DART), coupled with a time-of-flight (TOF) mass spectrometer (MS) (e.g. DART-TOF-MS), may improve routine analysis of fatty acids by massively cutting down on costs and shortening the time needed to less than an hour including data analyses (Black et al., 2016; Cajka et al., 2013). The biggest advantage of this technology is that almost no sample preparation is required, and the chemical profile of a suitable sample can be obtained within seconds.

Therefore, DART is already used in various fields of authenticity verification. Thus, the field of work ranges from the verification of the authenticity of foods such as olive oil (Vaclavik et al., 2009) to forensics (Sisco & Forbes, 2021) and the routine application of species verification in wood (Price et al., 2022). However, the application of this approach to the analyses of caviar is still at its infancy but the results of a preliminary analysis performed by Agroisolab (Berlin) show a remarkable potential of the approach in distinguishing the different origins.

In combination, DART-TOF-MS provides a typical mass fingerprinting of those reference samples. The differentiation of caviar from wild and farmed origin using Kernel Discriminant Analysis (KDA) of the fatty acid chemical fingerprint from the DART-TOF

analysis was performed on caviar obtained from farmed and wild origin (Figure 4). The analysis of the environment of origin (wild versus farmed) performed with DART-MS is highly concordant with the stable isotopes approach described in the next paragraph. The analyses of 58 test samples with both approaches yielded concordant results in 88% of cases (51 samples). In three samples the DART-TOF-MS gave no significant assessment, so the result could not be compared, and in four cases (7%) results were discordant. In these cases, further background information (i.e. on the feed used in aquaculture) might be helpful to draw the right conclusions. Regarding the costs, the method with a price of about 40 Euros per sample is quite inexpensive.

3.2 Stable isotopes

Stable Isotope Ratio Analysis (SIRA) is the analysis of variations in the bioelements C, O, H, N and S. Isotopes are atoms of the same element that have different masses, due to the variable numbers of neutrons in the core of the atom. In nature, different patterns of these isotopes occur in the tissue samples depending on geographic origin. These geographical patterns of isotopes composition can be used to trace the geographical origin of a sample through Isotope Ratio Mass Spectrometers (IRMS), allowing to verify the authenticity of materials (Primrose et al., 2010). There is a wide array of publications available on their use to determine the provenance of various products such as agricultural and animal products (see Divelos & Constantinos, 2012 for review).

The remarkable sensitivity of the method also allows for depicting temporal variations within individuals if applied to continuously growing tissues, for example to study animal migration patterns (Hobson, 1999). It was also applied to investigate interspecific trophic relationships and aquatic food webs (Middleburg, 2014). Given the great

geographic variability in isotopic patterns in water (Bowen & Wilkinson, 2002), water isotopes are largely used to assess the geographical origin of organisms. In fact, the location-dependent water isotope signatures are reflected in the tissue water of plants and animals (Boner & Foster, 2004) and in their products (Lesley et al., 2010). Most studies using stable isotopes for authentication purposes focus on terrestrial plants and animals. This is due to the fact that the first applications focused on the analysis of hydrogen and oxygen which have informative patterns in freshwater. In fact, $^2\text{H}/^1\text{H}$ and $^{18}\text{O}/^{16}\text{O}$ stable isotope ratios have been successfully used to verify the provenance of freshwater fish (Soto et al., 2016). The sea water shows reduced variability of these elements world-wide and their use is therefore less effective in sea organisms. As most commercial caviar is obtained from animals reared in freshwater aquaculture, the stable isotopes of oxygen and hydrogen can be applied to it. The comparison performed on more than 50 reference samples from 9 different countries is shown in Figure 5A, revealing a good distinction among caviar from different areas.

Further information can be obtained from the analyses of stable strontium isotopes ($^{87}\text{Sr}/^{86}\text{Sr}$) which have been successfully used to track the origin of salmon (Barnett-Johnson et al., 2008). Unfortunately, strontium isotopes are highly affected by the processing of sturgeon and caviar (e.g., salt effect). Nevertheless, new models have been developed to minimize the processing effect and finally to link the sturgeon sample with the origin (Tchaikovsky et al., 2019). Further stable isotopes of other bio elements such as carbon, nitrogen, and sulphur measured in animal tissue reflect indirectly the origin of the animals and of their products. For instance, carbon and nitrogen isotope ratios allowed to differentiate salmon of wild, organic, and conventional-farmed origin (Molkentin et al., 2007) and shrimps from wild-caught and aquaculture production

systems (Gamboa-Delgado et al., 2014). Unfortunately, these elements seem to have a limited potential for sturgeon identification with respect to the sturgeon farming system probably due to the complexity of sturgeon feed, as well as the varying feeding conditions. However, their combination with sulphur isotopes, even if still rarely used in aquatic studies (Hesslein, 1991; Kiyashko et al. 2011), results in a high discrimination power between wild and farmed origin (Figure 5B). Preliminary results based on 76 reference samples of caviar (34 wild and 42 from aquaculture) Analyzed by Agroisolab (Jülich, Germany), showed that the current separation probability using discriminate analysis (DA) by $\delta^{13}\text{C}$, $\delta^{15}\text{N}$ and $\delta^{34}\text{S}$ of wild and farm caviar is over 93%.

3.3 Framing pros and cons of isotope utilization for forensic analysis. To conclude, stable isotopes deliver a wide range of information about the geographical origin and the production method (wild/farm). Although a routine implementation of stable isotope analyses remains challenging and expensive, it is at the moment the leading standard method for authentication due to its potential for universal application to a broad range of products and has also been accepted in various court cases as an analytical proof (Camin, 2017).

The main restriction in application is that stable isotope technology is discriminative only if the diet of fish in the wild is different to commercial feed used in aquaculture. Therefore, it is important to keep in mind that the method cannot actually discriminate wild-caught from captive-bred sturgeon but rather sturgeon fed with natural versus commercial feed (e.g. American Paddlefish from an aquaculture facility in Ukraine was determined as wild-caught as it feeds on natural food only).

Finally, it should be stressed that the stable isotope applications require dedicated reference databases: the first made up by geolocated samples for the identification of

the geographical origin, and the second with known wild and farmed samples for the distinction of the two origins in unknown individuals.

4 IMPLICATIONS FOR CONSERVATION

Methods that allow distinguishing wild and farmed animals have direct implications for conservation as they provide the possibility to detect illegal trade or the illegal exploitation of natural populations and stable isotope methods have already been accepted as analytical proof in court cases (e.g. in Germany).

Furthermore, the possibility of correctly identifying sturgeon species and hybrids has a great conservation value. Most sturgeon populations are on the verge of extinction and many conservation programs include the support of restocking programs. The animals used for restocking are often produced in captivity through controlled reproduction. In many cases, the broodstock originates from aquaculture facilities. These fish should be checked genetically to assess their relatedness and their purity in terms of the species they belong to and the population of origin. Given the thriving production of different kinds of hybrids and exotic sturgeon species in aquaculture and since animals descending from a hybridization followed by different rounds of backcrosses cannot be morphologically distinguished from pure ones, the probability of accidental inclusion of impure individuals in the broodstock and the resulting risk of massive genetic pollution of already reduced natural populations is significant (Boscari et al., 2014b; Congiu et al., 2011). With the aim of selecting individuals whose purity is confirmed at as many loci as possible, the application of whole genome genotyping approaches is increasingly used. An enormous amount of information can be collected by these approaches dramatically increases the chances of detecting informative loci on which developing multilocus diagnostic tests of rapid and economic applicability. This is the direction

that will be followed in the coming years for the development of diagnostic markers in sturgeons.

5 CONCLUSIONS

The illegal trade in sturgeons is continuously putting remaining wild populations at risk. CITES listing of sturgeons in 1997 and the subsequent mandatory labeling of caviar in international trade provided a first set of tools to increase control of illegal trade. Despite this, recent investigations (Jahrl et al. 2021) show that illegal trade still comprises significant proportions of the commercial utilization of sturgeons on domestic markets, including poached products, fraud, as well as consumer deceit.

Improved regulations allowing the clear discrimination of legality of products and a consistent labelling technique must be put in place also on domestic markets, supported by systematic controls of the commodities.

The available methods for these controls have been outlined in this paper to provide a sound guidance on the information to be gained by these analyses. To make them effective tools to stop illegal trade in all its forms, national reference laboratories should be trained and inter-calibrated to deliver meaningful and jurisdictionally sound results acceptable by courts.

In addition, more communication and information efforts are necessary by public bodies but also by caviar producers to allow the consumers to make educated decisions, being aware about the conservation relevance of their behavior.

Credit authorship contribution statement

Leonardo Congiu: Conceptualization, Investigation, Writing original draft, Writing – review & editing. Beate Striebel-Greiter: Conceptualization, Writing – review & editing, Joem Gessner: Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Elisa Boscari: Investigation, Writing – original draft, Writing – review & editing, Markus Boner, Investigation, Writing – original draft, Writing – review & editing; Jutta Jahrl: Conceptualization, Writing – review & editing, Stefano Dalle Palle Formal analysis, Writing – review & editing, Arne Ludwig: Conceptualization, Investigation, Writing original draft, Writing – review & editing.

Acknowledgement

We thank all colleagues who contributed to the discussion during the preparation of this synthesis and Sabine Hofem and Claudia Erven for their technical support in biochemical and isotope analyses.

This research was funded by Italian PRIN 2017, grant number 201794ZXTL. The authors declare no conflict of interest in relation to this paper.

