UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UNIVERSITA’ DEGLI STUDI DI PADOVA

Dipartimento di Biologia

Scuola di dottorato di ricerca in Bioscienze e Biotecnologie

Indirizzo BIOLOGIA EVOLUZIONISTICA

Ciclo XXVI

Conservation genetics of Adriatic sturgeon (*Acipenser naccarii*)

Genetica della conservazione dello storione Cobice (*Acipenser naccarii*)

Direttore della Scuola : Ch.mo Prof. Giuseppe Zanotti

Coordinatore d’indirizzo: Ch.mo Prof. Andrea Pilastro

Supervisore :Ch.mo Dr. Leonardo Congiu

Co-Supervisore :Ch.mo Prof. Lorenzo Zane

Dottoranda: Elisa Boscari

Contents

List of manuscripts 5

Extended Abstract 7

1. General Introduction 11

1.1 Preamble 11

1.2 Conservation Genetics 12

1.3 Sturgeons and Conservation 14

1.4 The Adriatic Sturgeon (*Acipenser naccarii*) 16

1.5 Background on the status of the species 17

2. Aim of the Work 19

3. Results and Discussion 21

3.1 inheritance pattern 21

3.2 Sturgeon species and hybrids identification 23

3.3 Evaluation of past management activities 27

3.4 Characterization of the V.I.P. stock 31

3.5 Breeding Plan 35

4.5.1 Short-term breeding plan (F0xF0) 35

4.5.2 Long-term breeding plan (F1xF1) 36

4.5.3 Sustainability of the breeding plan 37

4. Conclusions 39

5. Other Publications 41

5.1 Transcriptome sequencing of the Adriatic sturgeon 41

5.2 Analysis of transposable elements in Acipenser naccarii 41

Annex A 43

Annex B 46

Annex C 52

Annex D 64

Appendix I 82

Appendix II 84

6. Riassunto 86

Introduzione 86

Risultati e Discussione 86

7. Reference List 90

Acknowledgements 98

LIST OF MANUSCRIPTS

The main results of the present study are reported in four manuscripts two of which were published on international journals, one is accepted and one in an advanced stage of drafting. The four papers, listed here below, are reported as annexes.

ANNEX A

Inheritance pattern of microsatellite loci in the polyploidy Adriatic sturgeon (*Acipenser naccarii*)

Elisa Boscari, Federica Barbisan, Leonardo Congiu

Aquaculture 321 (2011): 223-229

ANNEX B

Species and hybrid identification of sturgeon caviar: a new molecular approach to detect illegal trade

Elisa Boscari, Anna Barmintseva, Josè Martin Pujolar, Phaedra Doukakis, Nikolai Mugue, Leonardo Congiu

Molecular Ecology resources (2014), DOI: 10.1111/1755-0998.12203 *In press*

ANNEX C

The need for genetic support in restocking activities and *ex situ* conservation programs: the case of the Adriatic sturgeon in the Ticino River Park.

Elisa Boscari and Leonardo Congiu

Accepted on Journal of Applied Ichthyology (2014) *Accepted*

ANNEX D

Captive breeding program for the critically endangered Adriatic sturgeon (Acipenser naccarii).

Elisa Boscari, Isabelle Dupanloup, Riccardo Corradin, Leonardo Congiu

Manuscript

OTHER PUBLICATIONS

During the last three years I have been involved also in other investigations on the Adriatic sturgeon, although not directly focused on conservation aspects. The resulting publications, listed here below, are reported in Appendix I and II.

APPENDIX I

Transcriptome sequencing and *de novo* annotation of the critically endangered Adriatic sturgeon

Michele Vidotto, Alessandro Grapputo, Elisa Boscari, Federica Barbisan, Alessandro Coppe, Gilberto Grandi, Abhishek Kumar, Leonardo Congiu

BMC Genomics (2013) 14: 407

APPENDIX II

***Tana1,* a new putatively active *Tc1*-like transposable element in the genome of sturgeons**

Josè Martin Pujolar, Laura Astolfi, Elisa Boscari, Michele Vidotto, Federica Barbisan, Alice Bruson, Leonardo Congiu.

Molecular Phylogenetics and Evolution (2013) 66: 223-232.

EXTENDED ABSTRACT

Captive breeding represents the last chance of survival for many species faced with imminent extinction in the wild (Russello and Amato, 2004). The management of the Adriatic sturgeon (*Acipenser naccarii*), a highly imperiled species endemic of the Adriatic Sea, is strictly related to the *ex situ* conservation of the remnant genetic variability. The last remaining population of Adriatic sturgeon of certain wild origin (50 F0 individuals) is maintained in captivity since 1977 in a private aquaculture plant, named V.I.P.. This stock is currently decimated to only 13 breeders (Boscari and Congiu*,* 2014). Few alternative brood-stocks of F1 individuals, originated from the same F0 stock, have been retained from local administrations to be used in a near future and part of them were randomly chosen to be released in the wild disregarding their genetic composition. The consequences of conducting breeding programs without a genetic input was evaluated in the Ticino River Park (TRP), in Italy, where Adriatic sturgeon was restocked in the last years (Boscari and Congiu*,* 2014). The very low heterogeneity observed in TRP-F1 sample analyzed (Stock_1 and Stock_2) strongly suggests the need of a complete reorganization and coordination of the conservation efforts conducted on this species, paying special attention to the long-term preservation of the available genetic diversity based on more heterogeneous stocks. To this purpose, about 450 animals (V.I.P. F1 Stock), retained in captivity at the V.I.P. plant and obtained from several reproductions performed in the past 20 years, were analyzed. Currently, most of them have reached the sexual maturity and could be used as “new” breeders in captive programs for *ex situ* conservation.

The genetic variability of all F1 available was estimated based on mitochondrial and microsatellite analyses and compared with the variability of the F0 stock, taking into account the tetrasomic inheritance pattern observed (Boscari *et al.,* 2011). To improve the performance of the Band-sharing method used for parental allocation, a new tool for pedigree analysis was developed allowing the estimation for each individuals of a compatibility index for all putative parent pairs.

The pedigree analysis showed that the V.I.P. F1 Stock retained a relevant part of the genetic variability due to the several F0 parents used as founders. Moreover, the distribution of pairwise genetic distances was used to estimate threshold values of distance above which a

given degree of relatedness could be significantly excluded thus inferring putative familiar groups among animals with unknown pedigree.

To detect possible contamination within stocks of pure species, a new molecular marker for species and hybrids identification was also set up based on species-specific SNPs in the first intron of the nuclear S7 gene (Boscari *et al.,* 2014). The method was applied to all not-allocated animals; hybrids between *A. gueldenstaedtii* or *A. transmontanus* and *A. naccarii* were detected and excluded from the "*breeders* unit" used to planning the breeding scheme.

A long-term brood-stock management plan was proposed for the Adriatic sturgeon based on a breeding strategy optimized considering the tetraploid genome, the long life cycle and the high costs for aquaculture.

A two-step strategy is proposed in which a first short-term breeding program, relying on the only 13 remnant individuals of certain wild origin, is followed by a second series of crosses among already mature F1 families. In the latter case, F1 families were opportunely selected to form the *"breeders unit"* from the V.I.P. Stock, the major source of variability for this species, and the Stock_1. This action also encourages a strict collaboration among different aquaculture plants to plan more coordinated strategies.

The breeding program here proposed is not based on single individuals but on familiar groups instead. A priority order to follow for reproduction was given to each family ("priority family"). Then, estimation of genetic distances among familiar groups, genotyped at 24 microsatellite loci, allowed to select the different family combinations for the recruitment of individuals ("mating family") to be paired with the "priority family".

In order to test different strategies in pairing the mating families and with the aim of assessing optimal number of breeders per family to reproduce, simulations were performed under different scenarios and the efficiency of allele transmission was evaluated in F2 virtual generations. Economical and logistical aquaculture constraints were also simulated to evaluate the feasibility of the program and the necessary economical support.

Results generally suggest that, in the short-term, the best strategy to maximize the retention of genetic variation present in the F0 parental stock is the selection of the "mating family" based on genetic distance. Simulations showed that the recommended number of individuals to reproduce per family is three. Moreover, detailed information about the number of individuals to release were also reported.

This work represents the first involved in the planning of a coordinated strategy for the successful long-term preservation of the Adriatic sturgeon. Additionally, the plan here proposed is in line with the general consideration of the *ex situ* conservation as a temporary solution with the final aim to reach self-sustaining population to release.

Finally, this strategy might be an example to plan breeding schemes for other tetraploid sturgeon species.

1. General Introduction

1.1 Preamble

The present work arises from the need of an urgent and specific captive breeding plan for the safeguard of the critically endangered and possibly extinct in the wild Adriatic sturgeon (*Acipenser naccarii*). Overfishing, introduction of alien species and habitat degradation have caused the drastic decline of natural populations of this species, which represented an important endemism of the Adriatic region, since 1996 listed in the Red List of the IUCN.
Despite the several restocking programs performed in the last decades, the status of the Adriatic sturgeon is not significantly improved and no evidences of natural reproductions have been reported in the last 20 years
Today, the recovery of the *A. naccarii* mostly depends on the correct management of captive F1 stocks obtained by crosses, performed in the last 30 years, among individuals of the last captive stock of certain wild origin.
Since no further inputs from the wild are possible, and since all the available breeders directly descend from a limited number of wild animals, the correct management of the residual genetic diversity represents a priority task. To this purpose, the artificial reproduction should be performed after planning a breeding scheme aimed to maximise the fraction of genetic diversity transmitted to the following generations.
In order to propose a suitable breeding program for the Adriatic sturgeon, main goal of the present thesis, different aspects must be taken into account, such as the tetraploid level of the species, its biological features well as economical and logistical constraints.
The breeding plan here proposed should be considered as a solid starting point for needed actions that must necessarily compose a long term conservation program and that, besides the restocking activities, also include environmental restorations, monitoring strategies, and building of public awareness.
The results of this thesis could be considered as example for other tetraploid sturgeon species for which similar efforts are required.

> "*This widespread evidence of decline alerts us to the fact that despite protective fishing regulations, sturgeon are still in trouble due to historic fishing, current illegal fishing and habitat degradation*," says Dr Kent Carpenter, IUCN Global Marine Species Assessment Director. "*This is alarming given their unique lineage and particular vulnerability.*" (IUCN)

> "*Sturgeon have survived dramatic change over the past 250 million years only to face the serious threat of becoming extinct as a direct result of human activities. Illegal catch, over fishing, the breaking up of the migratory routes and pollution are the key elements that have driven almost all species to the brink of extinction,*" says Dr Mohammad Pourkazemi, chair IUCN/SSC Sturgeon Specialist Group. "*The latest Red List assessment shows an increasingly alarming status of natural populations.*" (IUCN)

1.2 Conservation genetics

During the last decades the application of genetic theories and techniques in elucidating aspects of species' biology relevant to conservation measures (*e.g.* habitat fragmentation, species adaptation, inbreeding depression) (Hedrick 2001; DeSalle and Amato, 2004; Pertoldi *et al.*, 2007) is becoming ever more important. Understanding the processes that lead to the loss of genetic variation represents an important step towards promoting adequate strategies to reduce the risk of extinction of endangered species. Conservation genetics, whose "*central dogma*" is centred on the preservation of the amount of genetic variability and thus the preservation of an important fraction of biodiversity (Witzenberger and Hochkirch, 2011), is a scientific discipline originated from the strict contribution of genetics and ecology with conservation biology (Pertoldi *et al.*, 2007; Van Dyke 2008).

The multiple causes that lead to the extinction of a species can be divided into two types, that identify two non-mutually exclusive categories of causes of the extinction process (Leus, 2011).

The first includes causes mainly associated with human impact. The human-threats can involve directly the species such as fishing, exploitation for commercial purposes and competition with exotic species artificially introduced in the wild; but more indirect threats can also impact natural environments such as pollution, habitat degradation and fragmentation.

The second category of causes of extinction is not linked to human activities; it is mainly represented by stochastic demographic and genetic processes such as genetic drift, inbreeding and natural environmental variations, which could cause fitness reduction or high mortality. Moreover, populations already small are much more susceptible to these events. In this case, additional management activities must be applied to support the *in situ* conservation, which alone would not be sufficient to avoid extinction (Leus, 2011; Witzenberger and Hochkirch, 2011). In this context, captive breeding (*ex situ* activities) becomes a fundamental approach for the safeguard of imperilled species (Fig. 1) and, in some cases, is considered as the last chance of survival for many species faced with imminent extinction in the wild (Russello and Amtao, 2004; Araki *et al.,* 2007; Frankham, 2008; Leus 2011).

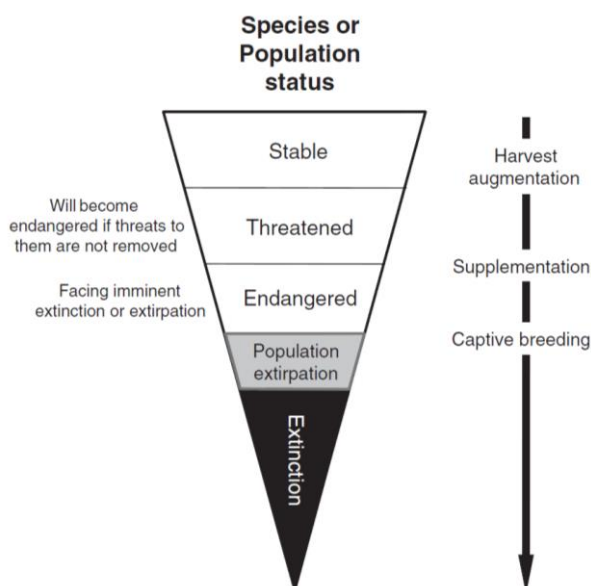


Figure 1: Schematic view of actions required in relation to the status of a specie or a population (Fraser, 2008).

In 2008, Frankham (2008) has estimated that the number of terrestrial vertebrates

species that would require a captive managing was around 2000-3000.
A standard captive breeding plan should be aimed to preserve the maximum proportion of gene diversity of the wild species in captive stocks (Frankham *et al.,* 2002; Leus 2011). To reach this goal, each plan must be based primarily on three criteria.
The first is the "*founder assumption*" that considers the group of founders as unrelated individuals and as representatives of the species (Russello and Amato, 2004; Rudnick and Lacy, 2008; Lacy 2012). The second criterion is that populations used for captive breeding must be sufficiently large. Crow and Kimura (1970) have estimated that a 97.5% of gene diversity of a population could be retained starting from 20/30 founders. The third criteria recommends a careful planning of the mating scheme (Leus 2011).
To carry out a correct planning of a breeding program, two main goal must be satisfied: the establishment of an adequate "*breeders unit*" in which the genetic diversity of the species (or population) is best represented (Goldenstein *et al.,* 2000), and the choice of breeding pairs and the definition of their priorities (Myers *et al.,* 2001). To this purposes a complete genetic characterization of all individuals available is necessary to perform an estimation of relatedness among individuals and a reconstruction of pedigrees for animals obtained by past reproductions performed in captivity (Kozfkay *et al.*, 2008; Congiu *et al.*, 2011). Molecular markers are generally used to this purpose and they allow also to infer possible relationships among animals with unknown ancestry (Russello and Amato, 2004).
Finally, considering the recommendation to use captive breeding only as a temporary measure (Philippart 1995; Snyder *et al.,* 1996), the new generations should be managed with the final aim to maximize the adaptive success of the individuals in the case of releases in adequate target environments, generating self-sustaining populations (Fraser 2008; Kozfkay *et al.,* 2008; Tzika *et al.,* 2009; Boscari and Congiu*,* 2014).

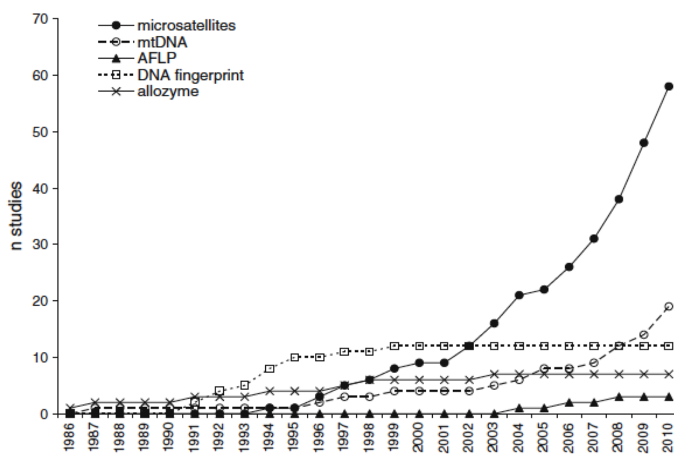


Figure 2: Cumulative number of studies with different molecular markers. The use of microsatellites started during the 1990s and then is exponentially increased (Witzenberger and Hochkirch, 2011).

In the last decades, the introduction of new techniques for genetic analysis such as *randomly amplified polymorphic DNA* (RAPD), *restriction fragment length polymorphisms* (RFLP), *amplified fragmented length polymorphisms* (AFLP), DNA sequencing, microsatellites and SNPs has significantly improved the efficiency of the genetic characterization. In detail, microsatellites are the most widely used markers for animal conservation (Fig. 2) (DeSalle and Amato, 2004; Russello and Amato, 2004; Rodzen *et al.,* 2004; Fopp-Bayat and Luczynski, 2006; Hayes *et al.,* 2006; Witzenberger

and Hochkirch, 2011). Many studies involved in parentage reassignment in aquaculture brood-stocks showed a 90% correct parental allocation using microsatellite data with over 100 possible parental pairs (Garcia de Leon *et. al,* 1998; Estoup *et al.,* 1998; Waldbeiser and Wolters, 1999; O'Reilly *et al.,* 1998; Norris *et al.,* 2000).
Unfortunately, no genetic guidelines were followed during some past rehabilitation programs (Teichert-Coddington and Smitherman, 1988; Huanh and Liao, 1990). The absence of an adequate genetic support to rehabilitation program might cause a detrimental impact on the residual variability of a species probably due to a biased selection of animals (*e.g.* employing a low number of breeders or mating related animals).
This is the case of sturgeons (Hayes *et al.,* 2006; Doukakis *et al.*, 2009).

1.3 Sturgeons and conservation

Sturgeons (order Acipenseriformes, superorder Chondrostei) are an ancient fish group of the northern hemisphere (Fig. 3), composed by about 25 species, considered evolutionary, ecologically and commercially important (Birstein and DeSalle, 1998; Anders *et al.*, 2011).
The evolutionary relevance derives from their key phylogenetic position with a separation from teleosts over 200 Mya and see them as one of the most archaic lineages of vertebrates (Bemis and Kynard, 1997; Ludwig *et al.,* 2001; Fontana *et al.,* 2001; Peng *et al.,* 2007; Krieger *et al.,* 2008). Moreover, their primitive characteristics, which allow naming them as "*Living Fossils*", are of particular interest from an evolutionary point of view (Billard and Lecointre, 2001; Peng *et al.,* 2007).
From a commercial point of view, these animals produce one of the most valuable and exclusive food commercialized worldwide: the caviar, composed by unfertilized eggs (Ludwig, 2008). Besides its tasteful delicacy, the high economic value is justified by the rarity of the producer species.

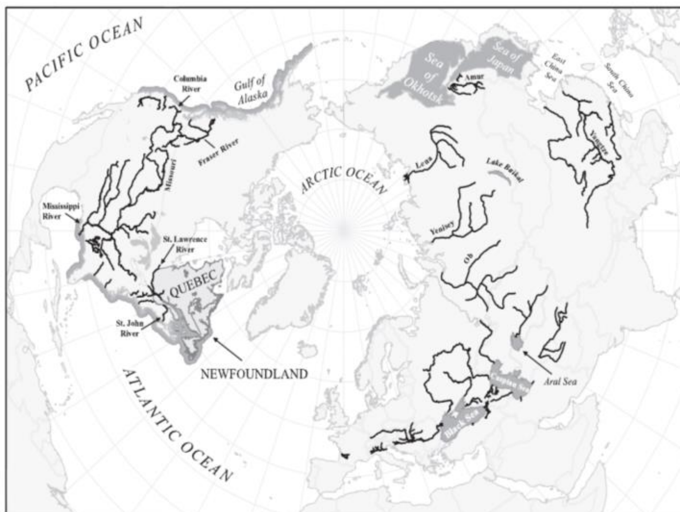


Figure 3: Distribution of sturgeons and paddlefish in the northern hemisphere (Pikitch *et al.,* 2005)

A recent report suggested that 85% of world's sturgeon species are currently at risk of extinction, making them the most threatened group of animals on the *Red List of threatened species* of the *International Union for Conservation of Nature* (IUCN). For these species the drastic decline is probably due to an over-exploitation for caviar production along with, in some cases, significant habitat degradation (Bronzi *et al.,* 2011).

“Sturgeon more critically endangered than any other group of species

18 March 2010 | International news release by IUCN

Eighty five percent of sturgeons, one of the oldest families of fishes in existence, valued around the world for their precious roe, are at risk of extinction, making them the most threatened group of animals on the IUCN Red List of Threatened Species™. The latest update of the Red List assessed the status of 18 species of sturgeon from all over Europe and Asia and found that all were threatened.”

As a result of the continuous deterioration of the status of the sturgeon species, several efforts to minimize this negative trend and many measures to preserve the remnant populations were considered in the last decades. At the same time, several different countries have shown a considerable increase of aquaculture practices. In Europe, mainly in France, Italy and Germany, a successful captive rearing was performed for *A. transmontanus, A. baerii, A. naccarii* and *A. ruthenus A. gueldenstaedtii* (Bronzi *et al.,* 1999; Williot *et al.,* 2001; Nathanailides *et al.,* 2002).

The number of studies on conservation biology has rapidly grown and for many sturgeon species several efforts to produce guidelines for their safeguard and to plan specifically *in situ* and *ex situ* recovery actions were developed. As an example, since 1999 China started the action plan for brood-stock collection for Kaluga (*H. dauricus*) and Amur sturgeon (*A. schrenckii*) in order to reduce the heavy fishing pressure on natural populations (Wei *et al.,* 2004). Similarly, 40 wild specimens per year of the Chinese sturgeon (*A. sinensis*) were collected in the Yangtze River to establish captive breeding units for subsequent releasing activities (Wei *et al.,* 2004; Wang *et al.,* 2011). Other examples are the restoration efforts for the Atlantic sturgeon (*A. oxyrinchus*) (Waldam and Wirgin, 1997), the conservation program involving the white sturgeon (*A. transmontanus)* in the Kootenai River (USA) started in 1990 (Ireland *et al.,* 2002) and the Adriatic sturgeon (*A. naccarii*) in the Po River (Italy) started in 1977 (Bronzi *et al.,* 1999; Congiu *et al.,* 2011).

Despite these actions, information about the real status of most species is still scarce and fragmented.

1.4 The Adriatic sturgeon (*Acipenser naccarii*)

2.2.1 Systematic classification:

PHYLUM	*Chordata*
SUBPHYLUM	*Vertebrata*
CLASS	*Osteichthyes*
SUBCLASS	*Actinopterygii*
SUPERORDER	*Chondrostei*
ORDER	*Acipenseriformes*
FAMILY	*Acipenseridae*
SUBFAMILY	*Acipenseinae*
GENUS	*Acipenser*

Figure 4: Adriatic sturgeon (*Acipenser naccarii*)

1.4.1 Distribution and status of the species

The Adriatic sturgeon (*Acipenser naccarii*) is a tetraploid species representing an important endemism of the Adriatic region where, two additional sturgeon species were also present until the late 70ies and are now locally extinct: the Atlantic sturgeon (*A. sturio*) and the Beluga (*Huso huso*) (Bemis and Kynard, 1997; Birstein and Bemis, 1997).
Once widely distributed in nearly all the main tributaries of the Adriatic Sea, the Adriatic sturgeon is currently considered to be at high risk of extinction. In fact, its presence in the wild has dramatically decreased as testified by catch records of 2000 Kg/year in the 1970s that became only 200 Kg/year in 1990/1991 and only 19 individuals captured in 1993 (Bronzi *et al.,* 1994).
Since 1998 it is included in the Annex II of the *Conservation in International Trade in Endangered Species* (CITES) and since 1996 it was also listed in the *Red List* of the IUCN.
Moreover, in a recent press released by the IUCN, the Adriatic sturgeon was reclassified from "*vulnerable*" to "*critically endangered and possibly extinct in the wild*" since no natural spawning has been recorded in the last 20 years (Congiu *et al.,* 2011).