Data statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

- Ahmed, N., Sangale, D., Tiknaik, A., Prakash, B., Hange, R., Sanil, R., Sajid, K. & Khedkar, G. (2018). Authentication of origin of meat species processed under various Indian culinary procedures using DNA barcoding. *Food Control*, 90, 259-265. <https://doi.org/10.1016/j.foodcont.2018.02.012>
- Arechavala-Lopez, P., Fernandez-Jover, D., Black, K.D., Ladoukakis, E., Baye-Sempere, J.T., Sanchez-Jerez, P., Dempster, T. (2013). Differentiating the wild or farmed origin of Mediterranean fish: a review of tools for sea bream and sea bass. *Reviews in Aquaculture* 5, 137-157. <https://doi.org/10.1111/raq.12006>
- Baetscher, D.D., Clemento, A. J., Ng, T.C., Anderson, E.C. & Garza, J.C. (2018). Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources*, 18, 296-305. <https://doi.org/10.1111/1755-0998.12737>
- Barmintseva, A.E. & Mugue, N.S. (2013). The use of microsatellite loci for identification of sturgeon species (*Acipenseridae*) and hybrid forms. *Genetika*, 49, 1093-105. <https://doi.org/10.7868/s0016675813090038>
- Barnett-Johnson, R., Pearson, T. E., Ramos, F.C., Grimes, C.B. & MacFarlane, R.B. (2008). Tracking natal origins of salmon using isotopes, otoliths, and landscape geology. *Limnological Oceanography*, 53, 1633-1642. <https://doi.org/10.4319/lo.2008.53.4.1633>
- Beridze, T., Boscari, E., Scheele, F., Edisherashvili, T., Anderson, C. & Congiu, L. (2022) Interspecific hybridization in natural sturgeon populations of the Eastern Black Sea: the consequence of drastic population decline? *Conservation Genetics*, 23, 211-216. <https://doi.org/10.1007/s10592-021-01413-7>
- Birstein, V.J., Doukakis, P., Sorkin, B. & DeSalle, R. (1998). Population Aggregation Analysis of Three Caviar-Producing Species of Sturgeons and Implications for the Species Identification of Black Caviar. *Conservation Biology*, 12, 766-775. <https://doi.org/10.1111/j.1523-1739.1998.97081.x>
- Birstein, V.J., Doukakis, P. & De Salle, R. (2000). Polyphyly of mtDNA lineages in the Russian sturgeon, *Acipenser gueldenstaedtii*: forensic and evolutionary implications. *Conservation Genetics*, 1, 81-8. <https://doi.org/10.1023/A:1010141906100>
- Black, C., Chevallier, O.P. & Elliot, C.T. (2016). The current and potential applications of ambient mass spectrometry in detecting food fraud. *Trends in Analytical Chemistry*, 82, 268-278. <https://doi.org/10.1016/j.trac.2016.06.005>
- Boner, M. & Förstel, H. (2004). Stable isotopes variation as a tool to trace the authenticity of beef. *Analytical and Bioanalytical Chemistry*, 378, 301-310. <https://doi.org/10.1007/s00216-003-2347-6>
- Boscari, E., Barmintseva, A. E., Pujolar, J.M., Doukakis, P., Mugue, N. & Congiu, L. (2014a). Species and hybrid identification of sturgeon caviar: a new molecular approach to detect illegal trade. *Molecular Ecology Resources*, 14, 489-98. <https://doi.org/10.1111/1755-0998.12203>
- Boscari, E., Pujolar, J.M., Dupanloup, I., Corradin, R. & Congiu, L. (2014b). Captive Breeding Programs Based on Family Groups in Polyploid Sturgeons. *PLoS ONE*, 9(10), e110951. <https://doi.org/10.1371/journal.pone.0110951>
- Boscari, E., Barmintseva, A.E., Zhang, S., Yue, H., Li, C., Shedko, V.S., Lieckfeldt, D., Ludwig, A., Qiwei, W., Mugue, N.S. & Congiu, L. (2017a). Genetic identification of the caviar-producing Amur and Kaluga sturgeons revealed a high level of concealed hybridization. *Food Control*, 82, 243-250. <http://dx.doi.org/10.1016/j.foodcont.2017.07.001>
- Boscari, E., Vitulo, N., Ludwig, A., Caruso, C., Mugue, N.S., Suciu, R., Onara, D.F., Papetti, C., Matino, I.A.M., Zane, L. & Congiu, L. (2017b). Fast genetic identification of the Beluga sturgeon and its sought-after caviar to stem illegal trade. *Food Control*, 75, 145-152. <https://doi.org/10.1016/j.foodcont.2016.11.039>
- Bowen, G.J. & Wilkinson, B. (2002). Spatial distribution of $\delta^{18}\text{O}$ in meteoric precipitation. *Geology*, 30, 315-318. [https://doi.org/10.1130/0091-7613\(2002\)030<0315:SDOOIM>2.0.CO;2](https://doi.org/10.1130/0091-7613(2002)030<0315:SDOOIM>2.0.CO;2)
- Bronzi, P., Chebanoc, M., Michaels, J.T., Wei, Q., Rosenthal, H. & Gessner, J. (2019). Sturgeon meat and caviar production: Global update 2017. *Journal of Applied Ichthyology*, 35, 257-266. <https://doi.org/10.1111/jai.13870>
- Boscari, E., Marino, I.A.M., Caruso, C., Gessner, J., Lari, M., Mugue, N., Barmintseva, A., Suciu, R., Onara, D., Zane, L. & Congiu, L. (2021). Defining criteria for the reintroduction of locally extinct populations based on contemporary and ancient genetic diversity: the case of the Adriatic Beluga sturgeon (*Huso huso*). *Diversity and Distributions*, 27(5), 816-827. <https://doi.org/10.1111/ddi.13230>
- Cajka, T., Danhelova, H., Vavrecka, A., Riddellova, K., Kocourek, V., Vacha & F., Hajslova, J. (2013). Evaluation of direct analysis in real time ionization-mass spectrometry (DART-MS) in fish metabolomics aimed to assess the response to dietary supplementation. *Talanta*, 115, 263-270. <https://doi.org/10.1016/j.talanta.2013.04.025>
- Camin, F. (2017). Stable isotopes techniques for verifying the declared geographical origin of food in legal cases. *Trends in Food Science & Technology*, 61, 176-187. <https://doi.org/10.1016/j.tifs.2016.12.007>
- Chassaing, O., Hanni, C. & Berrebi, P. (2011). Distinguishing species of European sturgeons *Acipenser* spp. using microsatellite allele sequences. *Journal of Fish Biology*, 78, 208-226. <https://doi.org/10.1111/j.1095-8649.2010.02852.x>
- Cohen, A. (1997). Sturgeon poaching and black market caviar: a case study. *Environmental Biology of Fishes*, 48, 423-426. <https://doi.org/10.1023/A:1007388803332>
- Congiu, L., Pujolar, J.M., Forlani, A., Cenadelli, S., Dupanloup, I., Barbisan, F., Galli, A. & Fontana F. (2011). Managing Polyploidy in Ex Situ Conservation Genetics: The Case of the Critically Endangered Adriatic Sturgeon (*Acipenser naccarii*). *PLoS ONE*, 6(3), e18249. <https://doi.org/10.1371/journal.pone.0018249>
- Cooper, J.E., Cooper, M.E. & Budgen, P. (2009). Wildlife crime scene investigation: techniques, tools and technology. *Endangered Species Research*, 9, 229-238. <https://doi.org/10.3354/esr00204>
- Congiu, L., Gessner, J. & Ludwig, A. (2023). IUCN Red List reassessment reveals further decline of sturgeons and Paddlefishes. *Oryx*, accepted.
- Czesny, S., Dabrowski, K., Christensen, J. E., Eenennamm, J. V. & Doroshov, S. (2000). Discrimination of wild and domestic origin of sturgeon ova based on lipids and fatty acid analysis. *Aquaculture*, 189, 145-153. [https://doi.org/10.1016/S0044-8486\(00\)00364-1](https://doi.org/10.1016/S0044-8486(00)00364-1)
- Daea, M. (2019, 14 October). Caviar, crime and corruption on Romania's Danube River. Earth Journalism Network. Retrieved from: <https://earthjournalism.net/stories/caviar-crime-and-corruption-on-romania-s-danube-river>
- Dalton, D.L. & Kotze, A. (2011). DNA barcoding as a tool for species identification in three forensic wildlife cases in South Africa. *Forensic Science International*, 207, e51-e54. <https://doi.org/10.1016/j.forsciint.2010.12.017>
- Dawnay, N., Ogden, R., McEwing, R., Carvalho, G.R. & Thorpe, R.S. (2007). Validation of the barcoding gene COI for use in forensic genetic species identification. *Forensic Science in international*, 173, 1-6. <https://doi.org/10.1016/j.forsciint.2006.09.013>
- Debus, L., Winkler, M. & Billard, R. (2002). Structure of micropyle surface on oocytes and caviar grains in sturgeons. *International Review of Hydrobiology*, 87: 585-603. [https://doi.org/10.1002/1522-2632\(200211\)87:5<585::AID-IROH585>3.0.CO;2-8](https://doi.org/10.1002/1522-2632(200211)87:5<585::AID-IROH585>3.0.CO;2-8)