1.4.2 Anatomy, habitat and biological cycle

As representative of the family Acipenseridae, the species *A. naccarii* has all the archaic characteristics of these "*Living Fossil*". Among them, it presents a cartilaginous skeleton, acentric vertebrae due to the persistence of the notochord and heterocercal caudal fin.
As all sturgeons, it presents very small eyes in compared to the large fish size that in this specie can reach two meters in length and 80 kg in weigh. The mouth is ventral "proctactile" and the rostrum with tactile barbels reveals its benthivore feeding habits. The rest of the

body has not acquired the typical specializations of a benthic life-style (Billard and Lecointre*,* 2001) probably because sturgeons are also long distance swimmers as required by their anadromous reproductive behaviour.
Its life cycle is quite long and individuals can live beyond 50 years. The sexual maturity occurs late, at 8-10 years for males and 10-12 years for females; moreover, reproduction cannot occur annually (Billard and Lecointre*,* 2001).

1.5 Background on the status of the species

The first conservation measure for the Adriatic sturgeon started in 1977, when about 50 immature wild animals were transferred from the Po River and its tributaries to a private aquaculture plant (named V.I.P.) located in Northern Italy (Orzinuovi – BS), thanks to the owner Giacinto Giovannini. This collection of wild animals represented the beginning of the *ex situ* conservation efforts for this species (Bronzi *et al.,* 1999).
At that time, this action was not aimed to establishing a captive brood-stock but, on the contrary, it was driven only by the passion of Giacinto Giovannini for these animals. However, the total absence of evidence of reproductive events occurring in the wild in subsequent years and the status of vulnerability of the species given by the IUCN, made this stock (from now on named F0 stock) exceptionally important. The F0 stock is presently considered the last remaining group of animals of certain wild origin (Congiu *et al.*, 2011). All *A. naccarii* reared in captivity worldwide directly descend from this limited F0 parental stock.
In 1988, the first captive reproduction performed by hormonal induction and squeezing was successful and since then, at the V.I.P. plant, crosses between F0 individuals were performed annually and several stocks of F1 generation were produced. During the last 20 years, many F1 stocks were purchased by local administrations (*e.g.* Ticino River Park and Province of Piacenza) with the aim of releasing fingerlings and establishing their own future F1 captive brood-stocks. Up to 2001 the estimated production for consumption only was of few hundred tons per year (Billard and Lecointre*,* 2001) while for restocking activities over 300.000 juveniles were released in the wild between 1988 and 2007.
Following most reproductions performed at the V.I.P. hatchery, part of the fingerlings produced were collected and reared in the same plant as future breeders. Most of these animals have now reached the sexual maturity (Congiu *et al.,* 2011).
In 1992, the European Council established the "Life Natura" as instrument for financial support to projects aimed to the conservation of species and habitats. In 2003/2004, the European Union has approved two LIFE projects aimed to the restoration of the Adriatic sturgeon in Italy: "*Conservazione di Acipenser naccarii nel fiume Ticino e nel medio corso del Po*" and "LIFE COBICE Project".
During the "Life Cobice project" a complete characterization of the parental F0 wild stock was performed in order to evaluate the remnant variability of this species. All F0 individuals were genotyped at 24 microsatellite loci and for all specimens the mitochondrial control

region (dLoop) was completely characterized identifying 6 mitochondrial haplotypes (Congiu *et al.,* 2011).

Although the F0 stock is composed by few animals, some of which already dead at the time of sampling for the characterization, genetic data show that a good diversification among animals is still present (Fig. 5) (Congiu *et al.,* 2011).

The parental F0 stock is now decimated to only 13 individuals due to their ripe age, making the F1 stocks an increasingly important source of remnant genetic diversity. This available diversity must be characterized by pedigree reconstruction in order to collect relevant information to be used in recovery plans.

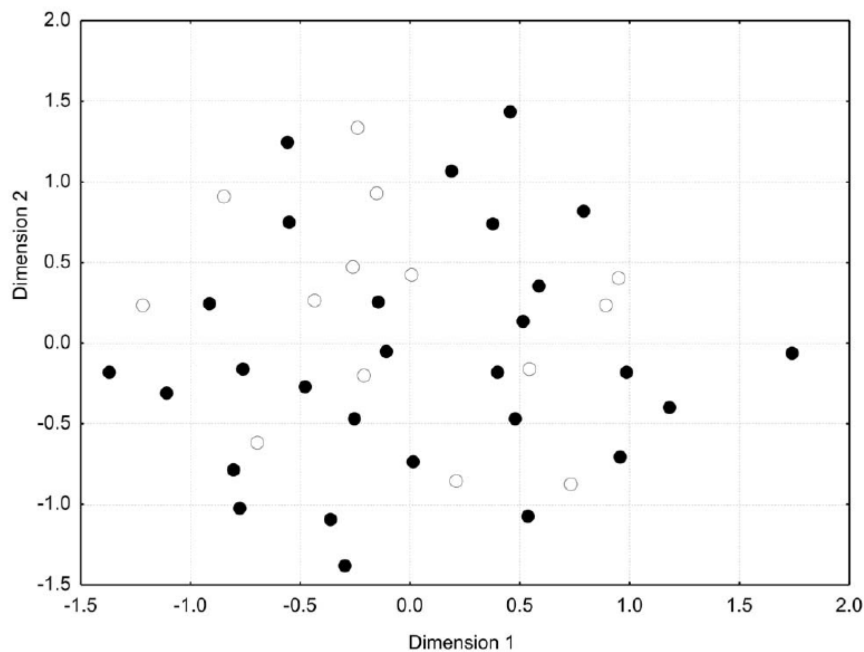


Figure 5: Multi-Dimensional Scaling (MDS) obtained from the pairwise genetic distances at 24 microsatellite loci among the individuals (N=42) of the F0 stock. The black and white colours indicate the mtDNA haplogroup Po1 and Po2, respectively. Stress value 0,29. (Congiu *et al.,* 2011)

2. AIM OF THE WORK

Considering that the last remaining population of certain wild origin (F0 stock), retained in captivity since 1977, was decimated to few individuals, the several F1 specimens available in different aquaculture plants are the most important source of variability for this species. Most of them have now reached the sexual maturity; hence, the *ex situ* conservation on the F1 generation is the only viable approach for the safeguard and the restoration of this endemic species.

The main goal of this project is the management of all the F1 stocks of *A. naccarii* through their complete genetic characterization at mitochondrial and microsatellite markers.

Pedigrees and relationships among individuals will be reconstructed based on the genotyping, and will be taken into account to plan a specific long-term breeding scheme.

Within the project different aspects of evolutionary and biological interest were also investigated in order to fill important gaps in knowledge and to develop methodological tools that have led to a more efficient characterization of stocks and to a more reliable breeding plan.

Projects goals

In detail, the following aspects were examined:

1. Investigation on segregation inheritance pattern in the tetraploid genome of *A. naccarii* (**Boscari *et al.,* 2011 – Annex A**).
2. Evaluation of the consequences of conducting breeding program without the support of genetic investigations (**Boscari and Congiu*,* 2014, *Accepted* – Annex B**).
3. Development of a tool for species and hybrid identification to identify possible contamination within stocks of pure species (**Boscari *et al.,* 2014, *In press* – Annex C**).
4. Development of a new allocation procedure for parental assessment and pedigree reconstruction in polyploid organisms.
5. Genetic characterization of the available stocks of *A. naccarii* reared in captivity for conservational purposes.
6. Development of a management plan for the Adriatic sturgeon and evaluation of its sustainability in aquaculture (**Boscari *et al, manuscript* – Annex D**, points 4, 5 and 6).

3. RESULTS AND DISCUSSION

The content of the four publications resulted from this study are summarized in the following chapters. At the end of the four summaries, the detailed original publications (or manuscripts) are annexed (Annexes A, B C and D).

3.1 Inheritance pattern

(original publication in Annex A)

Sturgeons present a very complex and fragmented genome with a high number of chromosomes and micro-chromosomes (Van Eenennaam *et al., 1998;* Ludwig *et al.,* 2001) and different levels of ploidy that make difficult genetic investigations necessary for a correct management (Ludwig 2006).

Different ploidy levels are the result of independent duplication events occurred from a common ancestor with 60 chromosomes (Birstein and Bemis*,* 1997; Ludwig *et al.,* 2001*;* Fontana *et al.,* 2008). Based on the number of chromosomes, two main groups of species with 120 and 240 chromosomes, respectively, originated from these duplication events. The ploidy degrees of these two groups are still debated. Some authors, in fact, assign a tetraploid and ottoploid condition to the two groups with 120 and 240 chromosomes, respectively (Birstein and Vasil'ev, 1987); while others consider the same two groups as diploid and tetraploid due to a functional reduction following the genome duplication (Fontana *et al.,* 2007).

The Adriatic sturgeon (*A. naccarii*) presents more or less 240 chromosomes and it is considered a functionally tetraploid species (Fig. 6) (Fontana *et al.,* 1999; Fontana 2002).

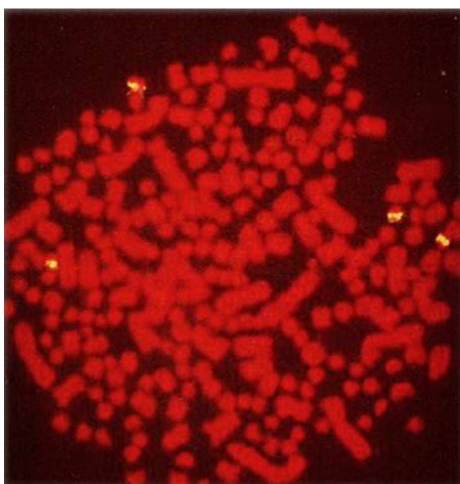


Figure 6: Metaphase of *A. naccarii* with *In situ* hybridization for the 5S rDNA (Fontana *et al.,* 1999).

The tetraploid condition can be generally reached by two mechanisms, an auto- or an allo-polyploidization, to which different modalities of inheritance are associated, namely tetrasomic and disomic, respectively (Stift *et al.,* 2008; Boscari *et al.,* 2011).

In the case of tetrasomic inheritance, four homologous chromosomes can randomly pair during meiosis and can segregate in six chromosome combinations. This pattern is usually associated to intraspecific genome duplications (auto-polyploidy). On the contrary, in the case of disomy, the four chromosomes cannot pair randomly due to the presence of homoeologues chromosomes. Thus, only the homologous

chromosomes will pair during meiosis but will not segregate into the same gamete, generating only four chromosome combinations. This pattern is associated to hybridization, very common in sturgeons, followed by genome duplication (allo-poliploidy) (Wendel 2000; Stift *et al.,* 2008; Meirmans and Van Tienderen, 2013).
Understanding whether *A. naccarii* follows a tetrasomic or a disomic inheritance pattern is important in a project aimed at planning allocation procedures. In fact, the variation of the alleles potentially inheritable from parents does change the expected genotypes in the offspring.
This aspect would further contribute to set up a procedure to generate virtual offspring and a more accurate pedigree useful for breeding plan; moreover, from an evolutionary point of view, this topic could be useful also to understand the origin of tetraploidy in this species.
For this purpose, microsatellite inheritance patterns were analyzed in 7 complete families at 7 microsatellite loci, for a total of 12 family/locus combinations and 174 individuals analyzed. For each available family, the loci considered were opportunely selected among the 24 loci genotyped during the characterization of the F0 stock (Congiu *et al.,* 2011). Each locus was considered informative only if one parent of each family pair was completely heterozygous in order to have the presence of each allele in single copy, while the other parent could share no more than 1 allele with the heterozygous parent. More details are reported in the publication (Boscari *et al.,* 2011 in Annex A).
In this way, supposing a sufficient number of offspring per family, it is possible to trace in the progeny the different allele combinations inherited from the heterozygous parent.
At all 7 loci analyzed, all six possible combinations were observed, allowing to unambiguously reject a strict disomic inheritance and to suggest tetrasomy as the more likely model in any of the cross combinations.
To compare the null hypothesis of tetrasomy (H_0) with disomic pattern and with different intermediate models of inheritance, a likelihood ratio test was applied, with 1 df following the model proposed by Stift *et al.* (2008). This method allowed to reject a strict disomic pattern, confirming the previous results, and also any intermediate models which resulted no significantly better than the null hypothesis of tetrasomy (Boscari *et al.,* 2011).
These results provide relevant information to the establishment of an optimal parental allocation procedure considering equally probable the six chromosome combinations expected in a tetrasomic inheritance pattern.
The high incidence of hybridization events among sturgeon species leads to consider allo-polyploidization as the more probable cause of the tetraploid condition in sturgeons. However, this study suggests that a current ploidy level in the Adriatic sturgeon was reached through an auto-polyploidization event, involving only one species.
A. naccarii shows a very high genetic similarity with other three sturgeon species (*A. gueldenstaedtii, A. persicus* and *A. baerii*) with which shares a common ancestor generating a single phylogenetic cluster. About the segregation pattern of these species there are no information but, considering that Ludwig *et al.* (2001) suggested a genome duplication occurred in the common ancestor of this group of species, it is possible to expect similar results in the three species *A. gueldenstaedtii, A. persicus* and *A. baerii*.

3.2 Sturgeon species and hybrids identification

(original publication in Annex B)

The ability to correctly identify sturgeon species is a critical prerequisite to fighting the illegal market of caviar (Knapp *et al.,* 2006) but has also a direct application to conservation since it would ensure the identification of individuals as belonging to the same species to build a management unit (Herran *et al.,* 2004).

In regard to sturgeons, there are known difficulties in genetic differentiation among different species. Most of them cannot be identified due to their strictly relatedness (Havelka *et al.,* 2011) such as *A. gueldenstaedtii, A. persicus, A. baerii* and *A. naccarii* which are known as *"gueldenstaedtii-complex"* and share the same common ancestor as explained above (Knapp *et al.,* 2006; Ludwig, 2008; Doukakis *et al.,* 2012).

Also hybrids and their products (meat and caviar) are not distinguishable from pure species and are often preferred in aquaculture for their higher growth performances and the consequent younger age of maturity that, in these long living fishes, correspond to a considerable saving of money (Wei *et al.,* 2011). Moreover, while hybrids obtained by species with different ploidy level are almost sterile, hybridization events between species with the same number of chromosomes are common in sturgeons and these hybrids can produce eggs (Havelka *et al.,* 2011; Dudu *et al.,* 2011).

The decrease of wild sturgeon populations has seen a parallel increase of aquaculture plants. Several species and hybrids are often reared in the same plants leading to accidental mixings and mis-identifications. When animals for commercial production and for restocking purposes are reared in the same plant, a possible consequence is the release of allochthonous species as testified by Congiu *et al.* (2011), who have identified a putative hybrid between *A. transmontanus* and *A. naccarii* within a brood-stock of "pure" *A. naccarii*. This reflects data reported by Williot *et al.* (2001) on the white sturgeon (*A. transmontanus*) as the dominant species reared in Europe (46%) for caviar production in 1999.

In Italy, the Adriatic sturgeon is reared for restocking but the hybrids with *A. baerii* are also produced for commercial purposes (Williot *et al.,* 2001).

Until now, a reliable method to easily discriminate among different sturgeon species does not exist and the majority of tools available are based on mitochondrial DNA, which are not suitable for hybrid identification due to its maternally inheritance (Ludwig 2008; Doukakis *et al.,* 2012).

Therefore, the availability of a good method for species and hybrid identification based on nuclear markers is of primary relevance for both conservation and commercial purposes (Havelka *et al.*, 2011).

In this work, a PCR-base tool was set up allowing a cost effective and a reliable identification of sturgeon species and hybrids of the most commercialized species in Europe and North America. The method is based on diagnostic SNPs on the first intron (RP1) of the Ribosomal

Protein S7 gene (PRS7) and incorporates data on the mitochondrial d-Loop and Vimentin gene for hybrids identification. S7 and Vimentin results are visualized exclusively on agarose gel with no need of sequencing instead required for d-Loop analysis. This study, whose publication is reported in Annex B, is the first that incorporates a validation test using a high number of individuals following the SNPs characterization and diagnostic primer design.

The initially characterization of the RP1 by universal primers revealed the presence of three putative loci (A, B and C in fig. 7)

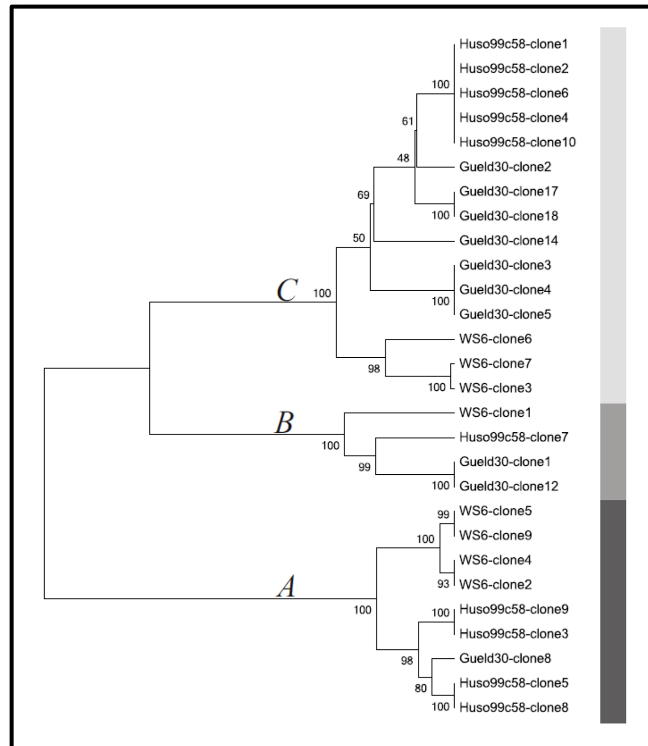


Figure 7: Phylogenetic tree (UPGMA) obtained with sequences from e different species of RP1 amplified with universal primers. The presence of three putative loci is supported.

Focusing on locus A, the intron 1-Locus A of the RPS7 gene was sequenced and cloned for 102 individuals of 13 species (*A. baerii, A. gueldenstaedtii, A. persicus, A. ruthenus, A. transmontanus, A. naccarii, A. fulvescens, A. sinensis, A. stellatus, A. sturio, A. schrenckii, Huso dauricus* and *H. huso*). Diagnostic SNPs or deletion were identified and used to build diagnostic primer pairs as explained in Boscari *et al.* (2014) (Annex B).

Diagnostic SNPs were observed for 5 species (*A. naccarii, A. fulvescens, A. stellatus, A. sinensis* and *A. transmontanus*), while 2 common deletions were detected in *A. baerii-A. ruthenus* and *A. schrenckii-H. dauricus,* respectively*.* A total of 8 selective primer pairs were designed.

Primer validation was performed for each primer pair on samples of target and non-target species for a total of 702 animals of the same species reported above (except the *A. sturio* for which sample were not available).

A 100% identification success was obtained for the five specific primer pairs because no amplification was observed in any of the non-target species analyzed.

Besides the pure species identification, a fundamental application of this tool is the possibility to identify interspecific hybrids here tested on two commercially relevant hybrids: the AL (*A. naccarii x A. baerii*) and the Bester (*H. huso x A. ruthenus*).

With the application of this new tool, the expected detection power for AL and Bester increased from 0 to 100% and from 0 to 80%, respectively.

In conclusion, this method represents a new, fast, inexpensive and reliable tool useful for sturgeon and hybrids identification. It is already suitable for law enforcement and provides important benefits in the fight against the international trade of illegal products.

Moreover, the method has important applications also in checking the purity of *ex situ* brood-stocks, an essential prerequisite in conservation programs.

Among the several diagnostic markers detected in this study and reported in detail within the publication in Annex B, especially relevant is the development of the first totally reliable marker (100% accuracy) for *A. naccarii* identification.

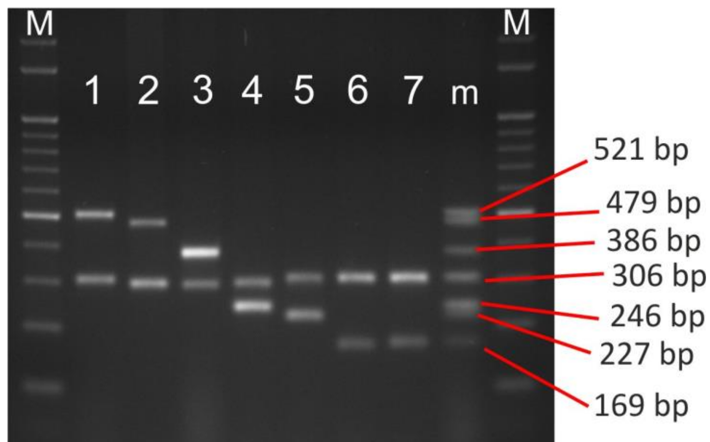


Figure 8: Expected profile and band sizes of the species *A. fulvescens (1), A. stellatus (2), A. sinensis (3), A. transmontanus (4), A. naccarii (5), A. baeri (6)* and *A. ruthenus (7).* The band of 306 bp corresponds to the positive control.

3.3 Evaluation of past management activities

(original manuscript in Annex C)

As widely explained in introduction, *ex situ* conservation in the past was seen as the maintenance of individuals in captivity from which obtain offspring to be released in the wild or to generate new brood-stocks. Several reproductions were performed without the support of genetic analysis and pedigree reconstruction but simply selecting individuals by chance.

In order to evaluate the consequences of breeding plans conducted without genetic support and to assess the effects of random selection of animals on stock genetic variation, two stocks involved in past conservation activities of the Adriatic sturgeon in the Ticino River Park (TRP) have been analyzed.

The two stocks, Stock_1 (N = 137) and Stock_2 (N = 116), were originally acquired from V.I.P plant by the Province of Piacenza and TRP, respectively. All F1 individuals derived from reproductions between individuals of the F0 parental stock. Since then, the Stock_1 was retained in captivity as future brood-stock. Now these animals have reached the sexual maturity and some of them have been already selected for uncontrolled reproductions (personal communication). Instead, the Stock_2 belongs to part of past restocking activities involved in the recovery of a land-locked population of Adriatic sturgeon confined upstream the dam of "*Isola Serafini*". All animals of the Stock_2 were sampled before the release.

In order to estimate the level of genetic diversity retained by these two stocks in comparison with the variability detected in the parental F0 stock (Congiu *et al.,* 2011), all individuals were genotyped at 7 microsatellite loci (LS-39, AoxD64, AnacE4, AnacC11, AoxD234, AnacA6, An20). These markers were opportunely selected among the 24 used for the parental characterization to obtain an 92% of allocation success, as expected from simulations performed with virtual offspring. In the case of unresolved multi-allocations, the number of loci was increase up to 10 (AoxD161, AocD241, An16); moreover, the mitochondrial d-Loop was sequenced from all individuals in order to solve some pedigree ambiguities.

Given the tetraploidy of this species, the usual methods to calculate allele frequencies are not suitable due to the presence of partial heterozygous individuals for which genotypes at different loci cannot be solved.

In this work (Annex C), Band Sharing (BS) was adopted as similarity index to infer relationships among animals. This value is calculated considering only the phenotypic information obtained from microsatellite loci and corresponds to the proportion of shared bands between two multi-locus phenotypic profiles (see Box 1).

Results of allocations of F1 animals to the corresponding F0 parents revealed the presence of few families with known pedigree within both stocks.

In details, for the Stock_1, a total of 8 families were identified with an uneven distribution of individuals (Table 1), most of them in fact are composed by only one fish. The unexpected result was the high presence of animals (more than half of the stock) for which parents

remain unknown. They probably are offspring of parents dead before sampling for the F0 genetic characterization.

BOX 1: THE BAND-SHARING ALLOCATION PROCEDURE

In the Adriatic sturgeon, the routinely genetic analyses on molecular markers such as microsatellites are complicated by its tetraploid condition that implies the presence of multiple alleles per locus. Each allele could be present in multiple copies, which cannot be distinguished from genotyping. Consequently, allelic frequencies cannot be perfectly calculated but only estimated with several different methods.
Some authors estimates allele frequencies extrapolating the number of allele copies from peak heights and area (Jenneckens *et al.,* 2001; Welsh and May, 2006; Zhao and Chang, 2006); however, this correlation was not observed in the Adriatic sturgeon.
To estimate genetic distances we have applied a phenotypic approach, exclusively based on phenotypes obtained from microsatellite information at different loci.
More in details, for each individual a multi-locus profile was created as string in binary code composed by 0s and 1s, corresponding to the absence and presence of alleles, respectively.
To perform a parental analysis, for each possible F0 parent pair, a cumulative profile was created and compared with the multi-locus profile of each F1 individual.