- DePeters, E. J., Puschner, B., Taylor, S. & Rodzen, J. (2013). Can fatty acid and mineral compositions of sturgeon eggs distinguish between farm raised versus wild white (*Acipenser transmontanus*) sturgeon origins in California? Preliminary report. *Forensic Science International*, 229, 128-132. <https://doi.org/10.1016/j.forsciint.2013.04.003>
- Dillman, C.B., Zhuang, P., Zhang, T., Zhang, L.-Z., Mugue, N. & Hilton, E.J. (2014). Forensic investigations into a GenBank anomaly: endangered taxa and the importance of voucher specimens in molecular studies. *Journal of Applied Ichthyology*, 30, 1300-1309. <https://doi.org/10.1111/jai.12568>
- Diveilos, S.A. & Constantinos, G.A. (2012). Multi-element and multi-isotope ratio analysis to determine the geographical origin of food in the European Union. *Trends in Analytical Chemistry*, 40, 38-51. <https://doi.org/10.1016/j.trac.2012.08.003>
- Doukakis, P., Pikitch, E.K., Rothschild, A., DeSalle, R., Amato, G. & Kolokotronis, S.-O. (2012). Testing the effectiveness of an international conservation agreement: Marketplace forensics and CITES Caviar trade regulation. *PLoS ONE*, 7, 1-9. <https://doi.org/10.1371/journal.pone.0040907>
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. & Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6(5), e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Environment and Climate Change (ECC) Canada (2001). CITES Identification Guide — Sturgeons and Paddlefish - Guide to the Identification of Sturgeon and Paddlefish species Controlled under the Convention on International Trade in Endangered Species of Wild Fauna and Flora. ISBN 0-660-61641-6
- Translated into Ukrainian by WWF. Available at: https://ua.danube-sturgeons.org/wp-content/uploads/2019/03/idguide_full.pdf
- Fain, S.R., Straughan, D. J., Hamlin, B. C., Hoesch, R. M. & LeMay, J.P. (2013). Forensic genetic identification of sturgeon caviars traveling in world trade. *Conservation Genetics*, 14, 855-874. <https://doi.org/10.1007/s10592-013-0481-z>
- Fontana, F., Congiu, L., Mudrak, V.A., Quattro, J.M., Smith, T.I.J., Ware, K. & Doroshov, S.I. (2008). Evidence of hexaploid karyotype in shortnose sturgeon. *Genome*, 51, 113-119. <https://doi.org/10.1139/g07-112>
- Gamboa-Delgado, J., Molina-Poveda, C., Godinez, D. & Villarreal-Cavazos, D.A. (2014). Application of stable isotope analysis to differentiate shrimp extracted by industrial fishing or produced through aquaculture practices. *Canadian Journal of Fisheries and Aquatic Sciences*, 71, 1520-1528. <https://doi.org/10.1139/cjfas-2014-0005>
- Gessner, J. & Friedrich, T. (2017). Sturgeon identification guide for Georgia. WWF Caucasus Programme, Tbilisi, Georgia. 44p.
- Gessner, J., Wirth, M., Kirschbaum, F. & Krüger, A. (2002). Caviar composition in wild and cultured sturgeons – impact of food sources on fatty acid composition and contaminant load. *Journal of Applied Ichthyology*, 18, 665-672. <https://doi.org/10.1046/j.1439-0426.2002.00366.x>
- Havelka, M., Fujimoto, T., Hagihara, S., Adachi, S. & Arai, K. (2017). Nuclear DNA markers for identification of Beluga and Sterlet sturgeons and their interspecific Bester hybrid. *Scientific reports*, 7, 1694. <https://doi.org/10.1038/s41598-017-01768-3>
- Havelka, M., Boscari, E., Sergeev, A., Mugue, N., Congiu, L. & Arai, K. (2019). A new marker, isolated by ddRAD sequencing, detects Siberian and Russian sturgeon in hybrids. *Animal Genetics*, 50, 115-116. <https://doi.org/10.1111/age.12733>
- Hesslein, R. H. (1991). Stable Isotopes of sulfur, carbon, nitrogen as indicators of trophic level and fish migration in the lower Mackenzie river basin, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, 48, 2258-2265. <https://doi.org/10.1139/f91-265>
- Hobson, K.A. (1999). Tracing origins and migration of wildlife using stable isotopes: a review. *Oecologia*, 120, 314-326. <https://doi.org/10.1007/s004420050865>
- Hruby, D. (2019, August 21). Craving for caviar is driving the Danube River's sturgeon to extinction. LA Times. Retrieved from <https://www.latimes.com/world-nation/story/2019-08-21/craving-for-caviar-is-driving-the-danube-rivers-sturgeon-to-extinction>
- Jahrl, J., Boner, M., Striebel, B. & Ludwig, A. (2021). Evidence for Trafficking of Critically Endangered Sturgeon in the Lower Danube Region. WWF Austria & WWF CEE.
- Jenneckens, I., Meyer, J.N., Debus, L., Pitra, C. & Ludwig, A. (2000). Evidence of mitochondrial DNA clones of Siberian sturgeon, *Acipenser baerii*, within Russian sturgeon, *Acipenser gueldenstaedtii*, caught in the River Volga. *Ecology Letters*, 3, 503-508. <https://doi.org/10.1111/j.1461-0248.2000.00179.x>
- Jenneckens, I., Meyer, J.N., Hörstgen-Schwark, G., May, B., Debus, L., Wedekind, H. & Ludwig, A. (2001). A fixed allele at microsatellite locus LS-39 exhibiting species-specificity for the black caviar producer *Acipenser stellatus*. *Journal of Applied Ichthyology*, 17, 39-42. <https://doi.org/10.1046/j.1439-0426.2001.00234.x>
- Khedkar, T., Sharma, R., Tiknaik, A., Khedkar, G., Naikwade, B.S., Ron, T.B. & Haymer, D. (2014). DNA barcoding using skin exuvates can improve identification and biodiversity studies of snakes. *Mitochondrial DNA*, 27, 499-507. <https://doi.org/10.3109/19401736.2014.905830>
- Kiyashlo, S. I., Velivetskaya, T.A. & Ignatiev, A.V. (2011). Sulfur, Carbon, and Nitrogen Stable Isotope Ratios in Soft Tissues and Trophic Relationships of Fish from the Near-Shore Waters of the Peter the Great Bay in the Sea of Japan. *Russian Journal of Marine Biology*, 37, 297-302. <https://doi.org/10.1134/S1063074011040079>
- Knight, J. (2017). Operation Roadhouse: How inter-agency collaboration stopped illegal Paddlefish depredation. The National Agency of Conservation Law Enforcement Chiefs. Available at: <https://www.naclec.org/press-pages/2017/9/17/operation-roadhouse-how-inter-agency-collaboration-stopped-illegal-paddlefish-depredation>
- Lesley, A.C., Valenzuela, L.O., O'Grady, S.P., Cerling, T.E. & Ehleringer, J.R. (2010). Hydrogen and Oxygen Stable Isotope ratios of milk in the United States. *Journal of Agricultural and Food Chemistry*, 58, 2358-2363. <https://doi.org/10.1021/jf904151c>
- Luca, A.M., Holdis, D., & Barberá, M.G. (2020). Decimated Danube: Sturgeon Revival Efforts Neglect Roots of Poaching. Balkan Insight. Retrieved from <https://balkaninsight.com/2020/05/20/decimated-danube-sturgeon-revival-efforts-neglect-roots-of-poaching/>
- Ludwig, A. (2008). Identification of *Acipenseriformes* species in trade. *Journal of Applied Ichthyology*, 24(S1), 2-19. <https://doi.org/10.1111/j.1439-0426.2008.01085.x>
- Ludwig, A., Lippold, S., Debus, L. & Reinartz, R. (2009). First evidence of hybridization between endangered sterlets (*Acipenser ruthenus*) and exotic Siberian sturgeons (*Acipenser baerii*) in the Danube River. *Biological Invasions*, 11, 753-760. <https://doi.org/10.1007/s10530-008-9289-z>
- Ludwig, A., Belfiore, N.M., Pitra, C., Svirsky, V. & Jenneckens, I. (2001). Genome duplication events and functional reduction of ploidy levels in sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). *Genetics*, 158, 1203-1215. <https://doi.org/10.1093/genetics/158.3.1203>
- Middleburg, J.J. (2014). Stable isotopes dissect aquatic food webs from the top to the bottom. *Biogeosciences*, 11, 2357-2371. <https://doi.org/10.5194/bg-11-2357-2014>
- Mugue, N.S., Barmintseva, A.E., Rastorguev, S.M., Mugue, V.N. & Barmintsev, V.A. (2008). Polymorphism of the mitochondrial DNA control region in eight sturgeon species and development of a system for DNA-based species identification. *Russian Journal of Genetics*, 44, 793-798. <https://doi.org/10.1134/S1022795408070065>
- Moltekin, J., Meisl, H., Lehmann, I. & Rehbein, H. (2007). Identification of organically farmed Atlantic salmon by analysis of stable isotopes and fatty acids. *European Food Research and Technology*, 224, 535-543. <https://doi.org/10.1007/s00217-006-0314-0>
- Molkentin, J., Lehmann, I., Ostermeyer, U. & Rehein, H. (2015). Traceability of organic fish – Authenticating the production origin of salmonids by chemical and isotopic analyses. *Food Control*, 53, 55-66. <https://doi.org/10.1016/j.foodcont.2015.01.003>
- Ogden, R., Gharbi, K., Mugue, N., Martinsohn, J., Senn, H., Davey, J.W., Pourkazemi, M., McEwing, R., Eland, C., Vidotto, M., Sergeev, A. & Congiu, L. (2013). Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. *Molecular Ecology*, 22, 3112-3123. <https://doi.org/10.1111/mec.12234>

- Pakstis, A.J., Gandotra, N., Speed, W.C., Murtha, M., Scharfe, C. & Kidd, K. (2021). The population genetics characteristics of a 90 locus panel of microhaplotypes. *Human Genetics*, 140, 1753–1773. <https://doi.org/10.1007/s00439-021-02382-0>
- Pappalardo, A.M., Petraccioli, A., Capriglione, T. & Ferrito, V. (2019). From fish eggs to fish name: caviar species discrimination by Coibar-RFLP, an efficient molecular approach to detect fraud in the caviar trade. *Molecules*, 24(13), 2468. <https://doi.org/10.3390/molecules24132468>
- Price, E.R., McClure, P. J., Huffman, A.N., Voin, D. & Espinoza, E.O. (2022). Reliability of wood identification using DART-TOFMS and the ForeST database: A validation study. *Forensic Science International: Animals and Environments*, 2, 1-9. <https://doi.org/10.1016/j.fsiae.2022.100045>
- Primrose, S., Woolfe, M. & Rollinson, S. (2010). Food forensics: methods for determining the authenticity of foodstuffs. *Trends in Food Science and Technology*, 21, 582-590. <https://doi.org/10.1016/j.tifs.2010.09.006>
- Savić, I., Gmizic, D. (2018, July 26). Serbia's Appetite for Sterlet Drives Unique Sturgeon to the Brink. *Balkan Insight*. Retrieved from <https://balkaninsight.com/2018/07/27/serbia-s-appetite-for-sterlet-drives-unique-sturgeon-to-the-brink-07-26-2018/>.
- Sisco, E. & Forbes T.P. (2021). Forensic applications of DART-MS: A review of recent literature. *Forensic Chemistry*, 22, 1-16. <https://doi.org/10.1016/j.forc.2020.100294>
- Soto, D.X., Hobson, K.A. & Wassenaar, L.I. (2016). Using hydrogen isotopes of freshwater fish tissue as a tracer of provenance. *Ecology and Evolution*, 6, 7776-7782. <https://doi.org/10.1002/ece3.2519>
- Tchaikovsky, A., Zitek, A., Irrgeher, J., Oppen, C., Scheiber, R., Moder, K., Congiu, L. & Prohaska, T. (2019). Chemometric tools for determining site-specific elemental and strontium isotopic fingerprints in raw and salted sturgeon caviar. *European Food Research and Technology*, 245, 2515-2528. <https://doi.org/10.1007/s00217-019-03363-4>
- Van Uhm, D. & Siegel, D. (2016). The illegal trade in black caviar. *Trends Organized Crime* 19:67–87. <https://doi.org/10.1007/s12117-016-9264-5>
- Vaclavik, L., Cajka, T., Hrbek, V. & Hajslova J. (2009). Ambient mass spectrometry employing direct analysis in real time (DART) ion source for olive oil quality and authenticity assessment. *Analytica Chimica Acta*, 645, 56-63. <https://doi.org/10.1016/j.aca.2009.04.043>
- Vasconi, M., Lopez, A., Galimberti, C., Rojas, J.M.M., Redondo, J.M.M., Bellagamba, F. & Moretti, V.M. (2019). Authentication of farmed and wild European eel (*Anguilla Anguilla*) by fatty acid profile and carbon and nitrogen isotopic analyses. *Food Control*, 102, 112-121. <https://doi.org/10.1016/j.foodcont.2019.03.004>
- Zabyelina, Y.G. (2014). The "fishy" business: A qualitative analysis of the illicit market in black caviar. *Trends in Organized Crime*, 17(3), 181-198. <https://doi.org/10.1007/s12117-014-9214-z>
- Zhang, X., Wu, W., Li, L., Ma, X. & Chen, J. (2013). Genetic variation and relationships of seven sturgeon species and ten interspecific hybrids. *Genetics Selection Evolution*, 45, 21. <https://doi.org/10.1186/1297-9686-45-21>
- Zhang, X., Ning, X., He, X., Sun, X., Yu, X., Cheng, Y., Yu, Ri-Q. & Wu, Y. (2020). Fatty acid composition analyses of commercially important fish species from the Pearl River Estuary, China. *PLoS ONE*, 15(1), e0228276. <https://doi.org/10.1371/journal.pone.0228276>

Sp	Ref	Marker	Locus	Criteria	Efficiency	Cross-species tested	Comments
BAE	1	SNP	RPS7 (Intron RP1)	Diagnostic PCR product	60.6%	FUL, GUE, NAC, PER, RUT, SCH, SIN, STE, TRA, DAU, HUS	Shared with RUT
			Vimentine	Diagnostic PCR product	78%	GUE, PER	-
	2	SNP	Unknown	Diagnostic PCR product	99%	FUL, GUE, NAC, PER, RUT, SCH, SIN, STE, TRA, DAU, HUS	Shared with GUE and highly present in PER
FUL	1	SNP	RPS7 (Intron RP1)	Diagnostic PCR product	100%	BAE, GUE, NAC, PER, RUT, SCH, SIN, STE, TRA, DAU, HUS	-
GUE	2	SNP	Unknown	Diagnostic PCR product	96%	BAE, GUE, NAC, PER, RUT, SCH, SIN, STE, TRA, DAU, HUS	Shared with BAE and highly present in PER
MED	3	Microsatellite	An20	Species-specific range: 185-197	>92%	BAE, GUE, MIK, NUD, PER, RUT, SCH, STE, DAU, HUS	-
			AfuG51	Diagnostic allele: 296	100%		
			AoxD161	Diagnostic allele: 114	97%		
			AoxD165	Species-specific range: 230-254	68%		
NAC	1	SNP	RPS7 (Intron RP1)	Diagnostic PCR product	100%	BAE, GUE, NAC, PER, RUT, SCH, SIN, STE, TRA, DAU, HUS	-
	4	Microsatellite	LS54, Aox23	specific allelic range	100%	OXY, STU	Low number of species compared RS possibly related
			LS19, LS68 AoxD161	SNPs on flanking region	100%		Need of sequencing; Low number of species compared
NUD	3	Microsatellite	An20	Diagnostic allele: 153	100%	BAE, GUE, MED, MIK, PER, RUT, SCH, STE, DAU, HUS	-
			AfuG51	Diagnostic allele: 272	100%		
			AoxD165	Diagnostic allele: 196	74%		
OXY	4	Microsatellite	Aox23, AoxD161, LS68	SNPs on flanking region	100%	NAC, STU	Need of sequencing; Low number of species compared RS possibly related
RUT	1	SNP	RPS7 (Intron RP1)	Diagnostic PCR product	96%	BAE, GUE, NAC, PER, SCH, SIN, STE, TRA, DAU, HUS	Shared with BAE
	5	SNP	Unknown	Diagnostic PCR product	100%	BAE, GUE, MIK, NAC, PER, SCH, STE, TRA, DAU, HUS	-
	3	Microsatellite	An20	Diagnostic allele: 177	>80%	BAE, GUE, MED, MIK, NUD, PER, SCH, STE, DAU, HUS	Partially shared with DAU
			AfuG51	Diagnostic allele: 252	>91%		
SCH	1	SNP	RPS7 (Intron RP1)	Diagnostic PCR product	100%	BAE, GUE, NAC, PER, RUT, SIN, STE, TRA, DAU, HUS	Shared with DAU
	3	Microsatellite	An20	Diagnostic allele: 137	100%	BAE, GUE, MED, MIK, NUD, PER, RUT, STE, DAU, HUS	-
			AfuG41	Species-specific range: 185-197	Moderate		
SIN	1	SNP	RPL8 (Intron RP5)	Diagnostic PCR product	100%	BAE, GUE, NAC, PER, RUT, SCH, STE, TRA, DAU, HUS	Partially shared with DAU; To be used in combination with RPS7 (1)
			RPS7 (Intron RP1)	Diagnostic PCR product	100%		-

STE	1	SNP	RPS7 (Intron RP1)	Diagnostic PCR product	100%	BAE, FUL, GUE, NAC, PER, RUT, SCH, SIN, TRA, DAU, HUS	-
	3	Microsatellite	An20	Diagnostic allele: 141	85%	BAE, GUE, MED, MIK, NUD, PER, RUT, SCH, DAU, HUS	-
			AfuG51	Diagnostic allele: 288	86%		
STU	4	Microsatellite	AoxD161	Specific allelic range	100%	NAC, OXY	-
			Aox23, AoxD161, LS68	SNPs on flanking region	100%		Need of sequencing; Low number of species compared. RS possibly related
TRA	1	SNP	RPS7 (Intron RP1)	Diagnostic PCR product	100%	BAE, FUL, GUE, NAC, PER, RUT, SCH, SIN, STE, DAU, HUS	-
DAU	1	SNP	RPS7 (Intron RP1)	Diagnostic PCR product	100%	BAE, FUL, GUE, NAC, PER, RUT, SCH, SIN, STE, TRA, HUS	Shared with SCH
	3	Microsatellite	AoxD165	Private allele: 206	73%	BAE, GUE, MED, MIK, NUD, PER, RUT, SCH, STE, HUS	-
			AfuG41	No amplification	100%		
			An20	Diagnostic allele: 169	93%		
	6	SNP	RPL8 (Intron RP4)	Diagnostic PCR product	100%	SCH	Partially shared with SCH To be used in combination with RPS7 (1)
HUS	7	SNP	RPS6 (Intron RP2)	Diagnostic PCR product	100%	BAE, FUL, GUE, NAC, PER, RUT, SCH, SIN, STE, TRA, DAU	-
	5	SNP	Unknown	Diagnostic PCR product	100%	BAE, GUE, MIK, NAC, PER, RUT, SCH, STE, TRA, DAU	-
	3	Microsatellite	An20	Diagnostic alleles: 145, 149	83%, 45%	BAE, GUE, MED, MIK, NUD, PER, RUT, SCH, STE, DAU	-
			AoxD161	Diagnostic allele: 98	98%		
			AoxD165	Diagnostic allele: 178	96%		
			AfuG41	Species-specific range: 249-277	Moderate		

TABLE 1. Summary of nuclear markers available for the sturgeon species ID and information addressing their efficiency and cross-species tests. Concerning the column “Efficiency” where the probability to detect the target species is estimated for each marker, the frequency of diagnostic alleles and the ploidy of the species have been considered for the calculation. References are numbered as follows: 1: Boscari et al., 2014a; 2: Havelka et al., 2019; 3: Barmintseva & Mugue, 2013; 4: Chassaing, 2011; 5: Havelka et al., 2017; 6: Boscari et al., 2017a; 7: Boscari et al., 2017b. The cross-species tested for each marker are reported by using the official CITES standard code for the identification of sturgeon species. Species codes: BAE = *Acipenser baerii*; FUL = *Acipenser fulvescens*; GUE = *Acipenser gueldenstaedtii*; MED = *Acipenser medirostris*; MIK = *Acipenser mikadoi*; NAC = *Acipenser naccarii*; NUD = *Acipenser nudiventris*; OXY = *Acipenser oxyrhynchus*; PER = *Acipenser persicus*; RUT = *Acipenser ruthenus*; SCH = *Acipenser schrenckii*; SIN = *Acipenser sinensis*; STE = *Acipenser stellatus*; STU = *Acipenser sturio*; TRA = *Acipenser transmontanus*; DAU = *Huso dauricus*; HUS = *Huso huso*. For all other species, no diagnostic markers are available.