Parent 1	0010100111011
Parent 2	1000110011001
Cumulative profile	1010110111011
Offspring	1000110001010

F1 animals are considered perfectly allocated only if their multi-locus profiles are entirely contained in the cumulative profile of a putative parent pair.
Starting from these phenotypic profiles it is also possible estimate pairwise genetic distances as (1-BandSharing). The Band Sharing is a coefficient of similarity estimated on the alleles shared by two profiles:

$$BS = 2a/(2a + b + c),$$

where a corresponds to bands shared by both profiles while b and c are bands exclusively present within the first and the second profile, respectively.

Stock_1 (F1 future breeders)			Stock_2 (F1 released)		
N° of F1 individuals	**Parental pairs**	**Haplotype**	**N° of F1 individuals**	**Parental pairs**	**Haplotype**
8	**Pelvienne (F) x NaccS18 (M)**	2	22	**NaccS8 (F) x 740 (M)**	2
18	**NaccS8 (F) x Matto (M)**	2	32	**NaccS3 (F) x 740 (M)**	3
26	**NaccS15 (F) x NaccS13 (M)**	4	17	**NaccS33 (F) x 740 (M)**	3
1	**NaccS3 (F) x Matto (M)**	3	1	**O2 (F) x Matto (M)**	5
4	**NaccS4 (F) x Matto (M)**	2	1	**NaccS26 (F) x NaccS13 (M)**	5
1	**NaccS3 (F) x NaccS18 (M)**	3	1	**O2 (F) x NaccS31 (M)**	5
1	**NaccS7 (F) x NaccS27 (M)**	3	1	**NaccS4 (F) x Matto (M)**	2
1	**NaccS7 (F) x NaccS6 (M)**	3	1	**NaccS33 (F) x NaccS29 (M)**	3
3	**Multi-allocated**	2	1	**NaccS4 (F) x Raspo (M)**	2
70	**Not allocated**	2	1	**Multi-allocated**	3
4	**Not allocated**	*	38	**Not allocated**	2/3/5

Table 1: results of parental allocation of the Stock_1 and Stock_2 using the Band-Sharing method. Haplotype for each family was also reported. * Indicates individuals which haplotypes not correspond to the 7 *A. naccarii* haplotypes.

To infer possible relationships among these not-allocated animals, the criteria used in Congiu *et al.* (2011) were followed. The distribution of pairwise genetic distances among individuals with a known degree of relatedness (sibs, half-sibs and unrelated animals) was estimated. Values corresponding to the 99th percentile of the observed distances were considered as threshold values to exclude the corresponding relatedness (0.52 for full-sibs and 0.62 for half-sibs).

The comparison of these values with the pairwise distance matrix among not-allocated individuals has revealed the presence of a single family of full-sibs. In synthesis, only four families enclosed the 90% of individual composing the Stock_1.

Similar results were obtained from the characterization of the Stock_2 (Table 1).

Also in this case 8 uneven families were identified while, among not-allocated, the presence of three families of full-sibs where estimated using the threshold values as previously explained.

Despite the same number of families detected in both stocks, the diversity indices appositely created to represent the diversity within stocks (reported in detail in the Annex C), showed a greater loss of familiar heterogeneity within the Stock_2 in comparison with the Stock_1. This could be explained by the presence of highly related individuals. The three most abundant families, in fact, shared the same sire (Table 1).

This highlights how careless conservation activities may potentially have a strong detrimental impact. The Stock_2 shows a case of releasing in the wild of animals strictly related. About the 65% of them share one of the two parents (half-sibs).

Moreover, within the Stock_1 that should be a brood-stock of pure *A. naccarii,* the mitochondrial analysis revealed a possible contamination or contribution of other species. Four not allocated individuals did not present haplotypes of *A. naccarii* and two of them present highly genetic distances (Fig. 9). A Blast research revealed a correspondence for two

of them with *A. gueldenstaedtii* haplotype while for the other two with *A. transmontanus* haplotype.

As indicate in the above chapter, the purity of a stock is a fundamental aspect in founding brood-stock; thus, the RPS7 tool for species and hybrid identification was applied to all not-allocated individuals.

Hybrids between the white sturgeon (*A. transmontanus*) or the Russian sturgeon (*A. gueldenstaedtii*) females and Adriatic sturgeon (*A. naccarii*) male were confirmed. These hybrids should be excluded from the stock to avoid their accidentally use in reproductions.

The representation within the Stock_1 of only few F0 breeders and the presence of hybrids are a direct consequence of a careless management of this species. This work emphasizes how the *ex situ* management must necessarily rely on the careful characterization of the residual genetic diversity. In fact, the Stock_1 resulted scarcely suitable as future brood-stock: by random selection of breeders the probability to mate full-sibs individuals is remarkable high (35%).

For further reproductions, a tied collaboration with others plants rearing animals genetically unrelated and a re-organization of brood-stocks are strongly encouraged.

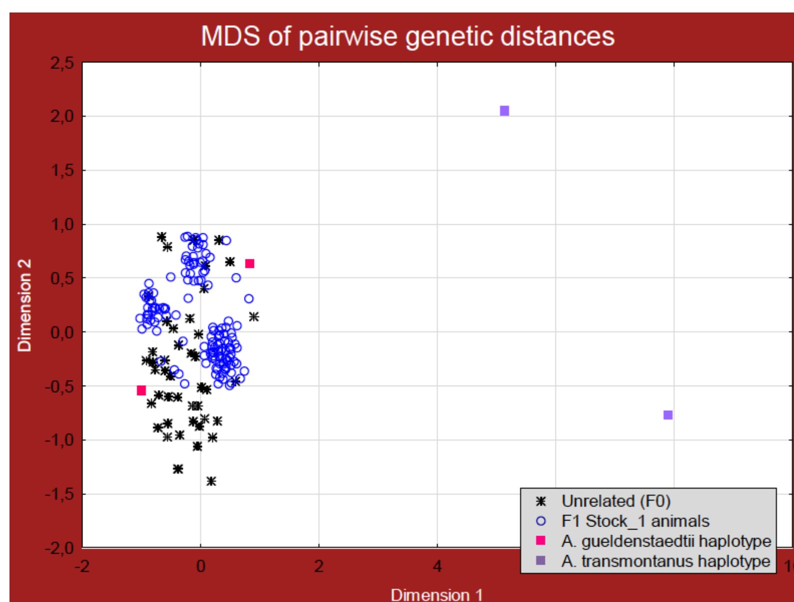


Fig. 9: MDS of pairwise genetic distances among all individuals of the Stock_1 and F0. The color squares represent the four individual resulted hybrids.

3.4 Characterization of the V.I.P. stock

(original manuscript in Annex D)

As seen in the above chapter, pedigree reconstruction represents an important step towards the correct management of the available diversity within captive brood-stocks (DeSalle and Amato, 2004).

In the present study the allocation procedure proposed by Congiu *et al.* (2011) based on a band-sharing (BS) approach and explained in detail in Box 1 was initially applied.

Besides the strong limitation of relying exclusively on the phenotypic information, the BS method raised some problems such as the absence of a probability test to verify the efficiency of loci used for parental allocation. Moreover, a specific pattern of inheritance was not specified and this could result in false positive allocations when all the alleles observed in a F1 were shared with a single parent.

Therefore, a new tool for pedigree analysis (named "*BreedingSturgeons*") was developed using c++ format to improve the performances of the BS method. The program allows estimating for each individual a compatibility index for all putative parent pairs under the assumption of tetrasomic inheritance, as demonstrated in the first chapter (Boscari *et al.,* 2011 – Annex A).

This index is estimated as follows: at each locus, starting from the observed microsatellite profile (phenotype), all possible tetraploid genotypes are inferred for both offspring and parents (p1 and p2). All possible pairs of inferred parental genotypes are screened to assess if all alleles observed in the progeny are justified, with the constraint that the offspring is assumed to inherit two allele copies from each parent. These allele copies can be identical except from completely heterozygous parental genotypes.

Thus, the single-locus compatibility index, *ci(x),* is estimated as follows:

$$ci(x) = \frac{\sum[\mathrm{A(p1)+A(p2)}]}{4},$$

where *ci(x)* is the compatibility index for that animal at a single locus ($0 \leq ci(x) \leq 1$), A(p1) is the number of allele copies inheritable from parental 1 genotype ($0 \leq A(p1) \leq 2$) and A(p2) is the number of allele copies inheritable from parental 2 genotype ($0 \leq A(p2) \leq 2$).

The same approach is extended to all loci estimating a multi-locus compatibility index (*CI*) as average of single locus probabilities:

$$CI = \frac{\sum_{x=i}^{j}[\mathrm{ci(x)}]}{n°loci}.$$

This procedure is applied to all possible parental pairs and the ones with higher *CI* are selected as more probable parents.

Once the pedigree reconstruction is completed, the composition of stocks is evaluated.

This part of the work, reported in the manuscript Annex D, concerns the genetic characterization of the V.I.P. Stock, represented by a group of F1 individuals (N = 445) retained in captivity after several reproductions performed during the last 20 years. This stock is the larger group of individuals available for this species.

Genetic analyses carried out on this stock are the same already described in the previous chapter, with the above improvement of the allocation procedure.
Results obtained from the genotyping at 7 microsatellite loci, sequencing of the mitochondrial control region and parental allocation to reconstruct the pedigree information showed that these individuals have retained a considerable proportion of the parental generation variability. Out of 445 F1 animals, 382 were successfully allocated to 30 families, while only 63 were not allocated (for more details see Annex D). Within this last group, 4 families with 4 different haplotypes were identified, for a total of 34 families (Fig. 10).
Unfortunately, due to a recent poisoning event, about half of the F1 animals died and 233 survived. Even if only four families were lost (Fig. 10), this dramatic event highlights the urgency for targeted actions such as splitting the stock in different locations, to minimize the effect of possible catastrophic events.

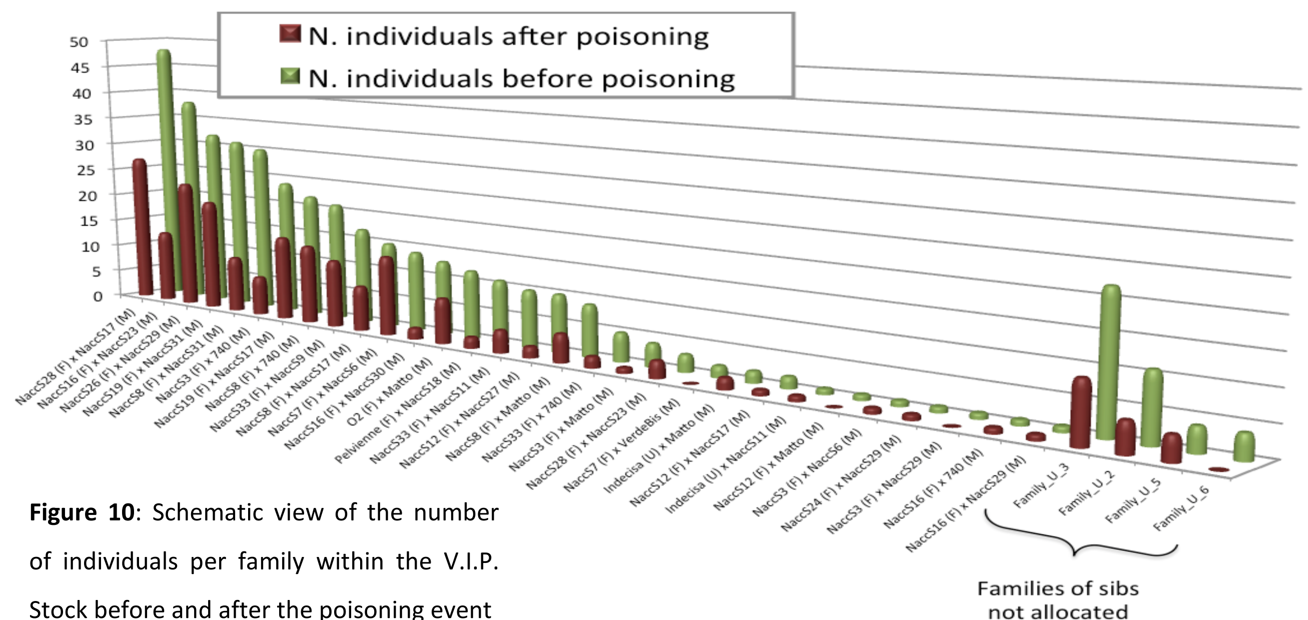


Figure 10: Schematic view of the number of individuals per family within the V.I.P. Stock before and after the poisoning event

Between the V.I.P. stock here analyzed and the previous Stock_1 retained by the Ticino River Park, the V.I.P. stock was identified as the most precious source of variability for *A. naccarii* and, therefore, it was used to plan the breeding program and to form the "*breeders unit*". However, in order to improve the heterogeneity of this "*breeders unit*", three families of the Stock_1 were also included in the breeding program, assuming the collaborative involvement of different aquaculture plants. Two of these additional families was not represented in the V.I.P. stock and one of them was not-allocated, while the third family was already available but under-represented.
Collectively, a total of 32 F1 families (28 with known and 4 with unknown pedigree) compose the "*breeders unit*" used for the planning of the breeding activities (Fig. 11).
As happened for the Stock_1 (previous chapter), one not-allocated animal of the V.I.P. Stock presented a haplotype of the white sturgeon (*A. transmontanus*). The application of the species and hybrid identification tool (Boscari *et al.,* 2014) confirmed this individual to be

hybrid between *A. transmontanus* female and *A. naccarii* male and, unexpectedly, revealed the presence of other three hybrids not previously identified due to the mitochondrial compatibility with *A. naccarii.* In this cases, in fact, the Adriatic sturgeon was the maternal species while the white sturgeon and Russian sturgeon represented the male contribution in two and one individuals, respectively.

Also these hybrids were discarded from the "*breeders unit*".

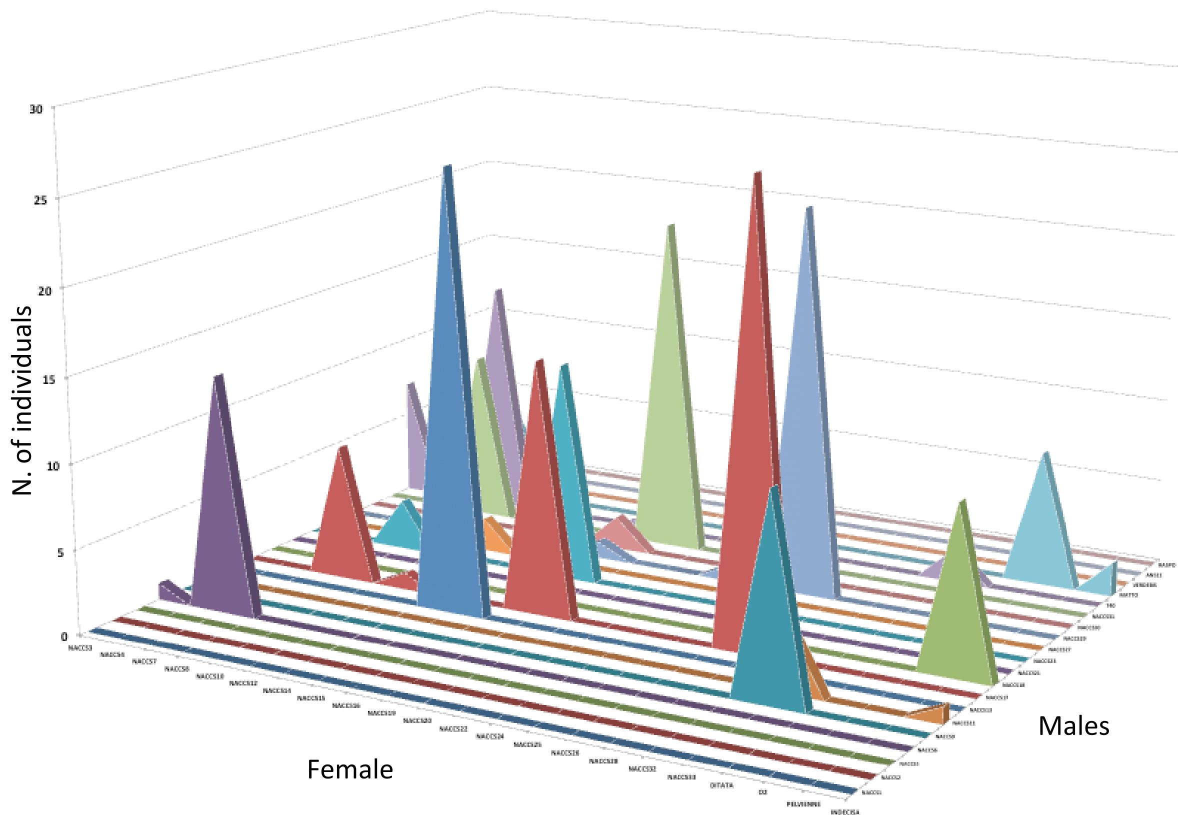


Figure 11: Schematic representation of the "*breeders unit*". The F1 families with known pedigree, obtained by crossing the individuals of wild origin ("males" and "females"), and their abundances, are reported.

3.5 BREEDING PLAN

(original manuscript in Annex D)

4.5.1 Short-term breeding plan (F0xF0)

As explained in the introduction, the F0 population of *A. naccarii* is the only available of certain wild origin. Thus, one of the main recommendations, usually reported in restocking guidelines, which is to introduce new wild animals (Witzenberger and Hochkirch, 2011) cannot be followed. However, a short-term breeding plan was proposed, involving the 13 still alive F0 of which 9 seem to be not represented in the F1 generation due to the absence of their offspring on the basis of parentage analysis.

The priority given to these animals allows starting the plan with wild individuals and is justified by the attempt to represent their genetic variability, which would be lost in the case of failure of reproduction also considering the advanced age of these animals (about 50 years old).

We tried to have the contribution of each remnant F0 breeder in at least three families within the F1 generation. The already existing families representatives of each F0 parent were counted and the corresponding number is reported in Fig. 12 as ScoreA. Doing this, we disregarded the families with less than 10 animals. The ScoreB represents the number of crosses per F0 to plan in order to generate at least three families per breeder in the F1 generation (ScoreA+B).

Mating pairs were selected among the more distant males and females basing on the pairwise genetic distances at 24 loci (Fig. 12) trying to minimize the number of crosses and maximize the number of parents represented. Following this strategy, 18 crosses were selected as priority (highlighted in Fig. 12).

ScoreA	ScoreB	Score A+B	Alive F0	Sex	Nac4	Nac7	Nac14	Nac20	Nac25	Nac28	Nac5	Nac6	Nac13	Nac17	Nac23	Nac30	Nac31
0	3	3	Nac4	F	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	3	Nac7	F	0,43802	0	0	0	0	0	0	0	0	0	0	0	0
0	4	4	Nac14	F	0,59259	0,4955	0	0	0	0	0	0	0	0	0	0	0
0	3	3	Nac20	F	0,53968	0,53488	0,58621	0	0	0	0	0	0	0	0	0	0
0	3	3	Nac25	F	0,57627	0,57025	0,5	0,50794	0	0	0	0	0	0	0	0	0
1	3	4	Nac28	F	0,47619	0,48837	0,62069	0,50746	0,57143	0	0	0	0	0	0	0	0
0	3	3	Nac5	M	0,48718	0,58333	0,60748	0,488	0,60684	0,472	0	0	0	0	0	0	0
1	2	3	Nac6	M	0,54955	0,5614	0,68317	0,54622	0,65766	0,29412	0,52727	0	0	0	0	0	0
1	2	3	Nac13	M	0,55	0,5122	0,6	0,46875	0,48333	0,45313	0,54622	0,48673	0	0	0	0	0
2	2	4	Nac17	M	0,54955	0,54386	0,44554	0,63025	0,54955	0,52941	0,54545	0,59615	0,61062	0	0	0	0
1	3	4	Nac23	M	0,56667	0,46341	0,49091	0,54688	0,46667	0,5	0,47899	0,61062	0,45902	0,50442	0	0	0
0	3	3	Nac30	M	0,52542	0,50413	0,44444	0,50794	0,49153	0,49206	0,4359	0,58559	0,55	0,47748	0,43333	0	0
2	3	5	Nac31	M	0,51724	0,59664	0,60377	0,5	0,53448	0,56452	0,25217	0,59633	0,61017	0,54128	0,52542	0,53448	0

Figure 12: Distance matrix at 24 microsatellite loci between still alive F0 males and females. Cells highlighted in grey indicate the mating selected following the criteria explained in the text. Score A+B represents the number of families per breeders in the next generation and it is given by the already existing families (ScoreA) and the crosses selected (ScoreB).

4.5.2 Long-term breeding plan (F1xF1)

The principle that underlies the usual approach in breeding plans is based on a selection of candidate breeders by genetic distances, founder importance coefficient or means kinship (Ballou and Lacy, 1995; Montgomery et al., 1997; Hayes et al., 2006; Sekino *et al.,* 2004; William and Hoffman, 2009). Subsequently, on the basis of available information a breeding scheme is selected. The more common strategies are the method by chance, the minimal kinship (*mk*) Strategy, the rotational or the circular mating systems (Nomura and Yonezawa, 1996; Honda *et al.,* 2004; Winding and Kaal, 2008; Theodorou and Couvet, 2010; Witzenberger and Hochkirch, 2011).

In the case of Adriatic sturgeon, the same criteria were applied to the different families composing the "*breeders unit*" (family selection) rather than follow an individual selection. In fact, the low probability that a single animal matures in a given year prevents the possibility to reason at "individual level", while having several adult animals per family, the chance of having at least one mature animal of a given family selected for reproduction is acceptable. Accordingly, genetic distances were not estimated among animals but among families (Fig. 13). The approach with kinship was discarded in this work due to the polyploid condition of the Adriatic sturgeon for which the allele frequencies in the stock cannot be calculated. In fact, individual genotypes are not known, thus, a phenotypic approach (1 - BS) was followed instead. This was done by considering for each family the cumulative profile at 24 loci of each possible parent pairs in the F0 stock.

All possible pairs of F0 breeders ("virtual families") were created and pairwise combined to obtain the distribution of genetic distances between all possible family pairs. All family pairs sharing one parent (half-sibs) showed a distance value lower than 0.35. This value was set as threshold above which the 100% of crosses between related families was excluded. Thus, all crosses with d<0.35 were discarded from the breeding plan matrix (red cells in fig. 13).

For what concerns the choice of the mating strategy, breeding schemes involving a large number of crosses (*e.g.* rotational mating system) are not feasible due to the long life cycle of this species and the high costs for its management in aquaculture. Thus, in the present work a specific breeding strategy was opportunely optimized considering several constraints. The order in which the families are selected was estimated based on how much the parents are already represented in the stock. This was calculated by averaging the number of progeny that the two parents have in the "*breeders unit*" (priority index). This strategy gives high priority (Fig. 13) to small families with under-represented parents whose alleles have less chance to be transmitted to the next generation. Following this priority, females of each "priority family" should be crossed with males of the more distant family (named "mating family"). Proceeding with the mate choices, crosses already performed cannot be re-selected.

Since we have included a total of 32 families in the "*breeders unit*" (considering also the not allocated animals), planning to do only one cross per "priority family", the complete breeding plan consists of 32 mates.

The plan here proposed is theoretical and should take into account the complex biology of this species and the constraints linked to the aquaculture plant where these conservation activities will be conducted.