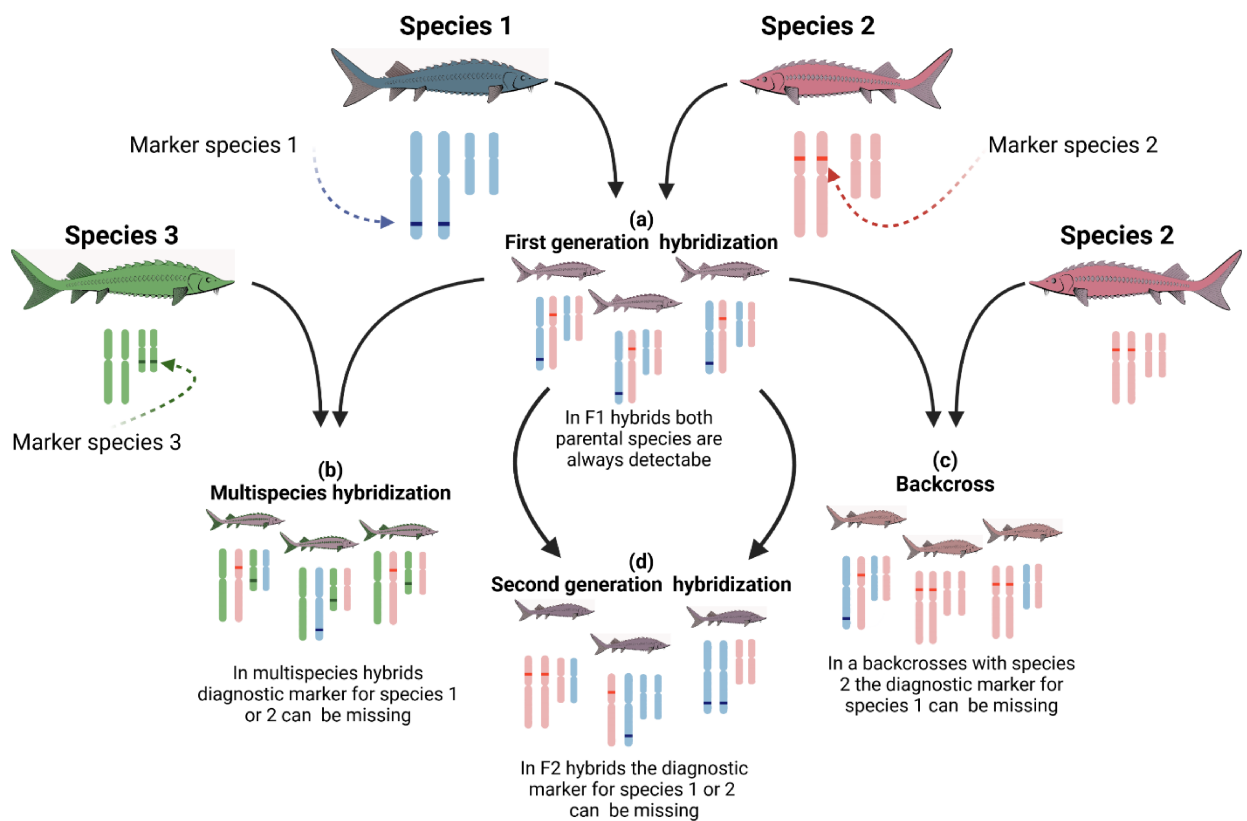


FIGURE 1

Schematic representation of different hybridization events: a) hybridization between two pure species (species 1 and 2) generating first generation hybrids (F1); b) hybridization between a first generation hybrid (F1) with a third pure species (Species 3) producing multispecies hybrids; c) hybridization between a first generation hybrid (F1) with one parental species (Species 2) producing a backcrossed hybrid, d) hybridization between two first generation hybrids (F1) generating second generation hybrids (F2). Genomes of the three species involved in the different hybridization events are schematized by chromosomes of the corresponding colors and bands on chromosomes highlight the presence of diagnostic markers for different species. Examples of chromosome inheritance and different possibilities to inherit the diagnostic markers are represented below each type of hybridization. The use of a single nuclear marker per species guarantees the detection of hybrids only in first generation. In the case of second (or more) generation hybrids, backcrosses or multispecies hybrids, the traceability of original contributing species can fail due to the loss of the corresponding marker. (Created with BioRender.com).

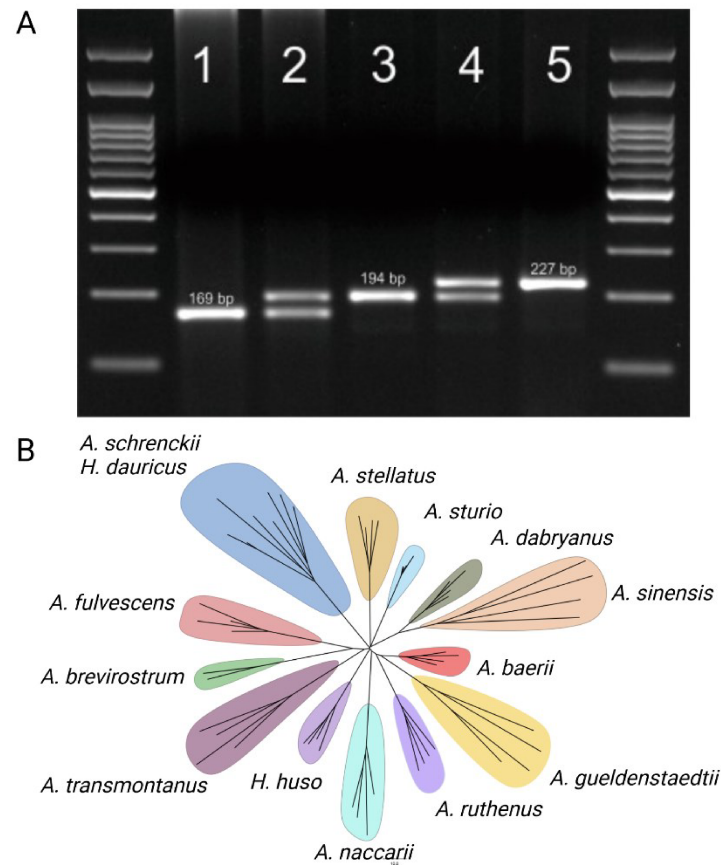


FIGURE 2

Different genetic approaches to species and hybrids identification. **A.** Identification of two sturgeon hybrids (*H. huso* x *A. ruthenus* and *A. naccarii* x *H. huso*) in lanes 2 and 4 and their pure parental species (*A. ruthenus*, *H. huso* and *A. naccarii*) in lanes 1, 3 and 5, based on specific PCR targeted to diagnostic SNPs in introns of RP2S6 and RP1S7 genes (Boscari et al 2017b). **B.** Preliminary results of multilocus intron analyses under investigation at the University of Padova showing a good potential as species identification tool. (Created with BioRender.com).

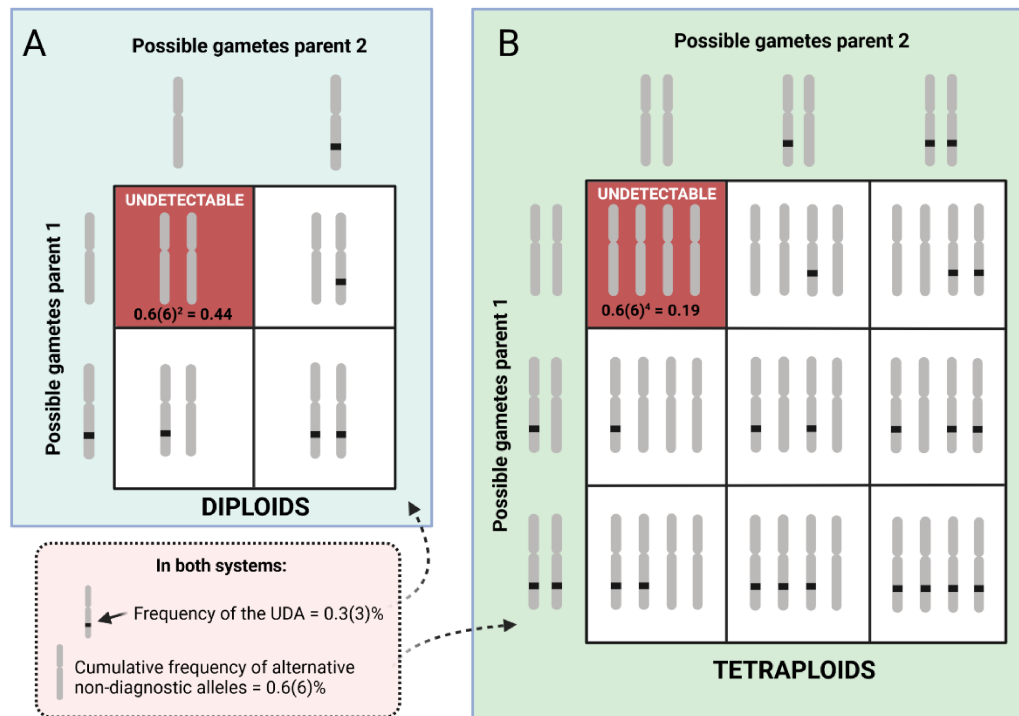


FIGURE 3

The effect of ploidy on markers' effectiveness. Schematic representation of possible genotypes resulting in diploids (A) and in tetraploids (B) from the inheritance of chromosomes carrying either an Unfixed Diagnostic Allele (UDA) or alternative non-diagnostic alleles. For both ploidies the UDA, represented by a black square on the chromosomes, is arbitrarily supposed to have a frequency of 0.33 to match the example reported in the text. All genotypes are detectable by a selective amplification of the UDA, with the only exception of those in the red squares. The undetectable genotype has a lower probability to occur in tetraploids, regardless of UDA frequency. (Created with BioRender.com)