Priority		pelv.Nac18	Nac8.matto	Nac7.Nac6	Nac8.Nac31	Nac33.Nac1	O2.matto	Nac26.Nac2	Nac19.Nac1	Nac28.Nac1	
9	pelv.Nac18	0	0	0	0	0	0	0	0	0	
22	Nac8.matto	0,3483	0	0	0	0	0	0	0	0	
15	Nac7.Nac6	0,4162	0,4386	0	0	0	0	0	0	0	
28	Nac8.Nac31	0,4033	0,2455	0,4253	0	0	0	0	0	0	
10	Nac33.Nac11	0,3516	0,3571	0,3257	0,3567	0	0	0	0	0	
11	O2.matto	0,1771	0,1798	0,373	0,3481	0,3407	0	0	0	0	
23	Nac26.Nac29	0,4696	0,4731	0,4483	0,4941	0,3918	0,4365	0	0	0	
32	Nac19.Nac17	0,3579	0,3068	0,3552	0,352	0,3222	0,3263	0,3743	0	0	
31	Nac28.Nac17	0,3797	0,4104	0,2667	0,3977	0,322	0,3797	0,3523	0,2865	0	
16	Nac16.Nac23	0,4211	0,3636	0,3989	0,352	0,4333	0,3368	0,3296	0,3617	0,3514	
29	Nac19.Nac31	0,3474	0,375	0,388	0,2514	0,3889	0,3579	0,4302	0,1915	0,4162	0,
8	Nac16.Nac30	0,4194	0,3953	0,419	0,3486	0,4091	0,3871	0,4057	0,3804	0,3591	0,
17	Nac33.Nac9	0,4144	0,4251	0,3333	0,3647	0,1813	0,3923	0,3294	0,3296	0,3295	0,
30	Nac8.Nac17	0,3667	0,2048	0,3757	0,2189	0,2824	0,3	0,4201	0,191	0,2457	0,
7	Nac12.Nac27	0,3115	0,3491	0,3864	0,3372	0,3873	0,3224	0,3721	0,337	0,3596	0,
18	Nac3.740	0,3862	0,3714	0,3736	0,3483	0,2737	0,3968	0,3708	0,2941	0,3587	0,
25	Nac8.740	0,3587	0,2	0,4124	0,2486	0,2989	0,3043	0,4798	0,3077	0,3855	0,
24	Nac15.Nac13	0,3583	0,3526	0,3556	0,3977	0,322	0,3155	0,3977	0,3405	0,3516	0,
26	Nac28.Nac23	0,3854	0,4157	0,2649	0,4254	0,3956	0,3542	0,326	0,3474	0,1444	0

Figure 13: portion of the distance matrix used for the selection of mates for the F1xF1 breeding plan. Distances are estimated using the cumulative profile of each family at 24 loci. Red cells correspond to d<0,35 and cannot be selected, while green cells are the recommended crosses.

4.5.3 Sustainability of the breeding plan

Initially, to assess the sustainability of this theoretical breeding plan, an R-script (“*CostsBreedingSturgeons*”) was compiled to estimate how many mates per year could be performed in a given hatchery, based on the features of a given plant, providing also an estimation of the years required to accomplish the whole plan. The software estimates the costs for the different grow phases in order to calculate the required financial support per year; moreover, it also provides an outline of the releases that can be made per year under different economical and logistical constraints.

The variables used in the simulation are referred to a hatchery located in Cassolnovo, where the “*breeding units*” will be reared in the near future. Detailed information about the outputs of the script are described in the manuscript (Annex D).

A second R-script (“*BreedingPlanSturgeons*”) was used to assess the recommended number of individuals to breed per family and to estimate the expected fraction of alleles of the parental generation that are successful inherited in the F2 virtual progeny under different strategies of family selection (reported in detailed in the Annex D).

4. CONCLUSIONS

Several restocking activities have been conducted in the last 20 years with the aim of avoiding the extinction of the Adriatic sturgeon. Nevertheless, in the same period no clear evidences of natural reproductions were recorded. The establishment of self-sustaining natural populations is still a distant goal and the future of this species presently depends on the careful management of the last captive stocks. Accordingly, the mail goal of the present study was to propose a breeding program that, if followed, will allow a long-term preservation of most residual genetic diversity of the Adriatic sturgeon.

To this purpose several preliminary investigations were performed in order to i) complete the genetic characterization of the main stocks reared in Italy, ii) assess the inheritance pattern of this species and optimize a parental allocation procedure, iii) reconstruct the pedigree information for all stocks, iv) optimize the first species and hybrids identification procedure based on a single nuclear marker and consequently v) check the stocks for the presence of alien species or interspecific hybrids.

All the above preliminary actions allowed to gain relevant information for the development of a complete breeding plan articulated in two phases: the first involving the last 13 animals of wild origin and the second focusing on the about 30 families of F1, trying to maximise the fraction of genetic diversity transmitted to the future captive generations of breeders as well as to the animals that will be released in the wild. Moreover, an assessment of sustainability of the plan considering economically and managerially constraints is estimated to ensure its feasibility.

In general, this plan represents the first active effort towards a responsible management of captive stocks of this species. The strategy here proposed includes also releases in the wild of such animals that cannot be retained in captivity due to spatial limitations. Despite the released animals are large enough to reduce the chances of predation, the expected success in terms of survival and fitness is not optimal mainly because the environment is not yet suitable for these animals.

The limited success of re-introductions is the main problem of several breeding plans and it is well documented in literature (Lynch and O'Hely, 2001; Woodworth *et al.,* 2002; Williams and Hoffman, 2009; Lyon *et al.,* 2012). As example, Beck *et al.,* (1994) have reviewed that only 11% of re-introduction plans had success, while for Fisher and Lindenmayer (2000) this value increased up to 13%. This might be due to different reasons in part related to the quality of the animals released and in part to habitat degradation.

Moreover, the animals released might not be fit for survival in the wild. In the past decades, Adriatic sturgeon recovery programmes have adopted standard aquaculture technologies for the production of fingerlings to be released, ignoring the need for producing animals acclimated or trained to survive in the wild. Avoiding crosses between related families and equalizing the F1s family contribution in the F2 generation, our strategy could be considered as a starting point that tries to maximize the level of diversity of stocks produced to be released. However, the establishment of a rearing protocol for the production and training of animals, according to the criteria of the fitness for survival is strongly recommended.

Concerning the habitat degradation, one of the main causes for the imperilled status of the Adriatic sturgeon is the fragmentation of the main rivers with dams, that prevents the access to spawning sites. For this reason, constructing adequate fish passages specifically projected for these big anadromous fishes represents a priority.

In general, all strategies, methods and recommendations proposed in this work are in line with the international FAO guidelines for sturgeon management and release. The present work represents the first breeding plan for the Adriatic sturgeon based on a detailed genetic characterization and can be considered as a reference guideline for all conservation actions based on controlled reproductions of this species.

Moreover, this is the first attempt to propose a coordinated strategy for the safeguard of the captive critically endangered Adriatic sturgeon, aimed to preserve its genetic diversity. Given the critical status of this species, the actions here proposed should be realized in the near future and should be complemented by activities at different levels such as the restoration of natural environments or the implementation of releasing protocols aimed to maximize the fitness of the animals.

Finally, this breeding plan and the related approaches, may represent a useful reference also for other imperilled tetraploid sturgeon species.

5. OTHER PUBLICATIONS

Here below I have resumed other publications involving the species *A. naccarii* on which I have participate during my PhD. The original publications are reported in Appendix I and II.

5.1 Transcriptome sequencing of the Adriatic sturgeon (original publication in Appendix I)

This work concerns the analysis of a 454-sequenced transcriptome obtained from brain and gonads cDNA libraries of two *A. naccarii* individuals (one male and one female full-sibs). These are the first transcriptome data for sturgeon species.
I have contributed to the phase of preparation of samples for libraries.
The sequencing yielded 55,000 high quality Expressed Sequence Tags (ESTs) given by the assembling of 182,066 reads for the male and 167,776 for the female.
Besides its evolutionary value for the amount of new information obtained for a member of the Chondrosteans, this work focused on the identification of gene related to sex differentiation, which could be differentially expressed between the two sexes. The transcriptome was screened for 32 genes related to sex differentiation of which 5 appear to be specifically for male and 2 for female.
Moreover, the identification of a large amount of SNPs (21,791) and microsatellites EST-linked (5,295) is an important step towards the establishment of a genome wide genetic markers panel that could be useful in the future in monitoring the effects of conservation activities in captivity and in the wild (releases).

5.2 Analysis of transposable elements in Acipenser naccarii (original publication in Appendix II)

During these years I also had the possibility to contribute to the discovery of *Tana1*, a new putatively active *Tc1-like* transposable element (Plasterk *et al.,* 1999) found in the genome of sturgeons.
The complete sequence of *Tana1* was first characterized in the 454-sequenced transcriptome of the Adriatic sturgeon (*A. naccarii*) by Dott. Michele Vidotto. The element found in the transcriptome was obtained by assembling many short reads but this cannot be a certain indication that the entire sequence is transcribed.
Therefore, the element was isolated and characterized from the cDNA of the same species (by PCR and cloning) to validate the bioinformatics data. Then, *Tana1* was isolated from the genome of Adriatic sturgeon and we demonstrated its occurrence also in the genome of 12

additional sturgeon species including 3 genera of the Acipenseridae (*Acipenser, Huso* and *Scaphirhynchus*).
The integrity of the native form (with the entire sequence of the transposase), the presence of all expected functional domains (Shao and Tu, 2001; Yuan and Wessler, 2011) and its occurrence in the sturgeon transcriptome suggest a current or recent activity of *Tana1*. The results of this study were recently published (Pujolar *et al.,* 2013; Appendix I).
Moreover, the discovery of DNA editing activity by ADAR (*Adenosine Deaminase Acting on RNA*) enzymes on the *Tana1* inverted repeats opens new interesting perspectives for the next year to study the putative activity of this element and the role of post-transcriptional editing on the activity of transposable elements in vertebrates.

ANNEX A

Aquaculture 321 (2011) 223–229

Contents lists available at SciVerse ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aqua-online

Inheritance pattern of microsatellite loci in the polyploid Adriatic sturgeon (*Acipenser naccarii*)

Elisa Boscari, Federica Barbisan, Leonardo Congiu *

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

ARTICLE INFO

Article history:
Received 8 April 2011
Received in revised form 31 August 2011
Accepted 8 September 2011
Available online 16 September 2011

Keywords:
ADRIATIC sturgeon
Acipenser naccarii
Inheritance pattern
Segregation
Tetraploidy
Microsatellites

ABSTRACT

The Adriatic sturgeon is a highly endangered tetraploid species whose conservation depends up ex-situ management of the remnant genetic variability. Understanding whether the species follows a tetrasomic or a disomic inheritance pattern is of primary importance to set up a parental allocation procedure and to establish a long-term breeding plan. Moreover, comprehending the inheritance modality can strongly contribute to understanding the origin of tetraploidy in this species. For this purpose, microsatellite inheritance patterns were analyzed in 7 complete families and at 7 loci for a total of 12 family/locus combinations. For each available family, a preliminary selection of loci was performed, in order to avoid ambiguities due to allele dosage, null alleles or interference between parental contributions. Results allowed to unambiguously reject a strict disomic inheritance pattern and to suggest tetrasomy as the more likely model. Accordingly, parental chromosomes can be expected to pair in the gametes in all possible combinations, though a certain degree of preferential pairing could not be excluded for the limited statistical power reached. This study represents the first investigation of the inheritance pattern in the Adriatic sturgeon and provides relevant information for the correct management of its residual genetic diversity.

Published by Elsevier B.V.

1. Introduction

Sturgeons (infraclass Chondrostei, order Acipenseriformes) are a very ancient fish group, distributed in the Palearctic hemisphere and represented by about 25 species (Peng et al., 2007). The international Union for Conservation of Nature (IUCN, March 2010) recently identified sturgeons as the most endangered group of animals with 85% of the species being at risk of extinction according to the Red List of Threatened Species (http://www.iucnredlist.org). This vulnerability is mostly due to the very high economic interest in these animals, overexploited for the production of caviar, one of the most valuable products on the international food market (Ludwig, 2008). Since 1998, international trade in all species of sturgeons has been regulated under CITES (http://www.cites.org) owing to concerns over the impact of unsustainable harvesting and illegal trade (Ludwig, 2008; Pikitch et al., 2005). Moreover, the dramatic decline in natural sturgeon populations in recent years prompted conservation efforts for most sturgeon species by means of restocking with animals produced by controlled reproduction (Congiu et al., 2011).

The correct management of highly imperiled populations by means of ex-situ conservation programs must be supported by adequate genetic investigations (Doukakis et al., 2010; Ludwig, 2006), which in sturgeons deal with a very complex genome and different levels of ploidy (Fontana et al., 2008; Ludwig et al., 2001). Two main groups of species can be identified based on their number of chromosomes, which is about 120 or 240, respectively. The level of ploidy to be ascribed to these chromosome numbers is still being debated. Some authors consider species of the two groups to be diploid and tetraploid, respectively (Fontana et al., 2007), while others attribute the tetraploid or octoploid conditions to these same groups (Birstein and Vasil'ev, 1987). The high number of chromosomes was likely reached through multiple polyploidization events starting from a 60-chromosome common ancestor (Fontana et al., 2008). After the first duplication event, the resulting tetraploid genome ($4n=120$) underwent a functional diploidization ($2n=120$). For this reason, species presenting about 240 chromosomes, considered to be octoploid by some authors, can be considered as functionally tetraploid, originating from a second event of chromosome doubling that likely took place more than once in Acipenseridae (Fontana et al., 2008) and possibly due to different mechanisms. Accordingly, the different tetraploid species ($2n=240$) distributed in all the northern hemisphere might be either auto- or allo-tetraploids (Fontana et al., 2008).

Depending on the origin of tetraploidy, different modalities of chromosome segregation into gametes are expected, namely tetrasomic and disomic (Stift et al., 2008). The tetrasomic pattern would be followed when each chromosome has four homologous copies ($A_1A_2A_3A_4$) and each copy may pair randomly with any other homolog during meiosis. This is usually observed in autotetraploids (derived from an intraspecific genome duplication). The six possible combinations all can segregate into gametes (A_1A_2, A_1A_3, A_1A_4, A_2A_3, A_2A_4, A_3A_4) (Stift et al.,

* Corresponding author. Tel.: +39 49 8276218; fax: 39 49 8276209.
E-mail address: leonardo.congiu@unipd.it (L. Congiu).

0044-8486/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.aquaculture.2011.09.015

2008). On the contrary, a disomic inheritance pattern would be observed in the presence of four homeolog (i.e., not properly homolog) chromosomes ($A_1A_2B_1B_2$) that cannot randomly pair. This generally occurs in allotetraploids in which tetraploidy is reached through hybridization. Each chromosome pairs only with its homolog in meiosis and only four different allele combinations can segregate into gametes (A_1B_1, A_1B_2, A_2B_1, A_2B_2) (Stift et al., 2008). Besides the pure tetrasomic and disomic modalities, different degrees of preferential pairing between chromosomes could lead to intermediate inheritance patterns in which all the possible chromosome combinations are expected with different frequencies (Stift et al., 2008).

Apart from providing information on the origin of their karyotype, understanding the inheritance pattern in tetraploid sturgeon species represents a necessary step toward the correct management of the residual genetic variability and the establishment of a breeding program for both conservation and aquaculture purposes (Rodzen et al., 2004). This is the case for the Adriatic sturgeon (*Acipenser naccarii*), endemic to the North Adriatic Sea and its tributaries. Since May 2010, it has been listed by IUCN as critically endangered and possibly extinct in the wild, on the basis of an estimated population decline greater than 80% (possibly 100) over the past three generations. Moreover, no evidence of natural spawning has been reported in the last 15 years. The few animals occasionally captured in the wild are probably of aquaculture origin and were released during the several restocking activities that have taken place since 1992 (Congiu et al., 2011; Ludwig et al., 2003). All Adriatic sturgeons released in the past were obtained by artificial reproduction starting from a limited broodstock of 50 animals of wild origin (hereinafter F_0) stocked in an aquaculture plant located in Orzinuovi (Brescia, Italy) since the early 1970s. Besides the F_0 stock, additional few stocks of F_1 animals, produced by artificial generation, were retained in captivity for up to 15 years and are now available as breeders. Currently, the conservation of the Adriatic sturgeon is totally dependent on the correct management of these captive breeders through the establishment of a long-term breeding plan aimed at preserving the genetic diversity available. Unfortunately, no information is available about the degree of relatedness existing among these future F_1 breeders and the establishment of a reliable parental allocation procedure is indispensable for a long-term breeding plan (Congiu et al., 2011).

However, the establishment of a parental allocation procedure on a tetraploid species cannot neglect the modality of inheritance. The genotypes expected in the progeny of a given parent vary under different transmission models. Moreover, the knowledge of the inheritance pattern is basic for the development of virtual offspring simulation procedures that can be very useful in assessing the expected range of genetic variation under different degrees of kinship (Congiu et al., 2011). In fact, knowing the inheritance model at microsatellite loci, allows simulating full or half sibs genotypes that can be used for estimating expected distribution of pairwise relatedness, under different degrees of kinship. This might be especially useful when comparing individuals whose parents are not sampled or no longer exist (Rodzen et al., 2004). The modality of inheritance has never been investigated before, in either the Adriatic sturgeon or in other Palearctic polyploid sturgeons. The only available studies were conducted on two North American tetraploid species: the white sturgeon (*Acipenser transmontanus*) (Rodzen and May, 2002) and the lake sturgeon (*Acipenser fulvescens*) (McQuown et al., 2002; Pyatskowit et al., 2001) by comparing microsatellite genotypes of parents and progeny. These authors inferred allele dosage from band intensity and compared the frequencies of observed genotypes with the ones expected under the pure tetrasomic or disomic inheritance patterns. However, as better developed in the Discussion section, this approach may present some ambiguities due to inference of allele dosage, presence of null alleles and interference between parental genotypes due to allele sharing. Furthermore, since the expected genotype frequencies under the two segregation models can be similar, the analysis of high numbers of fingerlings in order to reach statistical significance is required.

In the present study, we analyzed microsatellite segregation patterns in complete families; however, in order to avoid the possible ambiguities cited above, we performed a preliminary selection of the loci to be analyzed. Only loci for which each single allele could be identified and followed in the progeny were used. For this purpose, all parents of the available families were genotyped in advance at 24 loci, and for each family, only the loci that satisfied the conditions explained in the following section were further analyzed in the progeny. This permitted obtaining, for the first time, unambiguous results about the inheritance pattern in a sturgeon species.

2. Materials and methods

2.1. Preliminary selection of families and loci

All samples analyzed in the present study were obtained from the aquaculture plant Azienda Agricola VIP located in Orzinuovi (Brescia, Italy), where all the breeders of wild origin are reared and where some of them are reproduced every year. We focused on the mating pairs used for reproduction in 2007 and 2010, for which progenies had already been collected or were already available, respectively. All the breeders had been genotyped at 24 loci (Congiu et al., 2011) and the genotypes screened for the selection of loci to be analyzed. A locus was considered to be informative when the following conditions were satisfied: (a) at least one of the parents (the informative one) had a complete heterozygote genotype (four different alleles); (b) no more than one allele was shared by the two parents. In this way (point a), the four alleles of the informative parent are known to be present in a single copy, thus avoiding ambiguities due to allele dosage. Moreover, two of the four alleles are always expected to segregate in every F_1 animal, thus excluding the presence of null alleles. Accordingly, the transmission of each allele could be unambiguously followed in the progeny. With regard to point b, alleles transmitted by the informative parents can also be identified in the progeny when one allele is shared with the mating partner. An example is shown in Fig. 1 in which allele 1 is shared by the two parents. After preliminary selection, 7 families and seven loci were selected for further analyses. Since some families were informative at more than one locus, a total of 12 family-locus combinations were analyzed. In one family group (NaccS17♂ × NaccS19♀) we also had the opportunity to follow the genetic contribution of both parents, both complete heterozygotes without shared alleles.

2.2. Sample collection and laboratory procedures

Genomic DNA was extracted from breeders' fin clips (10–100 mg) and from F_1 larvae, using the DNA Easy Tissue Extraction Mini kit (Euroclone) and stored at −4 °C. A total of 174 F_1 animals were analyzed (Table 1). Loci analyzed are listed in Table 1, together with the optimized annealing temperatures, allelic size range and the number of fingerlings genotyped for each family.

Microsatellite loci were amplified from genomic DNA in 25 µl reactions containing: Taq buffer 1X (GE Healthcare), 1.5 mM $MgCl_2$, 0.4 µM of each primer, 200 µM dNTPs, 0.5 units of Taq (GE Healthcare), and about 50 ng of genomic DNA. All amplifications were performed on GeneAmp PCR System 9700 thermal cyclers (Applied Biosystems).

After checking amplifications on 1.8% agarose gel, genotyping was performed on ABI PRISM 3730XL or ABI 3100 automatic sequencers (external service, BMR Genomics). Scoring was conducted using the software Genotyper version 3.7 (Applied Biosystems).

2.3. Statistical analyses

Likelihood ratio tests with 1 *df* were applied to compare the null model of tetrasomy with the alternative models intermediate between disomic and tetrasomic, following the method proposed by

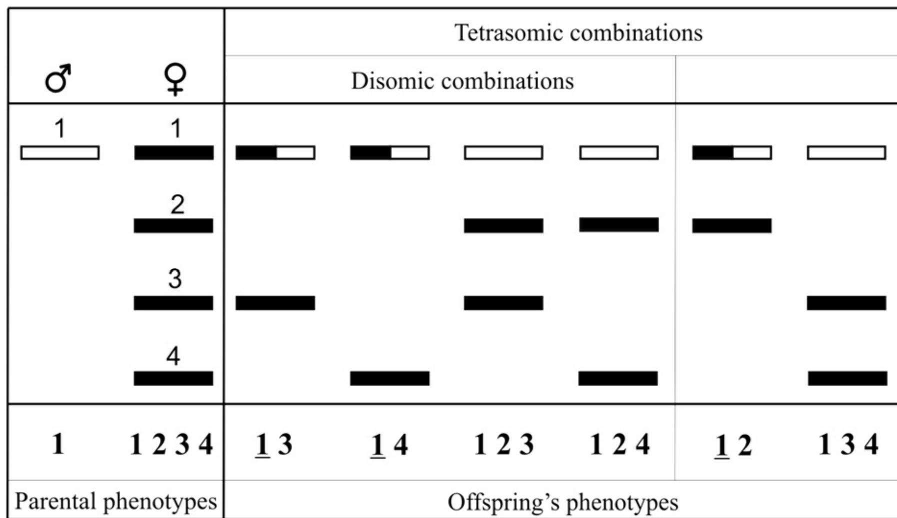


Fig. 1. Expected allele segregation under disomic and tetrasomic inheritance. The figure represents the possible expected phenotypes if one parent presents a completely heterozygote genotype and shares one single allele with the mating partner. Under the hypothesis of disomy, we have supposed pairs 1,2 and 3,4 to be homolog. Consequently allele combinations 12 and 34 are not expected under disomic inheritance while under tetrasomic inheritance all pairwise combinations of alleles are possible. Underlined numerals indicate biparental inheritance of allele 1.

Stift et al. (2008). For each cross-locus combination, we obtained log-likelihood estimates from constrained nonlinear regression models of the alternative inheritance models, using SPSS syntax provided by Stift et al. (2008). Significance levels for multiple comparisons across loci and families were adjusted using the sequential Bonferroni technique (Rice, 1989).

3. Results

At 23 of 24 loci previously analyzed in the F_0 animals, no individuals showed more than four alleles (data not shown). The only exception was represented by locus *D3* in which 23 of 42 F_0 animals showed 5, 6 or 7 alleles. Locus *D3* was already known to be duplicated (Forlani et al., 2008) and was accordingly discarded from this study.

Of the remaining 23 loci, 20 were potentially useful in investigating allele transmission patterns, having four or more alleles. The maximum number of alleles per locus observed was 16 with an average of 7.96 (SD = 3.78) alleles per locus.

Table 2 reports the different allele combinations (with the corresponding frequencies) observed in the progenies of all families at analyzed loci. For each family, the allelic profiles of both parents are reported along with six F_1 profiles, representative of the different allele combinations observed. Alleles inherited from the informative heterozygote parent are marked in bold. The alleles inherited from the other parent were not informative, except for the case of family NaccS17 × NaccS19 at locus *AoxD234*, in which both parents were completely heterozygote.

At all loci analyzed, the six allele combinations expected in the progeny under tetrasomic inheritance were observed. The only exception is represented by locus *AoxD241*, which showed only 5 allele combinations in the progeny of the family NaccS29 × NaccS26, probably due to the limited number of F_1 individuals available. However, the presence of five combinations is sufficient to exclude a disomic inheritance pattern. Moreover, at the same locus, all six allele combinations were observed in the progeny of NaccS30 × NaccS16, confirming tetrasomy. Results of Likelihood ratio tests are reported in Table 3 in which the best fitting model among the ones tested is compared with the null hypothesis of tetraploidy. Strict disomic inheritance was rejected at all loci and no intermediate pattern fitted the observed allele combination frequencies significantly better than the null model. The number of individuals analyzed did not allow us to significantly reject intermediate inheritance models, however, also in the case of a certain degree of preferential chromosome pairing, the rejection of the strict disomic model indicates that all possible pair combinations of parental chromosomes can be expected in the gametes.

Table 1
Technical details of microsatellite screenings of *Acipenser naccarii* families.

Parental pairs (F_0) analyzed	Microsatellite loci	Annealing temperature	Allele size range (bp)	Number of F_1 genotyped
NaccS30 ♂ × NaccS16 ♀	*AnacE4*	57 °C	326–354	28
	AoxD241	57 °C	156–198	25
NaccS31 ♂ × NaccS19 ♀	*AoxD234*	52 °C	215–275	29
	AnacB10	62 °C	212–258	30
NaccS9 ♂ × NaccS33 ♀	*AoxD234*	52 °C	215–275	31
NaccS29 ♂ × NaccS26 ♀	*AoxD241*	57 °C	156–198	20
NaccS31 ♂ × NaccS8 ♀	*AoxD234*	52 °C	215–275	28
NaccS17 ♂ × NaccS19 ♀	*AoxD234*	52 °C	215–275	14
NaccS23 ♂ × NaccS28 ♀	*AoxD64*	60 °C	216–252	23
	AnacE4	57 °C	326–354	22
	AnacA6	62 °C	289–313	23
	An20	62 °C	159–213	23

4. Discussion

4.1. Inheritance pattern

Preliminary genotyping performed on F_0 animals at 24 loci provided information useful for inferring the ploidy level of the Adriatic sturgeon. As mentioned above, the level of ploidy to be ascribed to sturgeon species with about 120 or 240 chromosomes has been debated. The observed presence of no more than four alleles per individual at 19 loci (out of the 20 loci with more than four alleles in the population) suggests that the Adriatic sturgeon should be considered functionally a tetraploid. Only one locus (*D3*) previously reported to be duplicated (Forlani et al., 2008), presented up to seven alleles per individual (data not shown). Locus *D3* is not the one with the highest number of alleles among the ones analyzed in this study, however, more than half the individuals showed five, six or seven alleles. If this observation were to be ascribed to octoploidy, individuals with more than four alleles should have also been observed at loci

Table 2
Single-locus microsatellite inheritance for each sturgeon family. A total of 12 family/locus combinations were studied. Phenotypic profiles of parents and six representative F_1 individuals are reported for each cross. Alleles are reported as allele sizes in bp. The 4 alleles of the complete heterozygote parents are highlighted in bold and marked with capital letters A, B, C or D, also used to label allele combinations observed in the progeny. The missing combination for the family NaccS29 × NeccS26 at locus AoxD241 is marked in gray. Since both NaccS17 and NaccS19 parents were informative at locus AoxD234, alleles observed in NaccS17 are underlined and labeled with lowercase letters. The umber of individuals observed for each allele combination is reported in brackets.

Locus *AoxD234*		Phenotypes					
NaccS9 ♂ ×				**A 235**	**B 243**	**C 247**	**D 275**
NaccS33 ♀		223	227		243		
Observed combinations	AB (5)	223	227	**235**	**243**		
	AC (6)	223	227	**235**		**247**	
	AD (4)	223		**235**	243		**275**
	BC (5)	223	227		**243**	**247**	
	BD (8)	223	227		**243**		**275**
	CD (3)	223	227			**247**	**275**

Locus *AoxD241*		Phenotypes					
NaccS29 ♂ ×		160		172	176		
NaccS26 ♀		**A 160**	**B 164**			**C 180**	**D 188**
Observed combinations	AB (0)						
	AC (2)	160		172		180	
	AD (6)	160		172			188
	BC (4)		164	172	176	180	
	BD (4)	160	164		176		188
	CD (4)			172	176	180	188

Locus *AoxD234*		Phenotypes					
NaccS31 ♂ ×		215	219				
NaccS19 ♀				**A 227**	**B 231**	**C 251**	**D 259**
Observed combinations	AB (9)		219	**227**	**231**		
	AC (5)		219	**227**		**251**	
	AD (4)		219	**227**			**259**
	BC (3)	215	219		**231**	**251**	
	BD (6)	215	219		**231**		**259**
	CD (2)		219			**251**	**259**

Locus *AnacB10*		Phenotypes					
NaccS31 ♂ ×		215			249	252	
NaccS19 ♀			**A 224**	**B 243**	**C 249**		**D 258**
Observed combinations	AB (1)		**224**	**243**	249		
	AC (2)	215	**224**		**249**		
	AD (6)	215	**224**			252	**258**
	BC (10)			**243**	**249**	252	
	BD (8)	215		**243**		252	**258**
	CD (3)	215			**249**	252	**258**

Locus *AoxD234*		Phenotypes					
NaccS31 ♂ ×		215	219				
NaccS8 ♀				**A 223**	**B 227**	**C 243**	**D 251**
Observed combinations	AB (8)	215	219	**223**	**227**		
	AC (4)		219	**223**		**243**	
	AD (6)		219	**223**			**251**
	BC (6)	215	219		**227**	**243**	
	BD (1)		219		**227**		**251**
	CD (3)	215	219			**243**	**251**

Locus *AoxD234*		Phenotypes							
NaccS17 ♂ ×		a 223			b 243	c 247		d 255	
NaccS19 ♀			**A 227**	**B 231**			**C 251**		**D 259**
Observed combinations	AB (1) /ad (1)	223	**227**	**231**				255	
	AC (2)/bd (1)		**227**		243		**251**	255	
	AD (2)/cd (4)		**227**			247		255	**259**
	BC (2)/ab (3)	223		**231**	243		**251**		
	BD (4)/cd			**231**		247		255	**259**
	CD (3)/bc (1)				243	247	**251**		**259**

Locus *AnacE4*		Phenotypes				
NaccS30 ♂ ×		332		344		
NaccS16 ♀			A **340**	B **344**	C **352**	D **354**
Observed combinations	AB (5)	332	**340**	**344**		
	AC (9)	332	**340**		**352**	
	AD (4)	332	**340**			**354**
	BC (2)	332		**344**	**352**	
	BD (7)	332		**344**		**354**
	CD (1)	332			**352**	**354**

Locus *AoxD241*		Phenotypes						
NaccS30 ♂ ×		A **160**			B **172**	C **176**	D **180**	
NaccS16 ♀			164	168				184
Observed combinations	AB (2)	**160**	164	168	**172**			
	AC (8)	**160**		168		**176**		184
	AD (2)	**160**					**180**	184
	BC (2)			168	**172**	**176**		184
	BD (2)			168	**172**		**180**	
	CD (9)		164			**176**	**180**	184

Table 2 (*continued*)

Locus *AnacA6*		Phenotypes				
NaccS23 ♂ × **NaccS28** ♀		A **297**	B **299**		C **307**	D **313**
			299	301		
Observed combinations	AB (2)	**297**	**299**			
	AC (7)	**297**	299	301	**307**	
	AD (4)	**297**	299	301		**313**
	BC (3)		**299**	301	**307**	
	BD (5)		**299**	301		**313**
	CD (2)		299	301	**307**	**313**

Locus *An20*		Phenotypes					
NaccS23 ♂ × **NaccS28** ♀				A **171**	B **185**	C **189**	D **193**
		159	163	171			
Observed combinations	AB (6)	159		**171**	**185**		
	AC (2)	159	163	**171**		**189**	
	AD (4)	159		**171**			**193**
	BC (1)	159	163		**185**	**189**	
	BD (7)		163	171	**185**		**193**
	CD (3)	159	163			**189**	**193**

Locus *AoxD64*		Phenotypes				
NaccS23 ♂ × **NaccS28** ♀		216		228		
		A **216**	B **224**		C **232**	D **326**
Observed combinations	AB (3)	**216**	**224**	228		
	AC (5)	**216**		228	**232**	
	AD (3)	**216**		228		**326**
	BC (2)		**224**	228	**232**	
	BD (3)	216	**224**	228		**326**
	CD (7)	216		228	**232**	**326**

Locus *AnacE4*		Phenotypes					
NaccS23 ♂ × **NaccS28** ♀		332	336		346		
		A **332**		B **344**		C **352**	D **354**
Observed combinations	AB (4)	**332**		**344**			
	AC (6)	**332**	336			**352**	
	AD (1)	**332**	336				**354**
	BC (4)		336	**344**	346	**352**	
	BD (4)	332		**344**	346		**354**
	CD (3)		336		346	**352**	**354**

with a comparable or higher level of polymorphism. The absence of individuals with more than four alleles at all remnant loci suggests that locus *D3* maps to a duplicated region in the genome, as already reported by Forlani et al. (2008).

At all seven loci analyzed, all possible allele combinations were observed, thus rejecting a pattern of strict disomic inheritance. At three loci (*AnacE4*, *AoxD241* and *AoxD234*) we had the possibility to analyze more than one family and these results were always confirmed. Results of the constrained nonlinear regression analyses revealed that intermediate inheritance models did not fit the observed allele combination frequencies significantly better than the null model of pure tetrasomy (TAU = 1) in any of the cross-locus combinations.

Therefore our data do not support the existence of any preferential pairing, although statistical power was limited and a larger sample size would be needed to obtain more confidence. The possibility to exclude the pure disomic pattern has important implications for the management of captive genetic diversity. For example, the possibility of observing all the parental allele combinations in the gametes confirmed the validity of the approaches previously used for the allocation of the animals to the corresponding familiar groups and for the

Table 3
Results of Likelihood ratio test between the null model of tetrasomy (TAU = 1) and the best fitting intermediate one (TAU estimated). In family NaccS17 × NaccS19 maternal and paternal contributions are indicated with (m) and (p) respectively.

Locus	Families	n	Null model (T = 1) LIKELIHOOD OBS	Best intermediate model			Model comparison: LRT	p-values
				Pairing alleles	T	LIKELIHOOD OBS		
AoxD234	NaccS9 × NaccS33	31	55,54	AB/CD	0,77	55,13	0,41	**0,26**
	NaccS31 × NaccS19	29	51,96	AD/BC	0,72	51,38	0,58	**0,22**
	NaccS31 × NaccS8	28	50,17	AC/BD	0,54	48,49	1,68	**0,10**
	NaccS17 × NaccS19 (p)	14	25,08	AB/CD and AD/BC	0,86	25,01	0,07	**0,39**
	NaccS17 × NaccS19 (m)	14	25,08	AD/BC	0,43	23,76	1,32	**0,13**
AoxD241	NaccS29 × NaccS26	20	35,84	AB/CD	0,60	34,96	0,87	**0,17**
	NaccS30 × NaccS16	25	44,79	–	0,48	42,88	1,92	**0,08**
AnacB10	NaccS31 × NaccS19	30	53,75	AB/CD	0,40	50,60	3,16	**0,04**[a]
AnacE4	NaccS30 × NaccS16	28	50,17	AB/CD and AD/BC	0,64	49,21	0,96	**0,16**
	NaccS23 × NaccS28	22	39,42	AD/BC	0,68	38,82	0,59	**0,22**
AnacA6	NaccS23 × NaccS28	23	41,21	AB/CD	0,52	39,74	1,47	**0,11**
AoxD64	NaccS23 × NaccS28	23	41,21	AD/BC	0,65	40,46	0,75	**0,19**
An20	NaccS23 × NaccS28	23	41,21	AD/BC	0,65	40,46	0,75	**0,19**

[a] Not significant after sequential Bonferroni correction.

simulation of virtual progenies under the assumption of tetrasomy (Congiu et al., 2011).

The decision to limit the analysis to fully informative loci provided, for the first time in a tetraploid sturgeon species, an indication of the inheritance pattern without the need for strong assumptions. The main problems of all the studies previously performed on tetraploid sturgeon species were due to the assumption that band intensity on a gel was proportional to the number of allele copies. However, a given profile might correspond to several different genotypes, making this approach less reliable for allele dosage. For example, a profile with two alleles (A and B) can be observed for six different genotypes: ABBB, AABB, AAAB, ØAAB, ØABB, ØØAB, where Ø represents a null allele. The corresponding A/B intensity ratios are 1/3, 1/1, 3/1, 2/1, 1/2 and 1/1 respectively, which are very hard or impossible to distinguish, especially considering the intrinsic variability of PCR efficiency. A second difficulty encountered by previously performed studies on sturgeon inheritance patterns was caused by the incomplete heterozygosity of the parental genotypes analyzed. Incomplete heterozygosity implies that a limited number of allele combinations is expected in the progeny. Consequently, the same genotypes are often expected under tetrasomy or disomy and the two transmission models can be distinguished only by comparing observed and expected genotype frequencies in the progeny. An example is given by a parental genotype with two alleles A and B in which we suppose that the genotype AABB could be deduced from band intensity and that no null alleles are present. The expected segregating genotypes are AA, AB and BB with the respective ratios of 1/6, 2/3 and 1/6 under tetrasomic inheritance and 1/4, 1/2 and 1/4 under disomy (when the homolog allele pairs are AB and AB). The only way to distinguish tetrasomy from disomy would be by analyzing genotype frequencies through the genotyping of a high number of individuals.

This background, together with the strong assumptions related to allele dosage and to the presence of null alleles, justifies the choice of the approach adopted in this study. Furthermore, the results presented here were obtained by analyzing a relatively low number of individuals compared to previous investigations, drastically reducing the experimental effort.

4.2. Evolutionary implications

The high incidence of natural interspecific hybridization among sturgeon species (Birstein and Bemis, 1997) might lead to considering allopolyploidization to be the more plausible cause of polyploidization in sturgeons. However, tetrasomic inheritance is usually associated with autotetraploidy (Stift et al., 2008). The results reported here suggest that the tetraploid condition of the Adriatic sturgeon could originate from an autopolyploidization event in which only one species was involved.

The Adriatic sturgeon belongs to a cluster composed of four 240-chromosome species that also includes the Russian sturgeon (*Acipenser gueldenstaedtii*), the Persian sturgeon (*Acipenser persicus*) and the Siberian sturgeon (*Acipenser baerii*) (Ludwig et al., 2001). These four species are characterized by high genetic similarity (Birstein and DeSalle, 1998; DeSalle and Amato, 2004; Ludwig, 2008). Unfortunately, no information about the inheritance patterns in the other species of this group is available. However, the duplication event that led to tetrasomy in the Adriatic sturgeon was suggested to have occurred in a common ancestor of this group of species (Ludwig et al., 2001). If this hypothesis is true, the approach used in the present study is expected to give analogous results also in the above listed close-related species.

The coexistence of tetrasomic and disomic inheritance models in the same genome (with different chromosomes following different patterns) cannot be excluded, as suggested for the lake sturgeon (*A. fulvescens*) by McQuown et al. (2002). This might be ascribed to a secondary functional diploidization (Fontana et al., 2008; Tagliavini et al., 1999). In allotetraploids, for example, if the original hybridization event took place between very similar genomes, different chromosomes may have different degrees of homology, with some of them being totally homolog (i.e. tetrasomic) and others different enough to have caused the distinction of the parental species (Stift et al., 2008). Alternatively, completely homolog chromosome tetrads arising from an autotetraploidization event may experience a secondary differentiation in two homeolog pairs (Stift et al., 2008). For this reason, extending the inheritance pattern observed at only seven loci to a whole genome composed of about 60 tetrads would be a strong assumption. The inheritance modality should be verified, if possible, at all loci used for any application concerning inheritance. Moreover it should be underlined that only tetraploid loci were considered in this study. We cannot exclude that the above described process of functional reduction of ploidy experienced by tetraploid sturgeon species (Fontana et al., 2008) have further differentiated some loci to diploidy, as observed in other species such as the lake sturgeon (Pyatskowit et al., 2001).

5. Conclusion

In conclusion, the analyses of previously selected family groups at fully informative loci yielded concordant indications about the microsatellite inheritance pattern in the tetraploid Adriatic sturgeon. The rejection of the pure disomic pattern provided useful information for the management of its genetic variability and suggests tetrasomy as the most likely inheritance pattern for this species. In order to exclude intermediate models of inheritance with a higher statistical confidence, more individuals should have been analyzed. However, for what concerns the practical management of the genetic variability, our results confirm the possibility to observe all the possible combinations of parental alleles in the progeny, which represents a relevant contribution to the establishment of an optimal parental allocation procedure. Our results also suggest that the current ploidy level in *A. naccarii* likely was reached through an autopolyploidization event. To our knowledge, these results represent the first data concerning the modality of segregation in the Adriatic sturgeon, and, more generally, in Palearctic tetraploid sturgeons.

Acknowledgments

We sincerely thank Marc Stift for his help in statistical testing of segregation models, Sara Vianello and Gaia Cecilia Santini for having preliminary genotyped some animals. This study was funded by University of Padova grant CPDA087543/08.

References

Birstein, V.J., Bemis, W.E., 1997. How many species are there within the genus *Acipenser*? In: Birstein, V.J., Waldman, J.R., Bemis, W.E. (Eds.), Sturgeon Biodiversity and Conservation. Kluwer Academic Publishers, Dordrecht, pp. 157–163.

Birstein, V.J., DeSalle, R., 1998. Molecular phylogeny of Acipenserinae. Molecular Phylogenetics and Evolution 9, 141–155.

Birstein, V.J., Vasil'ev, V.P., 1987. Tetraploid–octoploid relationships and karyological evolution in the order Acipenseriformes (Pisces): karyotypes, nucleoli, and nucleolus-organizer regions in four acipenserid species. Genetica 73, 3–12.

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. doi:10.1371/journal.pone.0018249.

DeSalle, R., Amato, G., 2004. The expansion of conservation genetics. Nature Reviews 5, 702–712.

Doukakis, P., Babcock, E.A., Pikitch, E.K., Sharov, A.R., Baimukhanov, M., Erbulekov, S., Bokova, Y., Nimatov, A., 2010. Management and recovery options for Ural River beluga sturgeon. Conservation Biology 24, 769–777.

Fontana, F., Zane, L., Pepe, A., Congiu, L., 2007. Polyploidy in Acipenseriformes: cytogenetic and molecular approaches. In: Pisano, E., Ozouf-Costaz, C., Foresti, F., Kapoor, B.G. (Eds.), Fish Cytogenetics. Science Publisher, Inc, Enfield, New Hampshire, USA, pp. 385–403.

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.

Forlani, A., Fontana, F., Congiu, L., 2008. Isolation of microsatellite loci from the endemic and endangered Adriatic sturgeon (*Acipenser naccarii*). Conservation Genetics 9, 461–463.
Ludwig, A., 2006. A sturgeon view on conservation genetics. European Journal of Wildlife Research 52, 3–8.
Ludwig, A., 2008. Identification of Acipenseriformes species in trade. Journal of Applied Ichthyology 24, 2–19.
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.
Ludwig, A., Congiu, L., Pitra, C., Fickel, J., Gessner, J., Fontana, F., Patarnello, T., Zane, L., 2003. Nonconcordant evolutionary history of maternal and paternal lineages in Adriatic sturgeon. Molecular Ecology 12, 3253–3264.
McQuown, E., Gall, G.A.E., May, B., 2002. Characterization and inheritance of six microsatellite loci in lake sturgeon. Transactions of the American Fisheries Society 131, 299–307.
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, 854–862.
Pikitch, E.K., Doukakis, P., Lauck, L., Chakrabarty, P., Erickson, D.L., 2005. Status, trends and management of sturgeon and paddlefish fisheries. Fish and Fisheries 6, 233–265.
Pyatskowit, J.D., Krueger, C.C., Kincaid, H.L., May, B., 2001. Inheritance of microsatellite loci in the polyploid lake sturgeon (*Acipenser fulvescens*). Genome 44, 185–191.
Rice, W.R., 1989. Analyzing tables of statistical tests. Evolution 43, 223–225.
Rodzen, J.A., May, B., 2002. Inheritance of microsatellite loci in the white sturgeon (*Acipenser transmontanus*). Genome 45, 1064–1076.
Rodzen, J.A., Famula, T.R., May, B., 2004. Estimation of parentage and relatedness in the polyploid white sturgeon (*Acipenser transmontanus*) using a dominant marker approach for duplicated microsatellite loci. Aquaculture 232, 165–182.
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, 2113–2123.
Tagliavini, J., Williot, P., Congiu, L., Chicca, M., Lanfredi, M., Rossi, R., 1999. Molecular cytogenetic analysis of the karyotype of the European Atlantic sturgeon, *Acipenser sturio*. Heredity 83, 520–525.

ANNEX B

MOLECULAR ECOLOGY RESOURCES
Molecular Ecology Resources (2014) doi: 10.1111/1755-0998.12203

Species and hybrid identification of sturgeon caviar: a new molecular approach to detect illegal trade

E. BOSCARI,* A. BARMINTSEVA,† J. M. PUJOLAR,*‡ P. DOUKAKIS,§ N. MUGUE†¶ and L. CONGIU*
*Department of Biology, University of Padova, I-35131 Padova, Italy, †Russian Federal Research Institute of Fisheries and Oceanography, Moscow 107140, Russia, ‡Department of Bioscience, Aarhus University, DK-8000 Aarhus, Denmark, §Scripps Institute of Oceanography, University of California San Diego, La Jolla, CA 92093, USA, ¶Institute of Developmental Biology RAS, Moscow 117808, Russia

Abstract

Overexploitation of wild populations due to the high economic value of caviar has driven sturgeons to near extinction. The high prices commanded by caviar on world markets have made it a magnet for illegal and fraudulent caviar trade, often involving low-value farmed caviar being sold as top-quality caviar. We present a new molecular approach for the identification of pure sturgeon species and hybrids that are among the most commercialized species in Europe and North America. Our test is based on the discovery of species-specific single nucleotide polymorphisms (SNPs) in the ribosomal protein S7, supplemented with the Vimentin gene and the mitochondrial D-loop. Test validations performed in 702 specimens of target and nontarget sturgeon species demonstrated a 100% identification success for *Acipenser naccarii*, *A. fulvescens*, *A. stellatus*, *A. sinensis* and *A. transmontanus*. In addition to species identification, our approach allows the identification of Bester and AL hybrids, two of the most economically important hybrids in the world, with 80% and 100% success, respectively. Moreover, the approach has the potential to identify many other existing sturgeon hybrids. The development of a standardized sturgeon identification tool will directly benefit trade law enforcement, providing the tools to monitor and regulate the legal trade of caviar and protect sturgeon stocks from illicit producers and traders, hence contributing to safeguarding this group of heavily threatened species.

Keywords: caviar, hybrids, nuclear markers, species identification, sturgeons

Received 19 August 2013; revision received 29 October 2013; accepted 30 October 2013

Introduction

Sturgeons (order Acipenseriformes) are large primitive fishes including about 25 species with particular life history traits. Sturgeons are long-lived organisms, with lifespans up to 100 years, show late maturation (5 to more than 30 years) and do not reproduce annually (2 to more than 10 years between reproductive cycles) (Pikitch *et al.* 2005). Multiple levels of ploidy exist in sturgeons, including diploid, tetraploid and hexaploid species with about 120, 240 and 360 chromosomes, respectively (Fontana *et al.* 2007). Once widely distributed and abundant in temperate waters of the northern hemisphere, sturgeons exist today in the wild only as fragmented and isolated populations with a limited geographical distribution (Pikitch *et al.* 2005; Ludwig 2008). The dramatic decline in sturgeon abundance observed in the last decades has lead the International Union for Conservation of Nature (IUCN) to identify sturgeons as the most endangered group of species in the world, with 85% of sturgeons being at risk of extinction according to the Red List of Threatened Species (www.iucnredlist.org). Together with pollution and habitat degradation, the principal reason for the highly threatened status of sturgeons is overexploitation of wild populations due to the high economic value of caviar, a luxury delicacy made of refined eggs of sturgeons, one of the most valuable food items of animal origin (Ludwig 2008; Fain *et al.* 2013; Jahrl 2013).

The severe decline of natural stocks in recent years has prompted a parallel increase in sturgeon aquaculture programmes in Europe, Asia and North America with a special interest in those species that either produce high-quality caviar or demonstrate enhanced growth rates in aquaculture. Since 2002, import of caviar from aquaculture has increased, while import of caviar from the wild has decreased (Bronzi *et al.* 2011). In 2006, global imports

Correspondence: Leonardo Congiu, Fax: +39(0)49-8276209;
E-mail: leonardo.congiu@biology.unipd.it

© 2013 John Wiley & Sons Ltd

of farmed caviar (20 tonnes) were for the first time nearly equivalent to those from wild production (24 tonnes) (Ishihara *et al.* 2010). Estimated world production of farmed caviar in 2008 was 110–120 tonnes originating from 16 countries. According to FAO data, current production of meat and caviar in aquaculture largely exceeds production from wild populations (Bronzi *et al.* 2011).