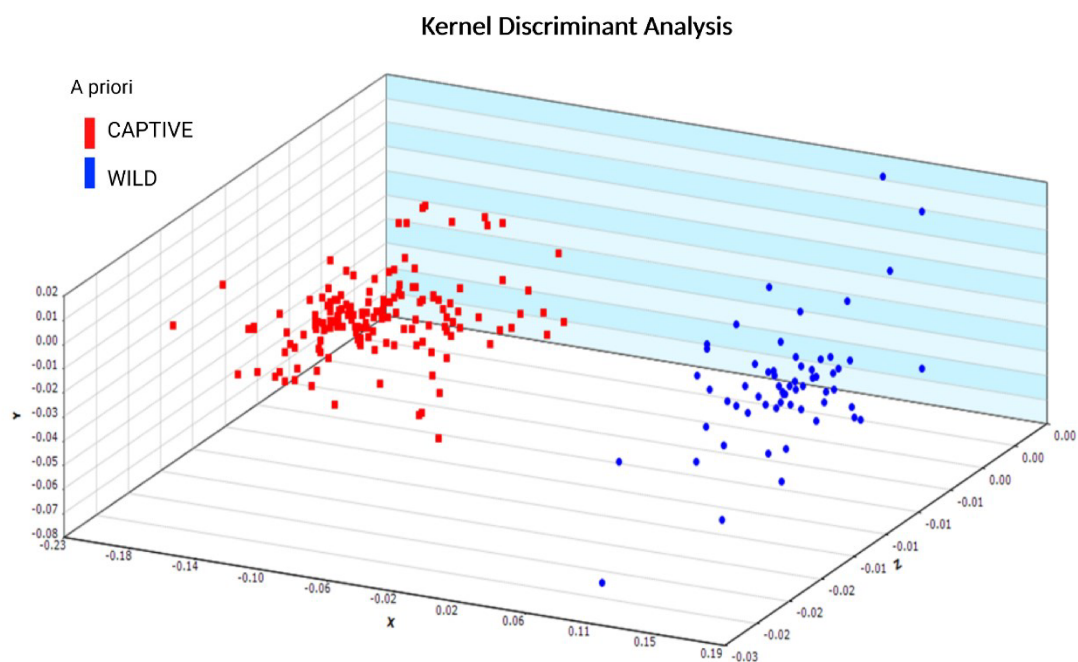


FIGURE 4

Example of fatty acids analyses. Differentiation of caviar from wild and farmed origin using KDA Kernel discriminant analysis of the fatty acid chemical fingerprint from the DART-TOF analysis performed on 54 samples, 38 of farmed and 16 of wild origin. All analyses were performed 4 times and results are independently reported in the graph. (Created with BioRender.com).

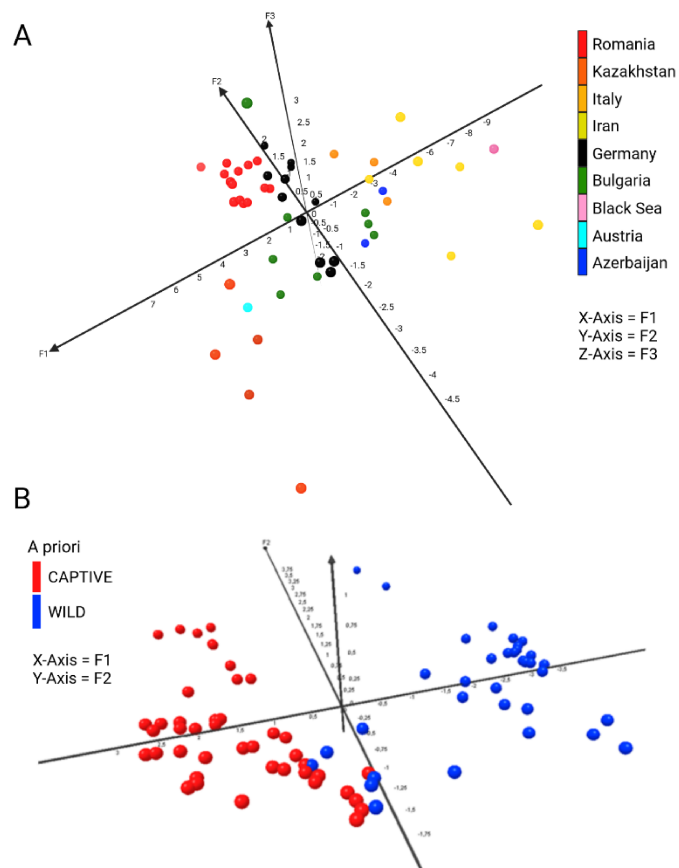


FIGURE 5

Example of Stable Isotope Ratio Analysis (SIRA) **A.** Agroisolab database of sturgeon caviar: $\delta^2\text{H}$ and $\delta^{18}\text{O}$ isotopic value (v.s. Vienna Standard Mean Ocean Water) of the tissue water and the organic fraction. **B.** PCA-Differentiation of caviar from wild / farmed origin using $\delta^{34}\text{S}$, $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ isotopic values. (Created with BioRender.com).

LIST OF REFERENCES

- Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature reviews genetics*, 11(10), 697-709.
- Arlati, G., & Poliakova, L. (2009). Restoration of Adriatic sturgeon (*Acipenser naccarii*) in Italy: situation and perspectives. In *Biology, Conservation and Sustainable Development of Sturgeons* (pp. 237-245). Springer, Dordrecht.
- Beridze, T., Edisherashvili, T., Anderson, C., & Scheele, F. (2021). Ship sturgeon rediscovered in the Rioni River in Georgia. *Oryx*, 55(1), 9-9.
- Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., [...] & Van Oosterhout, C. (2022). Genetic load: genomic estimates and applications in non-model animals. *Nature Reviews Genetics*, 23(8), 492-503.
- Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., [...] & Orti, G. (2017). Phylogenetic classification of bony fishes. *BMC evolutionary biology*, 17, 1-40.
- Billard, R., & Lecointre, G. (2000). Biology and conservation of sturgeon and paddlefish. *Reviews in Fish Biology and Fisheries*, 10, 355-392.
- Boscari, E., Barbisan, F., & Congiu, L. (2011). Inheritance pattern of microsatellite loci in the polyploid Adriatic sturgeon (*Acipenser naccarii*). *Aquaculture*, 321(3-4), 223-229.
- Boscari, E., & Congiu, L. (2014). The need for genetic support in restocking activities and ex situ conservation programmes: the case of the Adriatic sturgeon (*Acipenser naccarii* Bonaparte, 1836) in the Ticino River Park. *Journal of Applied Ichthyology*, 30(6), 1416-1422.
- Boscari, E., Barmintseva, A., Pujolar, J. M., Doukakis, P., Mugue, N., & Congiu, L. (2014a). Species and hybrid identification of sturgeon caviar: a new molecular approach to detect illegal trade. *Molecular ecology resources*, 14(3), 489-498.
- Boscari, E., Pujolar, J. M., Dupanloup, I., Corradin, R., & Congiu, L. (2014b). Captive breeding programs based on family groups in polyploid sturgeons. *PloS one*, 9(10), e110951.
- Boscari, E., Vidotto, M., Martini, D., Papetti, C., Ogden, R., & Congiu, L. (2015). 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). *Journal of Applied Ichthyology*, 31(6), 977-983.
- Boscari, E., Barmintseva, A., Zhang, S., Yue, H., Li, C., Shedko, S. V., [...] & Congiu, L. (2017a). Genetic identification of the caviar-producing Amur and Kaluga sturgeons revealed a high level of concealed hybridization. *Food Control*, 82, 243-250.
- Boscari, E., Vitulo, N., Ludwig, A., Caruso, C., Mugue, N. S., Suci, R., [...] & Congiu, L. (2017b). Fast genetic identification of the Beluga sturgeon and its sought-after caviar to stem illegal trade. *Food control*, 75, 145-152.
- Boscari, E., Marino, I. A., Caruso, C., Gessner, J., Lari, M., Mugue, N., [...] & Congiu, L. (2021). Defining criteria for the reintroduction of locally extinct populations based on contemporary and ancient genetic diversity: The case of the Adriatic Beluga sturgeon (*Huso huso*). *Diversity and Distributions*, 27(5), 816-827.
- Boscari, E., Wu, J., Jiang, T., Zhang, S., Cattelan, S., Wang, C., [...] & Wei, Q. (2022). The last giants of the Yangtze River: A multidisciplinary picture of what remains of the endemic Chinese sturgeon. *Science of The Total Environment*, 843, 157011.
- Bronzi, P., Chebanov, M., Michaels, J. T., Wei, Q., Rosenthal, H., & Gessner, J. (2019). Sturgeon meat and caviar production: Global update 2017. *Journal of Applied Ichthyology*, 35(1), 257-266.