At present, a substantial portion of aquaculture caviar production comes from sturgeon hybrids. For instance, in China, 38% of total biomass production is attributed to farmed hybrids (Wei *et al.* 2011). Hybrid production for caviar usually involves species with same level of ploidy, and although hybridization between species with different degree of ploidy is possible, the resulting hybrids are either sterile or show a reduced production of eggs (Birstein 2002). Early research into hybrid production focused on Bester, a hybrid between two diploid species: Beluga (*Huso huso*) females and Sterlet (*Acipenser ruthenus*) males. This hybrid produces high-quality caviar at a younger maturity age as compared to pure beluga. Many other hybrid combinations are now cultivated. Among these, the following ones obtained from tetraploid species are particularly relevant: AL, a hybrid between Adriatic sturgeon (*Acipenser naccarii*) females and Siberian sturgeon (*Acipenser baerii*) males; Rolik, a hybrid between Russian sturgeon (*Acipenser gueldenstaedtii*) females and Siberian sturgeon males and the bidirectional hybrid between Kaluga sturgeon (*Huso dauricus*) and Amur sturgeon (*Acipenser schrenckii*) (Wei *et al.* 2011). In recent years, additional hybrids such as *A. schrenckii* × *A. baerii* and *A. gueldenstaedtii* × *A. naccarii* (GUNA) have shown up for the first time in international trade (CITES trade statistics derived from the CITES Trade Database, UNEP World Conservation Monitoring Centre, Cambridge, UK).

The presence of hybrid-origin caviar or meat on the international market poses a challenge to enforcing international trade regulation. Since 1998, sturgeons have been listed under the Convention on International Trade in Endangered Species (CITES; www.cites.org). Most species are listed under CITES Appendix II, which allows for sustainable trade under strict guidelines regulated by governments. Trade in sturgeon parts or derivatives must be accompanied by an export permit that indicates the species harvested and is issued on the basis of sustainable harvest. The European Union has taken measures to fight illegal caviar trade and implemented in 2006 a universal labelling system for caviar traded under CITES that includes species and country of origin (EU regulation No. 865/2006). Hybrid caviar must be labelled and accompanied by an export quota even though CITES does not require the issuance of annual quotas for hybrid caviar in trade.

Illegal harvest and trade of caviar increased dramatically in the 1990s, flooding the international market with illegal, low-quality and inexpensive caviar (Knapp *et al.* 2006). At present, a large portion of the global caviar trade is thought to be illegal, possibly exceeding legal trade by several times (Knapp *et al.* 2006). Concern over fraudulent mislabelling of the species origin of caviar in international trade and domestic markets triggered the development of molecular markers for the identification of caviar products (DeSalle & Birstein 1996; Birstein *et al.* 2000; Congiu *et al.* 2001, 2002; Mugue *et al.* 2008; Doukakis *et al.* 2012). Most sturgeons can nowadays readily be identified to the species level using mitochondrial genes. However, those fail to identify some closely related and commercially important species such as the ones included in the so-called '*gueldenstaedtii complex*': Russian (*A. gueldenstaedtii*), Adriatic (*A. naccarii*), Siberian (*A. baerii*) and Persian (*Acipenser persicus*) (Knapp *et al.* 2006; Doukakis *et al.* 2012). Failure of species identification by mitochondrial markers is due to the high genetic variability of all four species of this complex as well as to ancient introgression events (Ludwig 2008). Only recently, the identification of a diagnostic polymorphism in the mitochondrial control region allowed distinction of Russian from Siberian sturgeon (Mugue *et al.* 2008). Proposed methods for genetic identification of sturgeon hybrids include random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphic (AFLP) DNA, but their use is discouraged because reproducibility is highly sensitive to experimental conditions and overall too expensive for routine application (Congiu *et al.* 2001, 2002; Rozhkovan *et al.* 2008; Yarmohammadi *et al.* 2012).

The objective of our study is to establish a new molecular approach for the identification of pure sturgeon species and hybrids of the most commercialized species in Europe and North America. Those included a total of 11 pure sturgeon species, including the four species of the '*gueldenstaedtii complex*' (*A. gueldenstaedtii*, *A. naccarii*, *A. baerii* and *A. persicus*) and two of the most commercially important hybrids in Europe and North America, Bester (*H. huso* female × *A. ruthenus* male) and AL (*A. naccarii* female × *A. baerii* male). The method is based on the discovery of species-specific diagnostic single nucleotide polymorphisms (SNPs) in the first intron of the nuclear encoded S7 ribosomal protein RPS7 and supplementary data from the mitochondrial D-loop and the Vimentin gene. The method is easy and fast, inexpensive, and reliable and reproducible across laboratories. Ultimately, the development of a standardized identification protocol for pure sturgeon species and hybrids provides the tools to control and regulate the global legal trade in sturgeon products and combat illegal and fraudulent caviar trade (e.g. low-value farmed caviar sold as

© 2013 John Wiley & Sons Ltd

top-quality hybrid-origin farmed caviar). Besides supporting trade law enforcement, the method is important for the management and conservation of this extremely threatened group of fish.

Materials and methods

Sampling and DNA extraction

A total of 98 individuals from 11 different sturgeon species (*Acipenser baerii*, *A. gueldenstaedtii*, *A. persicus*, *A. ruthenus*, *A. transmontanus*, *A. naccarii*, *A. fulvescens*, *A. sinensis*, *A. stellatus*, *A. sturio* and *Huso huso*) were selected for sequencing of the first intron RP1 of the RPS7 gene and species-specific primer design (Table S1). When possible, samples from different geographical locations were considered to have a better representation of the genetic variability of each species. Primer validation was performed for each primer pair on samples of target and nontarget species (excluding *A. sturio*) for a total of 686 animals. Results were checked by comparing the band sizes obtained with the expected specific band sizes as shown in Fig. 1. Additionally, a preliminary investigation of *Acipenser schrenckii* and *Huso dauricus* was conducted on 10 individuals (of aquaculture origin) per species and 90 individuals from 10 nontarget species.

Six individuals belonging to three different sturgeon species (one *A. baerii*, three *A. gueldenstaedtii*, two *A. persicus*) were selected for Vimentin sequencing and subsequent primer design. Vimentin validation was performed on a total of 309 individuals: 56 *A. baerii*, 96 *A. gueldenstaedtii* (48 from Caspian and 48 from Azov Sea), 48 *A. persicus*, 45 *A. naccarii*, 8 *A. ruthenus*, 8 *A. transmontanus*, 8 *A. fulvescens*, 8 *A. sinensis*, 8 *A. stellatus*, 8 *A. schrenckii*, 8 *Huso dauricus* and 8 *H. huso*.

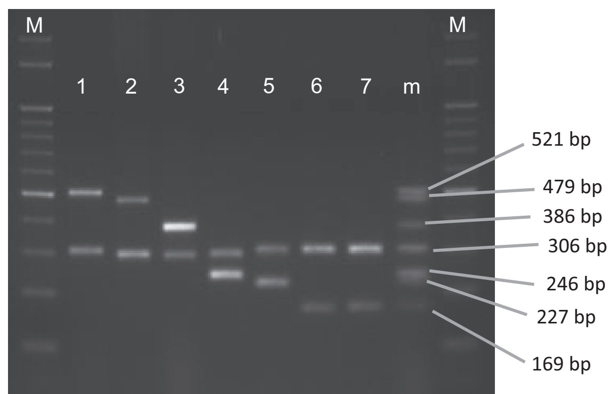


Fig. 1 Expected profiles and corresponding band sizes on agarose gel for pure species identification using RP1 including (1) *Acipenser fulvescens*, (2) *A. stellatus*, (3) *A. sinensis*, (4) *A. transmontanus*, (5) *A. naccarii*, (6) *A. baerii* and (7) *A. ruthenus*. The 306 pb band represents the positive control with slight variable sizes due to different deletions in the different species. The band of 223 bp identified as a potential marker for *Acipenser schrenckii* and *Huso dauricus* was not included.

Finally, a total of 91 individuals from two commercially important hybrids were used to develop a hybrid identification tool: 71 Bester (*H. huso* female × *A. ruthenus* male) and 20 AL (*A. naccarii* female × *A. baerii* male).

Genomic DNA was extracted from fin clips, muscle or caviar eggs (50 ng) using the standard protocol of the EuroGOLD Tissue DNA Mini Kit (Euroclone).

Development of RP1 marker

Initially, the first intron (RP1) of the nuclear encoded S7 ribosomal protein (RPS7) was amplified from three individuals of three different species (*A. gueldenstaedtii*, *A. transmontanus* and *H. huso*) using the universal primers for fish isolated by Chow & Hazama (1998), S7RPEX1F and S7RPEX1R. Following the identification of multiple loci (Figs S1 and S2), a new primer was designed to specifically amplify locus A (RP1_LocusA_R 5′ATCCAAGTACAAGCTTGAACA3′), which was used in combination with the universal primer S7RPEX1F.

All PCRs were performed in a total volume of 20 μL, consisting of 0.25 μM of each primer, 0.1 mM dNTPs, 1× buffer (500 mM KCl, 15 mM $MgCl_2$, 100 mM Tris-HCl pH 8.3, 1 mg/mL BSA, 100 mM $(NH_4)_2SO_4$, 0.025 U/μL of Taq DNA polymerase (Sileks or Resnova) and 10 ng of extracted DNA template. All amplifications were performed on an Applied Biosystem GeneAmp®PCR System 9700 set as follows: 2 min at 94 °C, 33–35 cycles at 94 °C for 45 s, 59 °C for 45 s and 72 °C for 45 s, followed by a 7-min extension at 72 °C. Products were checked by 1.8% agarose gel electrophoresis, purified using ExoSAP-IT® and directly sequenced on an ABI Prism 3730XL or an ABI 3100 automatic sequencer at BMR Genomics.

Sequences presenting no polymorphism were considered as single alleles (homozygote individuals). For sequences with double peaks, the consensus sequence was edited according to the IUPAC code. Heterozygote individuals, presenting insertions/deletions that resulted in chromatogram shifting, were cloned to identify all alleles for each individual. Cloning procedure was performed in JM109 competent cells using the P-GEM-T Easy Vector (Promega) following the manufacturer's recommendations. Considering that sequencing of cloned PCR products is known to generate errors through single-base mutations (Kobayashi *et al*. 1999), sequences were considered to be true alleles only if the comparison between cloned sequences and observed peaks in the genomic sequence profiles confirmed the presence of each polymorphism (Barbisan *et al*. 2009). Moreover, unclear sequences were cloned again from a

© 2013 John Wiley & Sons Ltd

second independent PCR. Locus A was sequenced for a total of 98 individuals from 11 species to detect species-specific SNPs. Sequences were edited using FinchTV (http://www.geospiza.com/Products/finchtv.shtml) and aligned using ClustalW (http://www.ebi.ac.uk/Tools/clustalw/index.html) (Thompson *et al.* 1994).

A multi-alignment was generated using all sequences obtained, and species-specific SNPs were identified. A panel of nine species identification primers were developed using Oligo-Explorer (Gene Link) and FastPCR (Primer Digital) (Kalendar *et al.* 2009). Diagnostic mutations were used for primer design following three general criteria: first, the last nucleotide at the 3′ primer-end must be complementary to the diagnostic mutation; second, the second-last nucleotide should be not complementary to the target sequence in order to establish a mismatch of two nucleotides if paired with nontarget sequences, which causes a detachment from the target DNA (Mugue *et al.* 2008); third, primers were chosen so that different species would differ in PCR product size and could be easily distinguishable on 2% agarose gel. All primer pairs were optimized and checked against target and nontarget species in uniplex reactions for a total of 702 animals. Validation experiments were performed independently in two laboratories located in Italy and Russia, respectively. The same PCR conditions reported above were used for all primer pairs to facilitate multiplexing.

Development of Vimentin marker

With the original aim of developing a marker to investigate a supposed ancient introgression of *A. baerii* into *A. gueldenstaedtii* in the Caspian Sea (Mugue *et al.* 2008) (thus expected to be absent in *A. gueldenstaedtii* from Azov Sea), a second nuclear marker (Vimentin) was analysed. The Vimentin sequence of *A. baerii* (AJ493266) was aligned to the *Danio rerio* sequence (NP_57194), and exon–intron boundaries were predicted. The primer pairs AbVim_F (5′GTCTACAATGAATCGGCAGTCGTC3′) and AbVim_R (5′TCCCCAAGGTTGTCTCGGTCTAC3′) were designed to amplify the first coding exon.

PCR products were amplified from one *A. baerii*, three *A. gueldenstaedtii* (Azov and Caspian Seas) and two *A. persicus* individuals, ligated in P-GEM-T Easy Vector (Promega) and cloned in *Escherichia coli*. A multi-alignment was generated using all sequences obtained, and polymorphisms were identified. One primer pair consisting of Bae154B7F (5′TCCAGGGTTTCCTACACCAGCCAAT3′) and Bae154B7R (5′CCACCCTCGCCTTTTCGTTGGTTTG3′) was designed to amplify alleles present in *A. baerii*, but rarely found in *A. gueldenstaedtii* individuals from Azov. The criteria for selective primer construction were the same as for RP1 with the reverse primer designed on the selected SNP. The expected amplification product is 373 bp long.

Vimentin mix reactions were the same as reported for RP1 with $MgCl_2$ increased to 2.5 mM. Amplifications were performed with the following cycle: 2 min of denaturation at 94 °C, 30 cycles at 94 °C for 15 s, 61 °C for 30 s and 72 °C for 30 s, followed by a 10-min extension at 72 °C. As in RP1, Vimentin was validated independently in two laboratories.

Hybrid detection tools

In the case of Bester, the paternal contribution was identified using primer RutBae_RP1F designed for Siberian (*A. baerii*) and Sterlet (*A. ruthenus*) sturgeons in combination with the reverse primer specific for RP1-Locus A (RP1_LocusA_R). This amplification was multiplexed with the positive amplification control following the same thermal cycle as for the uniplex-PCR. In the multiplex reaction, 0.25 μM of each primer forward and 0.5 μM of the reverse primer were used. The maternal contribution was identified by mitochondrial (mtDNA) DNA analysis following Mugue *et al.* (2008).

In the case of AL, the paternal (*A. baerii*) contribution was identified using the primer pair Bae154B7F/Bae154B7F for Vimentin. The maternal (*A. naccarii*) contribution was identified using the primer pair Nac_RP1F/NacFul_RP1R. The maternal contribution can be confirmed by analysis of the mitochondrial control region and comparison with the full list of existing haplotypes identified for *A. naccarii* (GenBank Accession nos KF771109–KF771115) (Congiu *et al.* 2011), with an additional haplotype referred to a putative case of sequence heteroplasmy.

Estimation of hybrid detection power

While for pure species, the expected identification power of the markers is directly inferred from the percentages observed in the validation experiments, the expected power in the case of hybrids depends on the probability that the diagnostic marker for the corresponding paternal species is inherited, which depends on its frequency and on the ploidy level. Dealing with dominant biallelic markers (presence/absence of a given band), the allelic frequencies of the 'absence' alleles can be estimated, in pure species, from the frequencies of the complete homozygous genotypes (e.g. individuals not presenting the amplification product) by applying the square root for diploids (for tetraploids, the fourth root should be applied). By doing this, we accepted the error given by the assumption of Hardy–Weinberg equilibrium in samples composed by individuals from different geographical areas. Based on allelic frequencies and taking into

© 2013 John Wiley & Sons Ltd

account that hybrids may inherit one or two paternal chromosomes (in diploids and tetraploids, respectively), we estimated the identification power as the proportion of hybrids expected to have inherited at least one copy of the 'presence' allele from the paternal species.

Results

RP1 characterization

Amplification of the RP1 by universal primers resulted in two bands for each individual distinguishable on agarose gel, of around 950 and 700 bp, respectively. After cloning of the PCR products, sequencing and cluster analysis showed 3 groups of sequences, A, B and C (Figs S1 and S2). The highest band on the agarose gel corresponded to the group B, while groups A and C presented the same size, but were characterized by different insertions/deletions as shown in Fig. S1.

The presence of at least one sequence per group in each individual suggests the existence of 3 different loci (A, B and C in Fig. S2). All subsequent analyses were focused on the characterization of locus A. To selectively amplify this group of sequences, a new locus A-specific reverse primer was designed (RP1_LocusA_R) to be used in combination with the universal primer (S7RPEX1F). A maximum of four different alleles for tetraploid species and two alleles for diploid species were observed, which points to the selective amplification of a single locus. A multi-species alignment of 583 bp was obtained including over 150 different sequences obtained from 11 species. A simplified alignment of one sequence per species is reported in Fig. S3.

Species-specific primers set and test validation

Diagnostic species-specific SNPs were detected for five species (*Acipenser naccarii*, *A. fulvescens*, *A. stellatus*, *A. sinensis*, *A. transmontanus*), in addition to a 10 bp deletion shared by *Acipenser baerii* and *A. ruthenus* (Table 1). On the basis of the diagnostic polymorphisms found, a panel of primers for species identification was designed. Primer combinations, species identified and expected band size are detailed in Table 2. Additionally, a common primer for all sturgeon species studied to be used as positive control (pc_RP1F) was designed on the basis of a conserved region shared by all individuals sequenced. No specific SNPs were identified for *Acipenser gueldenstaedtii*, *A. persicus* and *Huso huso* although nine specific polymorphisms were detected for *Acipenser sturio* (Table 1 and Fig. S3; *A. sturio* not studied further here).

A 100% identification success was obtained for the primer pairs designed to amplify *A. naccarii* (66 of 66 individuals), *A. fulvescens* (26/26), *A. sinensis* (13/13), *A. stellatus* (46/46) and *A. transmontanus* (16/16) (Table 3). For each of these five specific primer pairs, no amplification was observed for any of the nontarget species after a rigorous examination of (on average) 215 individuals: *A. naccarii* (0 amplifications of 306 individuals of nontarget species tested), *A. fulvescens* (0/178), *A. sinensis* (0/191), *A. stellatus* (0/204) and *A. transmontanus* (0/198). For the primer pair designed for *A. ruthenus*/*A. baerii* on the basis of the shared 10 bp deletion, the identification success was 67.6% (171 of 253 individuals). The efficiency was 96% for *A. ruthenus* (48/50) and 60.6% for *A. baerii* (123/203). This incomplete efficiency was unpredicted from the reference sequence data, but was confirmed a posteriori after

Table 1 Diagnostic single nucleotide polymorphisms (SNPs) for species identification (in bold)

Nucleotide position (bp) Species	62	75	83	112	126	172	204	223	353	364	400	406	427	430	439	470	532	542
Acipenser fulvescens	T	**G**	T	C	C	T	T	T	G	C	G	**G**	C	G	G	**A**	G	**T**
A. naccarii	*	T	*	*	*	*	*	*	*	**T**	*	*	*	*	*	C	*	*
A. transmontanus	*	*	*	*	*	*	**C**	*	*	*	*	*	*	*	*	C	*	G
A. sinensis	**C**	*	*	*	*	*	*	*	*	*	*	*	*	*	*	C	*	G
A. stellatus	*	*	*	*	**A**	*	*	*	*	*	*	A	*	*	*	C	*	G
A. ruthenus	*	*	*	*	*	*	*	*	*	*	*	A	**(–)$_{10}$**	–	*	C	*	G
Huso huso	*	*	*	*	*	*	*	*	*	*	*	A	*	*	*	C	*	G
A. baerii	*	*	*	*	*	*	*	*	*	*	*	A	**(–)$_{10}$**	–	*	C	*	G
A. gueldenstaedtii	*	*	*	*	*	*	*	*	*	*	*	A	*	*	*	C	*	G
A. persicus	*	*	*	*	*	*	*	*	*	*	*	A	*	*	*	C	*	G
A. sturio	*	*	**C**	**T**	*	**C**	*	**(–)$_{5}$**	*	*	**T**	A	*	**A**	**A**	**T**	**C**	G
A. schrenckii	*	*	*	*	*	*	*	*	**(–)$_{20}$**	–	*	A	*	*	*	C	*	G
Huso dauricus	*	*	*	*	*	*	*	*	**(–)$_{20}$**	–	*	A	*	*	*	C	*	G

SNPs shaded in grey were further selected for primer design. Numbers associated with each SNP refer to the nucleotide position in the multi-species alignment (Fig. S3). The symbol $(–)_n$ indicates a deletion of n bases.

© 2013 John Wiley & Sons Ltd

Table 2 Details of primers designed for sturgeon species identification including sequence, primer pairing, expected PCR product size and target species

Primer	Sequence	Used with	bp	Specificity
Reverse				
NacFul_RP1R	ACACGAGTTCCGATTTTATTATA			*Acipenser naccarii*; *A. fulvescens*
SinTran_RP1R	GGTCTGTAACACAGGTAATTATC			*A. sinensis*; *A. transmontanus*
Forward				
Nac_RP1F	GCCAATTTCAAAGTGTATTTGATT	NacFul_RP1R	227	*A. naccarii*
Ful_RP1F	CAGTGATCTGATTTTTAAGTAAAAAAG	NacFul_RP1R	521	*A. fulvescens*
Ste_RP1F	TGTCACCTTTCAAATTTGGTA	RP1_LocusA_R	479	*A. stellatus*
RutBae_RP1F	TTACATTAATTACCTGTGTTACAGATAG	RP1_LocusA_R	169	*A. ruthenus*; *A. baerii*
Sin_RP1F	GGAATTTTACAGTGATCTGATTAC	SinTran_RP1R	386	*A. sinensis*
Tran_RP1F	TGTGAAATTTAGATGATTTTAGTGC	SinTran_RP1R	246	*A. transmontanus*
SchDau_RP1F	GCCCATTTCAAAATGTATTGGGG	RP1_LocusA_R	223	*A. schreckii*; *Huso dauricus*
pc_RP1F	TGCAGGTAGTTATAAATAATTACTCTGC	RP1_LocusA_R	309	Positive control

sequencing the animals that gave unexpected results. Considering nontarget species, positive amplifications were observed in only 2.3% *A. gueldenstaedtii* (6/258) and 27.1% *A. persicus* (13/48).

An extra primer was designed to specifically amplify Kaluga sturgeon (*Huso dauricus*) and Amur sturgeon (*Acipenser schrenckii*) on the basis of a common shared 20 bp deletion (Tables 1 and 2 and Fig. S3). Although results are preliminary due to the low number of individuals tested and need further validation, a 100% positive amplification was obtained for both species (10/10 *A. schrenckii* and 10/10 *H. dauricus*), while no amplification was observed in 90 individuals of nine nontarget species (Table 3). Note that primer pc_RP1F designed as positive control for sturgeon species does not work for *H. dauricus* and *A. schrenckii* due to a deletion in the region in which the primer was designed that prevents annealing (Fig. S3).

Finally, the Vimentin primer tested on *A. baerii* and *A. gueldenstaedtii* individuals collected from Azov Sea showed that 44 of 56 *A. baerii* (78%) and 1 of 48 *A. gueldenstaedtii* (2%) resulted positive to the amplification of the expected band of 373 bp. On the contrary, *A. gueldenstaedtii* from Caspian Sea yielded positive amplification from 18 of 48 individuals, limiting the possibility to use this marker for the identification between Russian and Siberian sturgeon only to individuals from Azov Sea. With regard to the other species, no amplification was observed in *A. naccarii* (0/45), *A. fulvescens* (0/8), *A. sinensis* (0/8), *A. transmontanus* (0/8), *A. ruthenus* (0/8), *A. schrenckii* (0/8) and *H. dauricus* (0/8). However, positive amplifications were observed in *A. stellatus* (7/8), *H. huso* (8/8) and *A. persicus* (16/48).

Assays for hybrid identification

Bester (*H. huso* female × *A. ruthenus* male) and AL (*A. naccarii* female × *A. baerii* male) hybrids were identified using a combination of RP1 and mtDNA, supplemented in the case of AL with a second nuclear gene (Vimentin).

In the Bester identification (Fig. 2), the RP1 assay is able to distinguish Bester (two bands of 169 and 306 bp) from *H. huso* (one single band of 306 bp), but not from *A. ruthenus* (same two bands as Bester). However, Bester and *A. ruthenus* can be distinguished by mtDNA analysis, because Bester presents a 374 bp band and *A. ruthenus* presents an exclusive 190 bp band. The expected detection power for Bester based on the estimated allele frequencies is 80%. As *A. ruthenus* is a diploid species (Fontana *et al.* 2007), only one chromosome of this species is inherited by the hybrid; thus, the frequency of the band expected in Bester corresponds to the 'presence' allele frequency in *A. ruthenus.* In the validation test, observed results were similar to expectations and 53 of 71 (74.6%) Bester samples were correctly identified using the combined mtDNA and RP1 approach. The undetected hybrids (25.4%) presented the typical pattern of *H. huso* individuals, missing the *A. ruthenus*-specific band of 169 bp for RP1.

With regard to AL, a hybrid between Adriatic sturgeon (*A. naccarii*) females and Siberian sturgeon (*A. baerii*) males, the identification power is given by the application of RP1, totally effective in detecting the *A. naccarii* contribution, which is supplemented by the Vimentin marker for the identification of the *A. baerii* contribution (Fig. 2). As all AL commercialized worldwide have been produced in a single Italian hatchery by crossing Adriatic sturgeon females with the males of a limited stock of Siberian sturgeons, the expected power of AL detection directly depends on the percentage of correct identification on this paternal stock. The Vimentin marker amplification performed on all the 51 pure Siberian sturgeons composing the above stock yielded a clear band of the expected size from all individuals. Consequently, also the expected detection power on AL hybrids is 100%. In fact, in the AL identification (Fig. 2), the RP1 assay can distinguish

© 2013 John Wiley & Sons Ltd

Table 3 Validation tests of all primer pairs performed in 10 different sturgeon species presented as fraction of positive amplifications on the total number of individuals tested (also including animals tested in the RP1 primer development phase as first term of the denominator). Cells corresponding to target species for each primer are shaded in grey. Validation tests for the preliminary study of *Acipenser schrenckii* and *Huso dauricus* are also included. Accuracy refers to the validation success of the primer to identify the target species

Species tested Primer pairs	*Acipenser naccarii*	*A. baerii*	*A. gueldenstaedtii*	*A. persicus*	*A. ruthenus*	*A. transmontanus*	*A. fulvescens*	*A. sinensis*	*A. stellatus*	*Huso huso*	*A. schrenckii*	*H. dauricus*	Identified species	Accuracy (%)
Nac_RP1F NacFul_RP1R	66/(6 + 60)	0/(14 + 16)	0/(20 + 83)	0/(3 + 48)	0/(10 + 16)	0/(10 + 6)	0/(10 + 16)	0/(3 + 10)	0/(12 + 19)	0/(8 + 12)	0/(2 + 8)	0/(2 + 8)	*A. naccarii*	100
Ful_RP1F NacFul_RP1R	0/(6 + 10)	0/(14 + 23)	0/(20 + 30)	0/3	0/(10 + 8)	0/(10 + 6)	26/(10 + 16)	0/(3 + 10)	0/(12 + 11)	0/(8 + 4)	0/(2 + 8)	0/(2 + 8)	*A. fulvescens*	100
Ste_RP1F Stor1alS7	0/(6 + 15)	0/(14 + 8)	0/(20 + 21)	0/(3 + 24)	0/(10 + 8)	0/(10 + 6)	0/(10 + 16)	0/(3 + 10)	46/(12 + 34)	0/(8 + 12)	0/(2 + 8)	0/(2 + 8)	*A. stellatus*	100
RutBae_RP1F RP1_LocusA_R	0/(6 + 15)	123/(14 + 189)	6/(20 + 238)	13/(3 + 48)	48/(10 + 40)	0/(10 + 6)	0/(10 + 16)	0/(3 + 10)	0/(12 + 19)	0/(8 + 16)	0/(2 + 8)	0/(2 + 8)	*A. ruthenus* *A. baerii*	96 60.6
Sin_RP1F SinTran_RP1R	0/(6 + 10)	0/(14 + 23)	0/(20 + 30)	0/3	0/(10 + 8)	0/(10 + 6)	0/(10 + 16)	13/(3 + 10)	0/(12 + 11)	0/(8 + 4)	0/(2 + 8)	0/(2 + 8)	*A. sinensis*	100
Tran_RP1F SinTran_RP1R	0/(6 + 10)	0/(14 + 23)	0/(20 + 30)	0/3	0/(10 + 8)	16/(10 + 6)	0/(10 + 16)	0/(3 + 10)	0/(12 + 11)	0/(8 + 4)	0/(2 + 8)	0/(2 + 8)	*A. transmontanus*	100
SchDau_RP1F RP1_LocusA_R	0/(6 + 10)	0/(14 + 10)	0/(20 + 10)	0/3	0/(10 + 10)	0/(10 + 10)	0/(10 + 10)	0/(3 + 10)	0/(12 + 10)	0/(8 + 10)	10/(2 + 8)	10/(2 + 8)	*A. schrenckii* *H. dauricus*	100 100
pc_RP1F RP1_LocusA_R	66/(6 + 60)	203/(14 + 189)	258/(20 + 238)	51/(3 + 48)	50/(10 + 40)	16/(10 + 6)	26/(10 + 16)	13/(3 + 10)	46/(12 + 34)	38/(8 + 30)	0/(2 + 8)	0/(2 + 8)	All species tested	100

A. naccarii (two bands of 227 and 306 bp) from *A. baerii* (single 306 bp), but AL hybrids present the same two bands as *A. naccarii*. AL and *A. naccarii* can be distinguished using Vimentin, with AL presenting a single 373 bp band, while no band is amplified in *A. naccarii*. Validation was performed on 20 AL individuals. All individuals presented the two RP1 bands at 227 and 306 bp, respectively, plus the 373 bp Vimentin band, confirming their hybrid status. As a final step, the maternal *A. naccarii* contribution can be confirmed by mitochondrial control region analysis and comparison with the seven existing *A. naccarii* haplotypes.

Finally, we estimated the expected probability of correct hybrid detection for other hybrids not experimentally validated in our study, based on the positive and negative identification of target and nontarget species in the validation tests using RP1. Hybrids between species for which no specific polymorphism was detected are not distinguishable, such as the ones between *A. baerii* and *A. gueldenstaedtii* observed in the Volga River (Jenneckens *et al.* 2000) and of increasing interest for aquaculture (Ludwig 2008).

For all hybrids with *A. naccarii*, *A. fulvescens*, *A. stellatus*, *A. sinensis* or *A. transmontanus* as paternal species, a 100% identification success is expected. A 80% success is expected for those hybrids with *A. ruthenus* as paternal species. Although results are preliminary and need to be validated in more individuals, a 100% identification success is also expected for those hybrids obtained by crossing either *A. schrenckii* or *H. dauricus*.

Discussion

New DNA markers for sturgeon identification

We present a new molecular marker for fast, inexpensive and reliable identification of sturgeon species and hybrids, which represents a major step in the development of a standardized protocol for identifying illegal trade of sturgeon caviar and derivates. The method is based on simple PCR amplification of a single locus nuclear marker (RP1), the first intron of the RPS7 gene, which has traditionally been used for detecting polymorphisms at the intraspecific level due to its high mutation rate (e.g. atherinids) (Francisco *et al.* 2008). In sturgeons, the retention of species-specific SNPs can be attributable to their slowed evolutionary rate (Krieger & Fuerst 2002), predicted on the basis of particular life history traits (e.g. long generation time, large body size, ectotherm, late sexual maturation, low metabolic rate) and molecular data (i.e. limited karyotype and genetic differentiation). 'The approach here proposed is based on SNP-specific primers and is rapid, cheap and suitable for all laboratories; however, the same polymorphisms can also be

© 2013 John Wiley & Sons Ltd

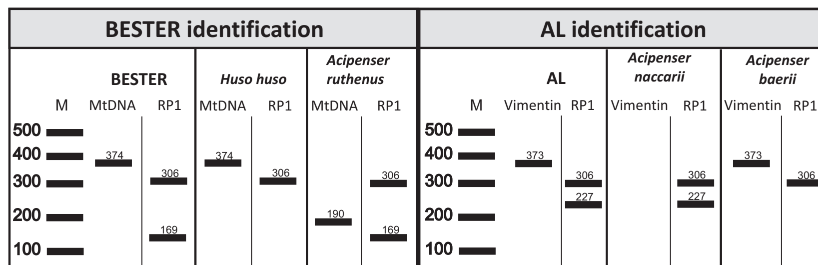


Fig. 2 Expected band patterns for Bester and AL hybrid identification using RP1 and mtDNA (Bester) and RP1 and Vimentin (AL).

genotyped following alternative approaches (e.g. pyrosequencing) that are less sensitive to DNA quality and become especially convenient in the case of high numbers of samples'.

In regards to pure species identification, besides full identification power for *Acipenser fulvescens*, *A. sinensis*, *A. transmontanus* and *A. stellatus*, we present for the first time a molecular tool to distinguish the Adriatic sturgeon with 100% accuracy. Until now, *Acipenser naccarii* was indistinguishable from the rest of species of the '*gueldenstaedtii complex*' (*Acipenser gueldenstaedtii*, *A. baerii* and *A. persicus*) using the mitochondrial cytochrome b and D-loop (Birstein *et al.* 2000). The complete reliability of the *A. naccarii* identification by both mtDNA and RP1 is guaranteed by the fact that all *A. naccarii* available worldwide descend from the stock analysed in our study (Congiu *et al.* 2011), allowing the complete description of all existing mitochondrial haplotypes. The Adriatic sturgeon is considered to be possibly extinct in nature, its persistence being strictly linked to the ex situ conservation of a single captive broodstock of 43 animals currently decimated to 13 individuals that represent the last remaining population of Adriatic sturgeon of certain wild origin (Congiu *et al.* 2011). Following the successful reproduction in captivity of the original F0 stock, all the Adriatic sturgeons reared worldwide for aquaculture purposes are F1 groups that directly descend from this limited stock. As all *A. naccarii* specimens used in the present study are either F0 individuals or F1 broodstock from the original stock, we can expect 100% reliability in correctly identifying *A. naccarii*.

The variability at the RP1 locus was only partially explored in our study. For example, several specific SNPs were available for *Acipenser sturio*, but primers were not developed because this species is not traded or reared for commercial purposes. It is reasonable to expect diagnostic SNPs for its sister species *Acipenser oxyrinchus*, which is the subject of re-introduction in Europe (Ludwig *et al.* 2008), but for which samples were not available. Limited sample availability was also an issue in the case of *Huso dauricus* and *Acipenser schrenckii*, and we only conducted a preliminary analysis. Further samples are needed for validation, but if the 100% of identification power of the RP1 primer is confirmed, it would represent the first reliable nuclear marker to distinguish these two commercially important sturgeon species from all others.

First molecular marker for hybrid identification

Besides species identification, one fundamental application of the RP1 and Vimentin markers is hybrid identification, significantly improving the identification power for Bester (from 0% to 80%) and AL (from 0% to 100%) and providing the tools to identify many other hybrids.

Bester is the most important hybrid in the world from a commercial point of view, produced by crossing Beluga (*Huso huso*) females and Sterlet (*Acipenser ruthenus*) males. The maternal species produces one of the most valuable caviar in the world known as Beluga, which is not distinguishable from Bester hybrid caviar using only the maternally inherited mtDNA, as both are produced by *H. huso* females. However, by combining mtDNA and RP1 analysis, we increased the identification success of Bester hybrids from 0% to 80%. Moreover, the possibility to analyse bester caviar obtained from unrelated females can significantly increase the probability of detecting the Sterlet paternal contribution in at least one of them. With only two or three available females, the detection power increases to 96% and 99.2%, respectively. The RP1 primer used for *A. ruthenus* detection also yields a band of the same size in 60.6% of Siberian sturgeons (*A. baerii*) as expected and in a very low fraction of *A. gueldenstaedtii/persicus* (6.15%). However, with regard to Bester caviar identification, this is not relevant because both *A. baerii* and *A. gueldenstaedtii/persicus* are not crossed with *H. huso* for hybrid caviar production due to different ploidy levels (Fain *et al.* 2013). The cooccurrence of the above band in *A. ruthenus* and *A. baerii also* makes their hybrids not identifiable. Again, the different ploidy level makes these hybrids, recently observed in the Danube River (Ludwig *et al.* 2009), not interesting for caviar production.

AL is another hybrid intensively produced for both caviar and meat production due to its high nutritional value (Vaccaro *et al.* 2005), resulting from crossing Adriatic sturgeon (*A. naccarii*) females and Siberian sturgeon (*A. baerii*) males. Our approach combining RP1, Vimentin and mtDNA is the first tool that allows the

© 2013 John Wiley & Sons Ltd

recognition of a hybrid between two of the most difficult species to identify, both included in the '*Gueldenstaedtii complex*' (Knapp *et al.* 2006; Ludwig 2008). In this case, the identification success increased from 0% to 100%. One caveat of using the Vimentin gene is its positive amplification in other sturgeon species including *H. huso*, *A. stellatus* and some populations of *A. gueldenstaedtii*. However, these species are not successfully hybridized with *A. naccarii* females for caviar production and do not interfere with AL identification. In any case, we suggest the use of Vimentin only when both mtDNA and RP1 have previously shown to be positive for *A. naccarii*.

Other newly produced hybrids involving *A. naccarii* as parental species have shown up recently in international trade and can be potentially identified using the RP1 marker and mtDNA with a 100% accuracy. This is the case of GUNA hybrids, obtained by crossing Russian sturgeon (*A. gueldenstaedtii*) females and Adriatic sturgeon (*A. naccarii*) males. Similarly, in the case of ADAM hybrids, obtained in Italy by crossing Adriatic sturgeon (*A. naccarii*) females with aquaculture-produced white sturgeon (*A. transmontanus*), a 100% identification accuracy is expected using RP1. Furthermore, any additional hybrids produced using *A. naccarii*, *A. fulvescens*, *A. stellatus*, *A. sinensis* or *A. transmontanus* as paternal species could also be potentially identified with a 100% identification power with the species-specific RP1 markers developed in our study, supplemented by mtDNA. For example, in a previous conservation study, some animals with *A. transmontanus* mitochondrial haplotypes were detected in a *A. naccarii* ex situ brood-stock reared for conservation purposes. These animals, hypothesized to be hybrids with the Adriatic sturgeon based on microsatellite information (Congiu *et al.* 2001), could have been unambiguously identified with RP1.

A molecular approach to regulate legal trade and combat fraudulent caviar trade

The development of a new molecular approach for the identification of pure sturgeon species and hybrids of the most commercially important species in Europe and North America based on diagnostic polymorphisms in the RPS7 gene ultimately offers new opportunities to regulate and control the international legal and illegal trade in caviar. Since the inclusion of all sturgeon species in the appendices of CITES, a number of conservation measures have been put in place to improve law enforcement of caviar trade. Despite these efforts, mislabelling of caviar products is common, and it is frequent to find low-value farmed caviar being sold as top-quality farmed caviar of hybrid origin (Knapp *et al.* 2006). It is thus essential for custom officials and trade authorities to have access to an accurate and reliable method of authenticating the caviar labelling system. The RP1 method allows for the identification of the source of caviar to assess the validity of the product label. Importantly, the method enables not only the identification of caviar at the species level but for the discovery of hybrid species, which have proven to be unidentifiable using other DNA-based methods. Our new molecular approach, based on the concordance between mtDNA and nuclear markers, is already suitable for law enforcement.

The expanding criminal caviar trade is a serious threat to the survival of wild sturgeon populations, diminishes the effectiveness of management and conservation programmes, undermines legal trade and threatens the sustainable production of caviar. Hence, the introduction of RP1 as a suitable marker for the identification of pure species and hybrids is not only useful to certify and corroborate the correct labelling of legal caviar by authorized producers, processors and traders, but will also have a strong dissuasive effect against fraudulent activity and could be a valuable support to conservation programmes in which the purity of the species is an essential prerequisite. Improving the control and monitoring of caviar trade will ensure its sustainability and will be the first step towards the rescue and safeguard of this group of threatened species.

Acknowledgements

We are grateful to Giacinto Giovannini and sons, Carlo Dalla Rosa for providing part of the samples, Dr. Federica Barbisan for her valuable help in the preliminary phase of this study.

References

Barbisan F, Savio C, Bertorelle G, Patarnello T, Congiu L (2009) Duplication polymorphism at MHC class II DRB1 locus in the wild boar (*Sus scrofa*). *Immunogenetics*, **61**, 145–151.

Birstein JV (2002) Sturgeon species and hybrids: can hybrids produce caviar? *Environmental Policy and Law*, **32**, 210–214.

Birstein JV, Doukakis P, DeSalle R (2000) Polyphyly of mtDNA lineages in the Russian sturgeon, *Acipenser gueldenstaedtii*: forensic and evolutionary implications. *Conservation Genetics*, **1**, 81–88.

Bronzi P, Rosenthal H, Gessner J (2011) Global sturgeon aquaculture production: an overview. *Journal of Applied Ichthyology*, **27**, 169–175.

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

Congiu L, Dupanloup I, Patarnello T *et al.* (2001) Identification of interspecific hybrids by amplification fragment length polymorphism: the case of sturgeon. *Molecular Ecology*, **10**, 2355–2359.

Congiu L, Fontana F, Patarnello T, Rossi R, Zane L (2002) The use of AFLP in sturgeon identification. *Journal of Applied Ichthyology*, **18**, 286–289.

Congiu L, Pujolar JM, Forlani A *et al.* (2011) Managing polyploidy in *ex situ* conservation genetics: the case of the critically endangered Adriatic sturgeon (*Acipenser naccarii*). *PLoS ONE*, **6**, e18249.

DeSalle R, Birstein JV (1996) PCR identification of black caviar. *Nature*, **381**, 197–198.

© 2013 John Wiley & Sons Ltd

Doukakis P, Pikitch EK, Rothschild A, DeSalle R, Amato G, Kolokotronis SO (2012) Testing the effectiveness of an international conservation agreement: marketplace forensic and CITES caviar trade regulation. *PLoS ONE*, **7**, e40907.

Fain SR, Straughan DJ, Hamlin BC, Hoesch RM, LeMay JP (2013) Forensic genetic identification of sturgeon caviars traveling in world trade. *Conservation Genetics*, **14**, 855–874.

Fontana F, Zane L, Pepe A, Congiu L (2007) Polyploidy in Acipenseriformes: cytogenetic and molecular approaches. In: *Fish Cytogenetics* (eds Pisano E, Ozouf-Costaz C, Foresti F, Kapoor BG), pp. 385–403. Science Publishers, Enfield.

Francisco SM, Congiu L, Stefanni S *et al*. (2008) Phylogenetic relationships of the North-eastern Atlantic and Mediterranean forms of Atherina (Pisces, Atherinidae). *Molecular Phylogenetics and Evolution*, **48**, 782–788.

Ishihara A, Kanari K, Saito T, Takahashi S (2010) *The State of Wildlife Trade in Japan*. TRAFFIC East Asia-Japan, Tokyo, Japan.

Jahrl J (2013) *WWF Report: Illegal Caviar Trade in Bulgaria and Romania – Results oa a Market Survey on Trade in Caviar from Sturgeon (Acipenseridae)*. WWF Austria & TRAFFIC, Vienna, Austria.

Jenneckens I, Meyer JN, 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.

Kalendar R, Lee D, Schulman AH (2009) FastPCR software for PCR primer and probe design and repeat search. *Genes, Genomes and Genomics*, **3**, 1–14.

Knapp A, Kitschke C, Von Meibom S (2006) Proceedings of the international sturgeon enforcement workshop to combat illegal trade in caviar. Prepared by TRAFFIC Europe for the European Commission, Brussels, Belgium, Contract No. N_070201/2006/433267/MAR/E2.

Kobayashi N, Tamura K, Aotsuka T (1999) PCR errors and molecular population genetics. *Biochemical Genetics*, **37**, 317–321.

Krieger J, Fuerst PA (2002) Evidence for a slowed rate of molecular evolution in the order Acipenseriformes. *Molecular Biology and Evolution*, **19**, 891–897.

Ludwig A (2008) Identification of Acipenseriformes in trade. *Journal of Applied Ichthyology*, **24**, 2–19.

Ludwig A, Arndt U, Lippold S *et al*. (2008) Tracing the first steps of American sturgeon pioneers in Europe. *BMC Ecolutionary Biology*, **8**, 221.

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.

Mugue NS, Barmintseva AE, Rastorguev SM, Mugue VN, Barmintseva VA (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.

Pikitch EK, Doukakis P, Lauck L, Chakrabarty P, Erickson L (2005) Status, trends and management of sturgeon and paddlefish fisheries. *Fish and Fisheries*, **6**, 233–265.

Rozhkovan KV, Chelomina GN, Rachelek EI (2008) Molecular identification and the features of genetic diversity in interspecific hybrids of Amur sturgeon (*Acipenser schrenckii* × *A. baerii*, *A. baerii* × *A. schrenckii*, *A. schrenckii* × *A. ruthenus*, and *A. ruthenus* × *A. schrenckii*) based on variability of multilocus RAPD markers. *Genetika*, **44**, 1453–1460.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalities and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Vaccaro AM, Bugga G, Messina CM, Santulli A, Mazzola A (2005) Fatty acid composition of a cultured sturgeon hybrid (*Acipenser naccarii* × *A. baerii*). *Food Chemistry*, **93**, 627–631.

Wei QW, Zou Y, Li P, Li L (2011) Sturgeon aquaculture in China: progress, strategies and prospects assessed on the basis of nation-wide surveys (2007–2009). *Journal of Applied Ichthyology*, **27**, 162–168.

Yarmohammadi M, Shabani A, Pourkazemi M, Noveiri B (2012) Identification of bester hybrids (female *Huso huso* and male starlet *Acipenser ruthenus*) using AFLP molecular technique. *Iranian Journal of Fisheries Sciences*, **11**, 415–423.

E.B. and L.C. had a major role in designing the experiments based on S7 markers, performing experiments and writing the paper. A.B. and N.M. designed the experiments based on Vimentin and performed experiments. J.M.P. had a relevant role in planning part of the experiments and in writing the manuscript. P.D. contributed to writing the manuscript and to frame it in the context of CITES enforcement.

Data Accessibility

DNA sequences: Adriatic sturgeon control region haplotypes, GenBank Accession nos KF771109–KF771115. RP1 sequences, GenBank Accession nos KF771068–KF771095. Locus A sequences GenBank Accessions nos KF771096–KF771108.

Sampling information for each species: Table S1 (Supporting information). Alignments for Figs S2 and S3 are available on Dryad entry doi: 10.5061/dryad.t3jd5.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Schematic view of the three groups of RP1 sequences (A, B and C) obtained after amplification with universal primers for fishes (S7RPEX1F and S7RPEX1R).

Fig. S2 Phylogenetic tree constructed with all RP1 sequences obtained in a total of three individuals of *Acipenser gueldenstaedtii*, *A. transmontanus* and *Huso huso*.

Fig. S3 Multi-species alignment of (Locus A) RP1, including one consensus sequence per species.

Table S1 Sampling details of all individuals and species used for sequencing and RP1-primer design and validation.

© 2013 John Wiley & Sons Ltd

Supporting Information

Table S1. Sampling details of all individuals and species used for sequencing and RP1-primer design and validation. U.G.O.= unknown geographical origin.