- Cabrera-Castro, R., Zabala, C., Soriguer, M. C., Domezain, A., & Hernando, J. A. (2018).** Morphological development in the first life phase of Adriatic sturgeon *Acipenser naccarii* under controlled conditions. *Journal of Fish Biology*, 92(6), 1956-1974.
- Camacho-Sanchez, M., Quintanilla, I., Hawkins, M. T., Tuh, F. Y., Wells, K., Maldonado, J. E., & Leonard, J. A. (2018).** Interglacial refugia on tropical mountains: novel insights from the summit rat (*Rattus baluensis*), a Borneo mountain endemic. *Diversity and Distributions*, 24(9), 1252-1266.
- Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015).** Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science advances*, 1(5), e1400253.
- Ceballos, G., & Ehrlich, P. R. (2018).** The misunderstood sixth mass extinction. *Science*, 360(6393), 1080-1081.
- Českleba, D. G., AveLallemant, S., & Thuemler, T. F. (1985).** Artificial spawning and rearing of lake sturgeon, *Acipenser fulvescens*, in Wild Rose State Fish Hatchery, Wisconsin, 1982–1983. *Environmental biology of fishes*, 14, 79-85.
- Chow, S., & Hazama, K. (1998).** Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular ecology*, 7(9), 1255-1256.
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., [...] & Ruden, D. M. (2012).** A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *fly*, 6(2), 80-92.
- Clo, J., & Kolář, F. (2022).** Inbreeding depression in polyploid species: a meta-analysis. *Biology Letters*, 18(12), 20220477.
- Congiu, L., Pujolar, J. M., Forlani, A., Cenadelli, S., Dupanloup, I., Barbisan, F., [...] & Fontana, F. (2011).** Managing polyploidy in ex situ conservation genetics: the case of the critically endangered Adriatic sturgeon (*Acipenser naccarii*). *PloS one*, 6(3).
- Congiu, L., Boscari, E., Pagani, S., Gazzola, M., & Bronzi, P. (2021).** Resumption of natural reproduction of the Adriatic sturgeon in the River Po. *Oryx*, 55(6), 816-816.
- Congiu, L., Gessner, J., & Ludwig, A. (2023).** IUCN Red List reassessment reveals further decline of sturgeons and paddlefishes. *Oryx*, 57(1), 9-10.
- Cooper, G. M., Stone, E. A., Asimenos, G., Green, E. D., Batzoglou, S., & Sidow, A. (2005).** Distribution and intensity of constraint in mammalian genomic sequence. *Genome research*, 15(7), 901-913.
- Côrte-Real, H. B. S. M., Dixon, D. R., & Holland, P. W. H. (1994).** Intron-targeted PCR: a new approach to survey neutral DNA polymorphism in bivalve populations. *Marine biology*, 120, 407-413.
- Corti, R. (2022).** Moving toward stock units identification based on spatial population structure of marine species in the Mediterranean Sea and adjacent waters through multidisciplinary and holistic approaches for the sustainability of fisheries resources. University of Bologna. PhD dissertation.
- Corti, R., Piazza, E., Mokhtar-Jamaï, K., Masnadi, F., Armelloni, E.N., Scarcella, G. & Cariani, A. (2022).** Seascape genomics approach to describe population structure of two marine species: *Solea solea* and *Merluccius merluccius* case studies. Proceedings of the 9th Congress of the Italian Society for Evolutionary Biology (SIBE), 4-7 September 2022, Ancona (Italy).
- Cowie, R. H., Bouchet, P., & Fontaine, B. (2022).** The Sixth Mass Extinction: fact, fiction or speculation?. *Biological Reviews*, 97(2), 640-663.

- Daguin, C., & Borsa, P. (1999).** Genetic characterisation of *Mytilus galloprovincialis* Lmk. in North West Africa using nuclear DNA markers. *Journal of experimental marine biology and ecology*, 235(1), 55-65.
- Dalle Palle, S., Boscari, E., Bordignon, S. G., Muñoz-Mora, V. H., Bertorelle, G., & Congiu, L. (2022).** Different chromosome segregation patterns coexist in the tetraploid Adriatic sturgeon *Acipenser naccarii*. *Diversity*, 14(9), 745.
- DeSalle, R., & Amato, G. (2004).** The expansion of conservation genetics. *Nature reviews genetics*, 5(9), 702-712.
- Doekes, H. P., Bijma, P., & Windig, J. J. (2021).** How depressing is inbreeding? A meta-analysis of 30 years of research on the effects of inbreeding in livestock. *Genes*, 12(6), 926.
- Du, K., Stöck, M., Kneitz, S., Klopp, C., Woltering, J. M., Adolphi, M. C., [...] & Scharthl, M. (2020).** The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nature ecology & evolution*, 4(6), 841-852.
- Dufresne, F., Stift, M., Vergilino, R., & Mable, B. K. (2014).** Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Molecular ecology*, 23(1), 40-69.
- Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N., & Hoban, S. (2018).** Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary applications*, 11(7), 1035-1052.
- Fontana, F., Lanfredi, M., Chicca, M., Congiu, L., Tagliavini, J., & Rossi, R. (1999).** Fluorescent in situ hybridization with rDNA probes on chromosomes of *Acipenser ruthenus* and *Acipenser naccarii* (Osteichthyes Acipenseriformes). *Genome*, 42(5), 1008-1012.
- Fontana, F., Zane, L., Pepe, A., & Congiu, L. (2007).** Polyploidy in *Acipenseriformes*: cytogenetic and molecular approaches. *Fish cytogenetics*, 385, 403.
- Fontana, F., Congiu, L., Mudrak, V. A., Quattro, J. M., Smith, T. I., Ware, K., & Doroshov, S. I. (2008).** Evidence of hexaploid karyotype in shortnose sturgeon. *Genome*, 51(2), 113-119.
- Forcina, G., Camacho-Sanchez, M., Tuh, F. Y., Moreno, S., & Leonard, J. A. (2021).** Markers for genetic change.
- Fox, D. T., Soltis, D. E., Soltis, P. S., Ashman, T. L., & Van de Peer, Y. (2020).** Polyploidy: a biological force from cells to ecosystems. *Trends in Cell Biology*, 30(9), 688-694.
- Frankham, R., Ballou, S. E. J. D., Briscoe, D. A., & Ballou, J. D. (2002).** Introduction to conservation genetics. Cambridge university press.
- Gemenet, D. C., Lindqvist-Kreuzer, H., De Boeck, B., da Silva Pereira, G., Mollinari, M., Zeng, Z. B., ... & Campos, H. (2020).** Sequencing depth and genotype quality: accuracy and breeding operation considerations for genomic selection applications in autopolyploid crops. *Theoretical and Applied Genetics*, 133, 3345-3363.
- Green, B. S. (2008).** Maternal effects in fish populations. *Advances in marine biology*, 54, 1-105.
- Hathaway, N. J., Parobek, C. M., Juliano, J. J., & Bailey, J. A. (2018).** SeekDeep: single-base resolution de novo clustering for amplicon deep sequencing. *Nucleic acids research*, 46(4), e21-e21.
- Havelka, M., Kašpar, V., Hulák, M., & Flajšhans, M. (2011).** Sturgeon genetics and cytogenetics: a review related to ploidy levels and interspecific hybridization. *Folia Zoologica*, 60(2), 93-103.

- Havelka, M., Hulák, M., Bailie, D. A., Prodöhl, P. A., & Flajšhans, M. (2013).** Extensive genome duplications in sturgeons: new evidence from microsatellite data. *Journal of Applied Ichthyology*, 29(4), 704-708.
- Hedrick, P. W. (2001).** Conservation genetics: where are we now?. *Trends in Ecology & Evolution*, 16(11), 629-636.
- Hu, Y., Liu, X., Yang, J., Xiao, K., Wang, B., & Du, H. (2019).** Development of 17 novel cross-species microsatellite markers for Dabry's Sturgeon (*Acipenser dabryanus*) from Chinese Sturgeon (*Acipenser sinensis*) via next-generation sequencing. *Pak. J. Zool*, 51, 2381-2384.
- Huang, Z. (2019).** Drifting with Flow versus Self-Migrating—How Do Young Anadromous Fish Move to the Sea?. *Iscience*, 19, 772-785.
- Huber, C. D., Kim, B. Y., & Lohmueller, K. E. (2020).** Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. *PLoS genetics*, 16(5), e1008827.
- Jombart, T. (2008).** adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403-1405.
- Jombart, T., Devillard, S., & Balloux, F. (2010).** Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, 11(1), 1-15.
- Keller, L. F., & Waller, D. M. (2002).** Inbreeding effects in wild populations. *Trends in ecology & evolution*, 17(5), 230-241.
- Kim, D. S., Nam, Y. K., Noh, J. K., Park, C. H., & Chapman, F. A. (2005).** Karyotype of North American shortnose sturgeon *Acipenser brevirostrum* with the highest chromosome number in the *Acipenseriformes*. *Ichthyological Research*, 52, 94-97.
- Kim, E. J., Park, C., & Nam, Y. K. (2020).** Behavioral characteristics of a chondrosteian sturgeon species *Acipenser baerii* prelarvae in response to different environmental light intensities in a diel photoperiodic cycle. *Journal of Animal Reproduction and Biotechnology*, 35(3), 245-257.
- Krieger, J., & Fuerst, P. A. (2002).** Evidence for a slowed rate of molecular evolution in the order *Acipenseriformes*. *Molecular Biology and Evolution*, 19(6), 891-897.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018).** MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, 35(6), 1547.
- Lessa, E. P. (1992).** Rapid surveying of DNA sequence variation in natural populations. *Molecular Biology and Evolution*, 9(2), 323-330.
- Ludwig, A., Belfiore, N. M., Pitra, C., Svirsky, V., & Jenneckens, I. (2001).** Genome duplication events and functional reduction of ploidy levels in sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). *Genetics*, 158(3), 1203-1215.
- Ludwig, A. (2008).** Identification of *Acipenseriformes* species in trade. *Journal of applied ichthyology*, 24, 2-19.
- Ludwig, A., Lieckfeldt, D., & Jahrl, J. (2015).** Mislabelled and counterfeit sturgeon caviar from Bulgaria and Romania. *Journal of Applied Ichthyology*, 31(4), 587-591.
- Marino, I. A. M., Benazzo, A., Agostini, C., Mezzavilla, M., Hoban, S. M., Patarnello, T., [...] & Bertorelle, G. (2013).** Evidence for past and present hybridization in three Antarctic icefish species provides new perspectives on an evolutionary radiation. *Molecular ecology*, 22(20), 5148-5161.
- Matias, F. I., Xavier Meireles, K. G., Nagamatsu, S. T., Lima Barrios, S. C., Borges do Valle, C., Carazzolle, M. F., ... & Endelman, J. B. (2019).** Expected genotype quality

- and diploidized marker data from genotyping-by-sequencing of *Urochloa* spp. Tetraploids. *The plant genome*, 12(3), 190002.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... & DePristo, M. A. (2010).** The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), 1297-1303.
- Meirmans, P. G., Liu, S., & van Tienderen, P. H. (2018).** The analysis of polyploid genetic data. *Journal of Heredity*, 109(3), 283-296.
- Mirrasooli, E., Ghorbani, R., Gorgin, S., Aghilinejad, S. M., & Jalali, A. (2019).** Factors associated with illegal fishing and fisher attitudes toward sturgeon conservation in the southern Caspian Sea. *Marine Policy*, 100, 107-115.
- Mukai, T., Chigusa, S. I., Mettler, L. E., & Crow, J. F. (1972).** Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics*, 72(2), 335-355.
- Nguyen, R. M., & Crocker, C. E. (2006).** The effects of substrate composition on foraging behavior and growth rate of larval green sturgeon, *Acipenser medirostris*. *Environmental biology of fishes*, 76, 129-138.
- Ortega, J., & Maldonado, J. E. (2020).** Conservation Genetics in Mammals. Integrative Research Using Novel Approaches. Cham: Springer International Publishing.
- Osborne, M. J., Caeiro-Dias, G., & Turner, T. F. (2023).** Transitioning from microsatellites to SNP-based microhaplotypes in genetic monitoring programmes: Lessons from paired data spanning 20 years. *Molecular Ecology*, 32(2), 316-334.
- Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K., & Hedrick, P. W. (2010).** Conservation genetics in transition to conservation genomics. *Trends in genetics*, 26(4), 177-187.
- Palumbi, S. R., & Baker, C. S. (1994).** Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution*, 11(3), 426-435.
- Peng, Z., Ludwig, A., Wang, D., Diogo, R., Wei, Q., & He, S. (2007).** Age and biogeography of major clades in sturgeons and paddlefishes (Pisces: Acipenseriformes). *Molecular phylogenetics and evolution*, 42(3), 854-862.
- Pertoldi, C., Bijlsma, R., & Loeschcke, V. (2007).** Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodiversity and Conservation*, 16, 4147-4163.
- Rajkov, J., Shao, Z., & Berrebi, P. (2014).** Evolution of polyploidy and functional diploidization in sturgeons: microsatellite analysis in 10 sturgeon species. *Journal of Heredity*, 105(4), 521-531.
- Raymakers, C. (2002).** International trade in sturgeon and paddlefish species—the effect of CITES listing. *International Review of Hydrobiology: A Journal Covering all Aspects of Limnology and Marine Biology*, 87(5-6), 525-537.
- Schiavon, L., Dulière, V., La Mesa, M., Marino, I. A. M., Codogno, G., Boscari, E., [...] & Papetti, C. (2021).** Species distribution, hybridization and connectivity in the genus *Chionodraco*: Unveiling unknown icefish diversity in antarctica. *Diversity and Distributions*, 27(5), 766-783.
- Schreier, A. D., Van Eenennaam, J. P., Anders, P., Young, S., & Crossman, J. (2021).** Spontaneous autopolyploidy in the Acipenseriformes, with recommendations for management. *Reviews in Fish Biology and Fisheries*, 31, 159-180.
- Schweizer, R. M., Saarman, N., Ramstad, K. M., Forester, B. R., Kelley, J. L., Hand, B. K., [...] & Luikart, G. (2021).** Big data in conservation genomics: boosting skills, hedging bets, and staying current in the field. *Journal of Heredity*, 112(4), 313-327.

- Sheraliev, B., Rozimov, A., Ludwig, A., & Peng, Z. (2021).** Rediscovery of rare shovelnose sturgeons in the Amu Darya River, Uzbekistan. *Oryx*, 55(3), 332-332.
- Soares, N. R., Mollinari, M., Oliveira, G. K., Pereira, G. S., & Vieira, M. L. C. (2021).** Meiosis in polyploids and implications for genetic mapping: a review. *Genes*, 12(10), 1517.
- Sørensen, T. A. (1948).** A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol. Skar.*, 5, 1-34.
- Stift, M., Berenos, C., Kuperus, P., & van Tienderen, P. H. (2008).** Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to *Rorippa* (yellow cress) microsatellite data. *Genetics*, 179(4), 2113-2123.
- Sulak, K. J., & Randall, M. (2002).** Understanding sturgeon life history: enigmas, myths, and insights from scientific studies. *Journal of Applied Ichthyology*, 18(4-6), 519-528.
- Thioulouse, J., Dray, S., Dufour, A. B., Siberchicot, A., Jombart, T., & Pavoine, S. (2018).** Multivariate analysis of ecological data with ade4.
- Tortonese, E. (1989).** *Acipenser naccarii* Bonaparte, 1836. The freshwater fishes of Europe, 1(Part II), 285-293.
- Wang, J., Sun, Z., Jiang, L., & Hu, Y. (2022).** Developing microsatellite duplex PCR reactions for sterlet (*Acipenser ruthenus*) and their application in parentage identification. *Scientific Reports*, 12(1), 12036.
- We, Q. (2020).** Conservation of Chinese sturgeon (*Acipenser sinensis*) based on its life history: Dilemma and breakthrough. *J. Lake Sci*, 32(5), 1297-1319.
- White, S. L., Kazyak, D. C., Darden, T. L., Farrae, D. J., Lubinski, B. A., Johnson, R. L., [...] & Wirgin, I. I. (2021).** Establishment of a microsatellite genetic baseline for North American Atlantic sturgeon (*Acipenser o. oxyrinchus*) and range-wide analysis of population genetics. *Conservation Genetics*, 22, 977-992.
- Wickham, H. (2016).** *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA; Available online: <https://ggplot2.tidyverse.org>
- Williot, P., Arlati, G., Chebanov, M., Gulyas, T., Kasimov, R., Kirschbaum, F., [...] & Zholdasova, I. M. (2002).** Status and management of Eurasian sturgeon: an overview. *International Review of Hydrobiology: A Journal Covering all Aspects of Limnology and Marine Biology*, 87(5-6), 483-506.
- Wu, J. M., Wang, C. Y., Zhang, H., Du, H., Liu, Z. G., Shen, L., [...] & Rosenthal, H. (2015).** Drastic decline in spawning activity of Chinese sturgeon *Acipenser sinensis* Gray 1835 in the remaining spawning ground of the Yangtze River since the construction of hydrodams. *Journal of Applied Ichthyology*, 31(5), 839-842.
- Yue, H., Li, C., Du, H., Zhang, S., & Wei, Q. (2015).** Sequencing and de novo assembly of the gonadal transcriptome of the endangered Chinese sturgeon (*Acipenser sinensis*). *PLoS One*, 10(6), e0127332.
- Zhang, H., Yang, D. G., Wei, Q. W., Du, H., & Wang, C. Y. (2013).** Spatial distribution and spawning stock estimates for adult Chinese sturgeon (*Acipenser sinensis* Gray, 1835) around the only remaining spawning ground during the trial operation of the newly constructed Three Gorges Project in the Yangtze River, China. *Journal of Applied Ichthyology*, 29(6), 1436-1440.
- Zhang, H., Wu, J. M., Wang, C. Y., Du, H., Liu, Z. G., Shen, L., [...] & Wei, Q. W. (2016).** River temperature variations and potential effects on fish in a typical Yangtze River reach: implications for management. *Appl. Ecol. Environ. Res*, 14, 553-567.

- Zhang, H., Jarić, I., Roberts, D. L., He, Y., Du, H., Wu, J., [...] & Wei, Q. (2020).** Extinction of one of the world's largest freshwater fishes: Lessons for conserving the endangered Yangtze fauna. *Science of the Total Environment*, 710, 136242.
- Zhou, J., Zhao, Y., Song, L., Bi, S., & Zhang, H. (2014).** Assessing the effect of the Three Gorges reservoir impoundment on spawning habitat suitability of Chinese sturgeon (*Acipenser sinensis*) in Yangtze River, China. *Ecological Informatics*, 20, 33-46.

ACKNOWLEDGEMENTS

I want to thank all the people who helped me during my Ph.D.

First of all, I would like to thank my supervisor Leonardo Congiu for the support and help received from the beginning to the end of this Ph.D.

I also want to thank Elisa Boscari and Ilaria Marino, who have accompanied me and taught me from the beginning.

I thank my advisors Alessandro Grapputo and Andrea Pilastro for their useful advice and Lorenzo Zane and Chiara Papetti for their help and kindness.

I especially thank Silvia Cattelan and Federica Poli for their support and advice both inside and outside the laboratory.

I want to thank all of my current and past laboratory mates for their help and the time we spent together: Federica Barca, Chiara Samassa, Simone Giulio Bordignon, Annalisa Scapolatiello, Irene Gregori, Francesco Martino, and Luca Schiavon.

I want to thank Giorgio Bertorelle and the members of the genomics group in Ferrara for the help and support for the bioinformatic analyses: Maëva Gabrielli, Victor Muñoz, Silvia Fuselli, and Andrea Benazzo.

Finally, I want to thank my family and all my friends who have supported me during these years.