Species	Sample origin	Individuals used for sequencing	Individuals used for primer validation
A. baerii	Lena river (Russian)	6	189
	Lake Baikal (Siberia, Russian)	8	5
	Ob' river (Siberia, Russian)	-	10
A. gueldenstaedtii	VIP (Northern Italy fish farm) –*U.G.O.*	4	-
	Azov sea (Eastern Europe)	3	111
	Caspian sea	7	127
	Volga river (Europe and Russian)	6	-
A. persicus	Caspian sea	3	48
A. ruthenus	VIP (Northern Italy fish farm)- *U.G.O.*	5	-
	Danube river (Central Europe)	4	8
	Fishery Research Institute, Wollershof , (Germany)- *U.G.O.*	1	-
	U.G.O.	-	32
A. transmontanus	VIP (Northern Italy fish farm)- *U.G.O.*	2	4
	North America	8	2
A. naccarii	Po River (Italy)	4	-
	VIP (Northern Italy fish farm) – *U.G.O.*	2	45
	Buna river (Bosnia and Herzegovina)	-	10
	U.G.O.	-	5
A. fulvescens	Little wolf river (Wisconsin)	2	2
	Wolf river (Wisconsin)	2	3
	Lake Winnebago (Wisconsin)	2	-
	Embarrass river (Wisconsin)	1	3
	Bad river (Wisconsin)	1	-
	North America - *U.G.O.*	2	-
	Upper Fox river (Wisconsin)	-	8
A. sinensis	Yangtze river (China)	3	10
A. stellatus	Danube River (Romania)	8	3
	USSR	1	-
	Kazakhstan	1	-
	Iran	1	-
	Romania	1	-
	U.G.O.		23
	U.G.O.		8
A. sturio	Gironde River (France)	2	-
H. huso	VIP (Northern Italy fish farm) – *U.G.O.*	4	3
	German	1	-
	Fishery Research Institute, Wollershof , (Germany)- *U.G.O.*	1	-
	Kazakhstan	1	-
	Iran	1	-
	Romania	-	1
	U.G.O.	-	26
A. schrenckii	Yangtze River Fisheries Research Institute - China	2	8
H. dauricus	Yangtze River Fisheries Research Institute - China	2	8
	TOTALE	**102**	**702**

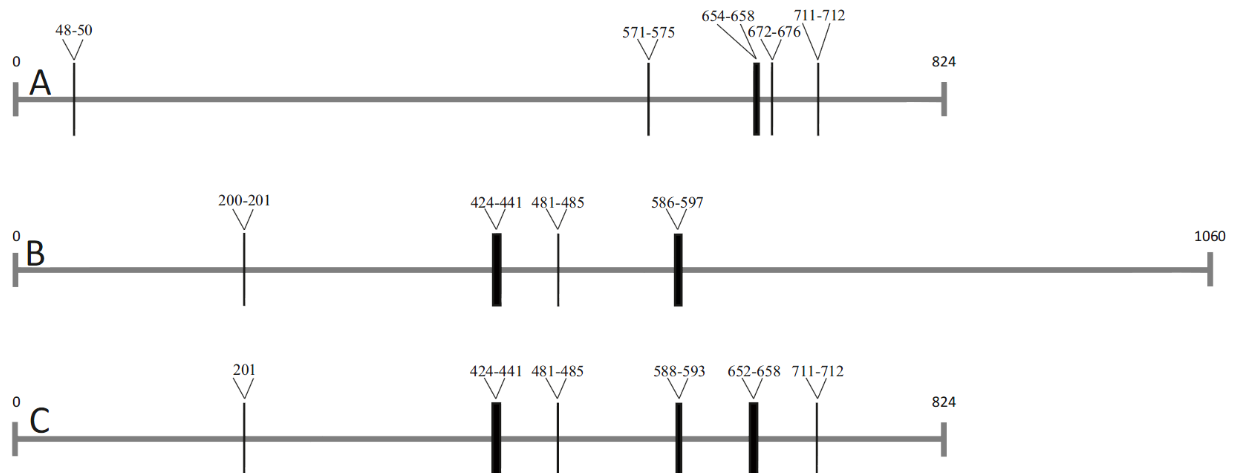


Figure S1. Schematic view of the three groups of RP1 sequences (A, B and C) obtained after amplification with universal primers for fishes (S7RPEX1F and S7RPEX1R). Black bars represent the distinctive deletions for each group. Nucleotide positions refer to the alignment of all sequences.

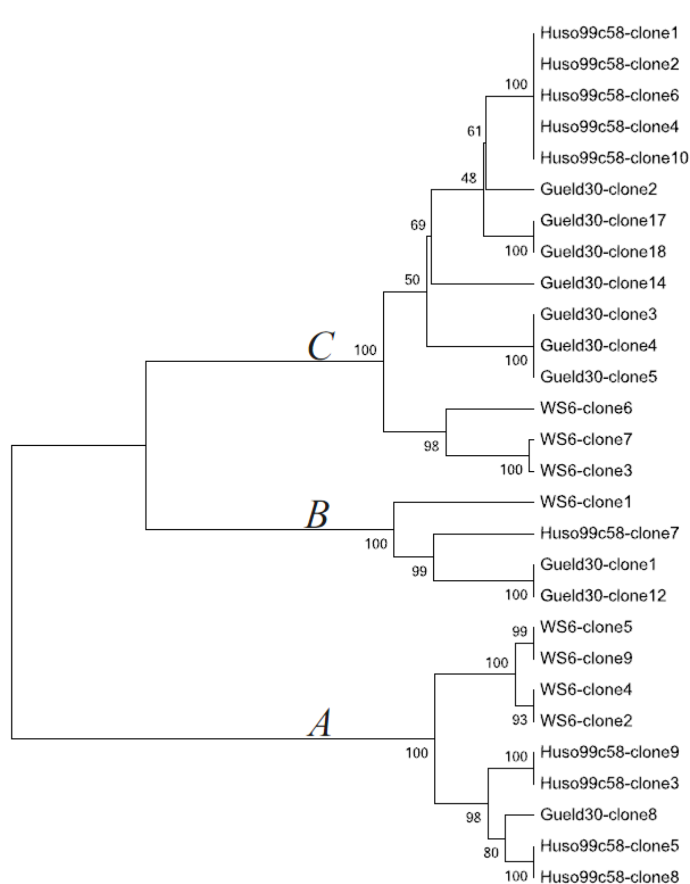


Figure S2. Phylogenetic tree constructed with all RP1 sequences obtained in a total of three individuals of *A. gueldenstaedtii*, *A. transmontanus* and *H. huso*. The UPGMA cluster analysis supports the presence of three different loci (A= blue, B= green and C= red). Boostrap was performed with 1,000 replications. Alignment available on DRYAD entry doi:10.5061/dryad.t3jd5

Figure S3. Multi-species alignment of (Locus A) RP1, including one consensus sequence per species. Regions in which primers were designed are shaded in grey. Diagnostic nucleotides at the 3' primer-end are underlined. Additional diagnostic positions for one or two species are marked in bold and underlined. Available on DRYAD entry doi:10.5061/dryad.t3jd5.

ANNEX C

The need for genetic support in restocking activities and *ex situ* conservation programs: the case of the Adriatic sturgeon in the Ticino River Park.

By E. Boscari[1] and L. Congiu[1]

[1]*Department of Biology, University of Padova, Via Ugo Bassi 58/b, 35121 Padova, Italy*

elisa.boscari@gmail.com

Summary

The Adriatic Sturgeon (*Acipenser naccarii*), endemism of the North Adriatic region, was recently reclassified by IUCN as *Critically Endangered* and *possibly extinct in the wild* since no natural spawning has been recorded in the last 20 years. Its persistence relies on restocking activities starting from a single captive broodstock collected in the 1970s in a private aquaculture plant (VIP) and currently decimated to 13 individuals. Few alternative brood-stocks of F1 animals have been retained to be used in a near future. So far, future breeders, as well as all stocks released in the wild, have been randomly chosen disregarding their genetic composition. The consequences of conducting breeding programs without a genetic input was evaluated in the Ticino River Park (TRP) in Italy. A parental allocation procedure based on microsatellite markers useful for tetraploids was used following a Band-Sharing approach. Levels of relatedness within the TRP F1 captive breeders (Stock_1) and among animals released by TRP in the past (Stock_2) were explored and compared with the stock of wild origin. The dramatic decrease in genetic diversity observed in the sample analyzed strongly suggests the need of a complete reorganization and coordination of the conservation efforts conducted on this species, paying special attention to the long term preservation of the available genetic diversity. We also identify the only potentially suitable stock of F1 animals that should be used as source of future breeders.

Introduction

Captive breeding represents the last chance of survival for species faced with imminent extinction in the wild (Russello and Amato 2004). Maintaining the offspring of wild brood-stock until maturity in the hatchery as future captive brood-stock (*ex situ* conservation) is considered a potentially useful strategy, widely applied in the last decades. In a correct strategy proposed for conservation management of threatened species, every *ex situ* population must be genetically investigated to determine if it is a representative sample of the available genetic diversity (Russello and Amato 2004; Doukakis *et al*. 2009; Theodorou and Couvet 2010; Christie *et al*. 2012) and managed with the aim of maximizing its adaptive success after release (Kozfkay *et al*. 2008) in adequate target environments.

Some past sturgeon rehabilitation programs were conducted without adequate genetic management plan, employing a limited number of breeders in restoration activities of declining populations (Hayes *et al.* 2006; Doukakis *et al*. 2009) with a potential detrimental effect on the residual genetic diversity of remnant wild stocks.

In order to evaluate the consequences of conducting breeding plans without a genetic input, we present the case of the Adriatic sturgeon in the Ticino River Park (TRP).

The tetraploid Adriatic sturgeon (*A. naccarii*) has long been considered as a highly imperiled species by the IUCN, which identified sturgeons as the most endangered group of species worldwide (Congiu *et al*. 2011; Anders *et al*. 2011; IUCN, 2011). According to a recent press released in 2010, 85% of sturgeon species are considered on the brink of extinction in the Red List of Threatened Species (http://www.iucnredlist,org); moreover, the IUCN has recently reclassified *A. naccarii* from *Vulnerable* to *Critically Endangered* and *possibly extinct in the wild* since no natural spawning has been recorded in the last 20 years (Congiu *et al*. 2011).

The last remaining population of Adriatic sturgeon of certain wild origin (more or less 50 individuals named F0) is maintained in captivity since 1977 in

the private plant "*VIP*" located in Northern Italy, thanks to owners Giacinto Giovannini and sons (Congiu *et al*. 2011). In 1988 the first captive reproduction was successful at the *VIP* plant. Since then several F1 stocks have been retained by *VIP* and purchased by different local administrations with the aim of establishing future brood-stocks composed by F1 individuals. The parental stock of wild origin is today decimated to 13 individuals due to senescence. To date, all individuals of *A. naccarii* reared in Europe directly descend from the limited parental stock genetically characterized in Congiu *et al.* (2011). However, it is possible that some F0 individuals used during the first years for crossings died before being sampled for genetic characterization, consequently, information on genetic variability of this wild stock might be incomplete (Congiu *et al.* 2011).

The aim of the present study is the characterization of genetic diversity in two stocks (one captive and one released) involved in the conservation activities of the landlocked Adriatic sturgeon population of the Ticino River, mainly conducted by the TRP. The first one (Stock_1), originally acquired from *VIP* by the Province of Piacenza, is now reared in captivity by TRP as potential brood-stock. The second (Stock_2) was purchased from *VIP* by TRP and released in the Ticino River as part of past restocking activities. The choice of animals composing the two above stocks was performed without any genetic support. In order to assess the effects of random selection of animals on stock genetic variation, we genotyped all individuals at mitochondrial DNA and at 7 microsatellite loci and compared them with the parental stock already analyzed in Congiu *et al* (2011). Relatedness among individuals was assessed when possible through parental allocation of each animal to the respective parental pair or, alternatively, analyzing the distribution of pairwise genetic distances.

The scenario here presented and discussed raises some cues to be developed for the *ex situ* conservation of *A. naccarii*, providing important information for the correct management of the captive stocks of this species, in accordance with the guidelines provided by FAO (2011).

Materials and Method

Sampling collection and laboratory procedures

Two F1 stocks of Adriatic sturgeon, obtained by unrecorded artificial reproductions using the parental stock (F0), were obtained from the Ticino River Park (Italy).

A total of 137 F1 individuals of potential future breeders (Stock_1) and a total of 116 F1 animals (Stock_1), a sub-set of the stock released in the Ticino River immediately after sampling, were analyzed.

Genomic DNA was extracted from fin clips (10-100 mg) using the DNA Easy Tissue Extraction Mini Kit (Euroclone) and stored at -4°C.

The entire mitochondrial control region was amplified using Pro1F – Phe1R primer pair following the experimental conditions described in Congiu *et al*. (2011). Following enzymatic purification with ExoSAP-IT™ (Usb), sequencing reactions were performed using an ABI Prism3730XL automatic sequencer (BMR Genomics).

All F1 individuals were genotyped at a sub-set of 7 microsatellite loci (LS-39, AnacE4, AnacC11, AnacA6, An20, AoxD234 and AoxD64) selected among the 24 loci genotyped during the parental characterization (Congiu *et al*. 2011). In cases of allocation to more than one parent pair (multi allocated individuals), the number of loci genotyped was increased up to 10 (An16, AoxD241 and AoxD161). Microsatellite loci were amplified following the conditions reported in the original references (Table 1).

Data analysis

Mitochondrial DNA sequences were aligned using ClustalW in MEGA5. A haplotype network was constructed using the statistical parsimony approach of Templeton et al. (1992) using TCS v. 1.13 (Clemetnt *et al.* 2000).

For F0 and F1 stocks, genetic diversity was measured using haplotype and nucleotide diversity estimated from number of segregating sites and mean number of pairwise differences using DnaSP software v. 5 (Librado and Rozas, 2009).

Scoring of microsatellite loci was conducted using the GeneMarker software version 1.95 (Soft Genetics LLS®). Analysis of nuclear data was

performed by a phenotypic approach based on Band-Sharing (BS) proposed by Congiu *et al*. (2011). Accordingly, BS between individuals was used as a measure of similarity. Pairwise genetic distances were calculated following the Sorenson distance and graphically represented by Multi-Dimensional Scaling (MDS) analysis
using STATISTICA version 7.1 (StatSoft). The allocation procedure was performed with the method proposed by Congiu et al. (2011) based on phenotypic multi-locus profiles.

Diversity indices

Two diversity indices, analogous to Gene Diversity index, were developed in order to record different aspects of familiar heterogeneity.
In the first index (H_f), frequencies of the families in which a given parent was involved were estimated as follows:

$$H_f = 1 - \Sigma(P_i/F)^{2},$$

where P_i is the number of families having the ith breeder as one of the parents and F is total number of families in the stock.
The H_f index represents the heterogeneity composition of a sample due to different familiar groups detected. Given a stock composed by F1 animals successfully allocated to the respective parents, H_f index evaluates how many parents were crossed to produce the observed familiar groups. In this way, the higher the number of families or the number of breeders involved in their constitution the higher the index, ranging within 0 and 1. In the case of samples composed by the same number of families, H_f decreases if some families are related (half-sibs). However, H_f does not weigh for the number of individuals composing each family, which is better described by the second index (B_i).
In $B_{i,}$ the frequencies of the offspring having a given breeder as a parent were estimated as follows:

$$B_i = 1 - \Sigma(n_i/2N)^{2},$$

where n_i is the number of offspring generated by the ith breeder and N is the number of individuals composing the stock. B_i index also ranges between 0 and 1. This value is less sensitive to the relatedness among familiar groups but is weighted for their relative abundance (evenness).
In the F0 stock, the indices were estimated assuming that each individual of wild origin was unrelated, thus representing the only component of its family.

Locus	Motif	Size (bp)	Species	Reference
LS-39	$(GTT)_{10}$	116-154	*A. fulvescens*	May et al, 1997
AnacE4	$(CA)_{20}$	326-354	*A. naccarii*	Forlani et al, 2007
AnacC11	$(TCTA)_{12}$	167-193	*A. naccarii*	Forlani et al, 2007
AnacA6	$(CA)_{15}$	289-313	*A. naccarii*	Forlani et al, 2007
An16*	$(ATCT)_{24}$	171-217	*A. naccarii*	Zane et al, 2002
An20	$(ATCT)_{10}(TG)_5$	159-213	*A. naccarii*	Zane et al, 2002
AoxD234	$(TAGA)_{17}$	215-275	*A. oxyrinchus*	Henderson-Arzapalo & King, 2002
AoxD64	$(TAGA)_{16}$	216-252	*A. oxyrinchus*	Henderson-Arzapalo & King, 2002
AoxD241*	$(TAGA)_{36}$	156-198	*A. oxyrinchus*	Henderson-Arzapalo & King, 2002
AoxD161*	$(CTAT)_{15}$	111-155	*A. oxyrinchus*	Henderson-Arzapalo & King, 2002

Table 1. List of microsatellite loci analyzed in the two F1 stocks of *A. naccarii* including repeat motif, size range, species in which microsatellites were developed and reference. *= loci genotyped for a limited number of individuals multi-allocated.

Haplotypes	1	2	3	4	5	6
F0 male (N = 20)	**2 (10%)**	**3 (15%)**	**8 (40%)**	**- (0%)**	**4 (20%)**	**3 (15%)**
F0 female (N = 23)	**- (0%)**	**8 (35%)**	**5 (22%)**	**1 (4%)**	**8 (35%)**	**1 (4%)**
F1 Stock_1 (N = 133*)	**- (0%)**	**103 (77%)**	**4 (3%)**	**26 (20%)**	**- (0%)**	**- 0 (%)**
F1 Stock_2 (N = 116)	**- (0%)**	**48 (41%)**	**62 (54%)**	**- (0%)**	**6 (5%)**	**- (0%)**

Table 2. list of number and percentage of individuals with different haplotypes for F0 sample (considering male and female separately) and the two F1 stocks for which the individual gender is unknown. Percentages are rounded. * The four individuals with haplotypes different than *A. naccarii* were not considered.

Sample	N	S	S_i	H	H_d	Θ_w	Θ_π
F0	**42**	**12**	**2**	**6**	**0.772**	**0.004**	**0.006**
F0 - present	**13**	**9**	**0**	**4**	**0.744**	**0.004**	**0.006**
F1 Stock_1	**133***	**4**	**0**	**3**	**0.364**	**0.001**	**0.001**
F1 Stock_2	**116**	**9**	**0**	**3**	**0.545**	**0.002**	**0.002**

Table 3. Diversity indices for all samples. Number of individuals (N), number of segregating site (S), number of singletons (S_i), number of haplotypes (H), haplotype diversity (H_d), nucleotide diversity estimated from number of segregating sites (Θ_w) and nucleotide diversity estimated from mean number of pairwise differencies (Θ_π) were included in the table. *Individuals with allochthonous haplotypes were discared from the analysis.

Results

Mitochondrial DNA analyses

The parental stock showed a moderate high level of genetic variation at mitochondrial control region with a total of seven haplotypes characterized (GenBank accession numbers KF771109 - KF771115) (Boscari *et al.* 2014). These seven haplotypes represent the entire variability of the mitochondrial control region of this species in captivity.

The number and percentage of wild parents showing the different haplotypes are reported in table 2.

Both F1 stocks showed a 50% decrease in haplotype diversity (from 6 to 3 haplotypes at each sample); number and percentages of individuals showing the different haplotypes are reported in Table 2. Accordingly, a drop in nucleotide diversity (H_d), more apparent in the Stock_1, was observed for both samples (Table 3).

In Stock_1, 4 out of 137 individuals presented haplotypes of other sturgeon species than *A. naccarii*. A Blast research revealed that two haplotypes belong to White sturgeon (*A. transmontanus;* ID = 98-99%) while the other two are compatible with the Russian sturgeon (*A. gueldenstaedtii;* ID = 99%). These individuals were discarded from the analysis of haplotype diversity.

Nuclear DNA and parental allocation analysis

For both stocks analyzed (N_{Stock_1} = 137 and N_{Stock_2} = 116), table 4 summarizes results of parental allocation obtained by the Band-Sharing procedure. In the case of multi-parental allocation, mitochondrial haplotypes were used to solve maternal contributions. For few ambiguous cases the analysis of three additional loci was necessary.

In Stock_1, 60 out of 137 F1 analyzed were successfully assigned to 8 families, of which 3 represented by a large number of individuals (table 4). Three cases of multi-parental assignment were observed. These results can be easily visualized in a Multidimensional Scaling analysis performed using pairwise genetic distances at the 7 loci genotyped, in which individuals cluster according to the families identified by the pedigree analysis (fig. 1).

More than half of individuals (70) were not assigned to any putative parental pair (table 4). However, the close position in the MDS (fig. 1) and

the shared haplotype 2 strongly suggest these individuals to be full sibs.

With the aim of assessing a correct degree of relatedness among not allocated individuals, the distribution of pairwise genetic distances among animals with known degree of relatedness were estimated (Congiu *et al*. 2011) and threshold values above which a given degree of relatedness can be excluded were estimated to be 0.52 for full sibs and 0.62 for half sibs. Based on these values, a single group of full sibs was identified among not allocated individuals of Stock_1. This result can easily be visualized by plotting the pairwise genetic distances of not-allocated individuals on the distribution of groups with known relationship (fig. 2).

The four individuals with haplotypes corresponding to other species than *A. naccarii* were not allocated, as expected. The analysis of RPS7 Locus A proposed by Boscari *et al.* (2013) confirmed two of these animals to be hybrids between *A. transmontanus* female *x A. naccarii* male and the other two between *A. gueldenstaedtii* female *x A. naccarii* male. These animals were excluded from the data set.

For the Stock_2 released in the Ticino River, 76 out of 116 individuals were unambiguously allocated to 9 families, of which 3 included the major part of the individuals and shared the same sire (table 3). Only one case of multi-parental assignment to two putative parent pairs was observed. According to the parental allocation procedure, the MDS analysis shows a close clustering of the 3 familiar groups (fig. 3). Also in this case, it was possible to infer the relationship between not allocated individuals (3 potential familiar groups) following the threshold value approach reported above.

Results obtained from parental allocation for the two F1 stocks in which a low number of familiar groups were detected are well summarized by the heterogeneity indices, H_f and B_i, reported in table 5. For F0 stock, the values of both indices are close to 1, which is explained by the assumption of no relationship among wild individuals. On the other hand, F1 stocks showed remarkably lower values for both indices due to the very low number of familiar groups represented and their uneven sizes.

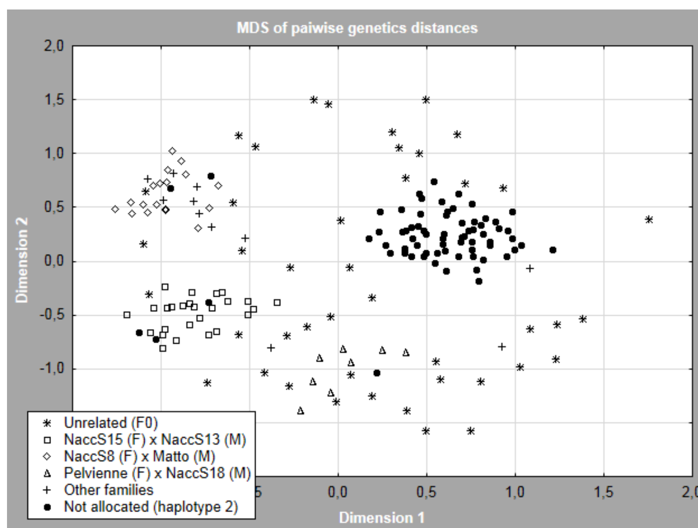


Figure 1. MDS obtained from the pairwise genetic distances among all F1 individuals of Stock_1 and the parental sample (asterisks). Triangles, squares and diamonds represent the 3 most abundant families identified while other allocated individuals are symbolized as crosses. Full black circles represent not allocated individuals, all sharing the same haplotype. Putative hybrids individuals were discarded from the analysis. Stress value = 0.23.

Index	H_f	B_i
F0 (N = 43)	**0.95**	**0.99**
F0 – present (N = 13)	**0.92**	**0.96**
F1 Stock_1 (N = 60)	**0.63**	**0.84**
F1 Stock_2 (N = 77)	**0.51**	**0.72**

Table 5. diversity indices estimates for parental sample and the two F1 stocks. For the F1 stocks, only allocated individuals were considered.

Stock_1 (F1 future breeders)			Stock_2 (F1 released)		
N° of F1 individuals	**Parental pairs**	**Haplotype**	**N° of F1 individuals**	**Parental pairs**	**Haplotype**
8	**Pelvienne (F) x NaccS18 (M)**	2	22	**NaccS8 (F) x 740 (M)**	2
18	**NaccS8 (F) x Matto (M)**	2	32	**NaccS3 (F) x 740 (M)**	3
26	**NaccS15 (F) x NaccS13 (M)**	4	17	**NaccS33 (F) x 740 (M)**	3
1	**NaccS3 (F) x Matto (M)**	3	1	**O2 (F) x Matto (M)**	5
4	**NaccS4 (F) x Matto (M)**	2	1	**NaccS26 (F) x NaccS13 (M)**	5
1	**NaccS3 (F) x NaccS18 (M)**	3	1	**O2 (F) x NaccS31 (M)**	5
1	**NaccS7 (F) x NaccS27 (M)**	3	1	**NaccS4 (F) x Matto (M)**	2
1	**NaccS7 (F) x NaccS6 (M)**	3	1	**NaccS33 (F) x NaccS29 (M)**	3
3	**Multi-allocated**	2	1	**NaccS4 (F) x Raspo (M)**	2
70	**Not allocated**	2	1	**Multi-allocated**	3
4	**Not allocated**	*	38	**Not allocated**	2/3/5

Table 4. Parental allocation to F0 stock of all F1 individuals analyzed. The number of individuals for each family and the mitochondrial haplotype are reported. * Haplotypes of other sturgeon species .

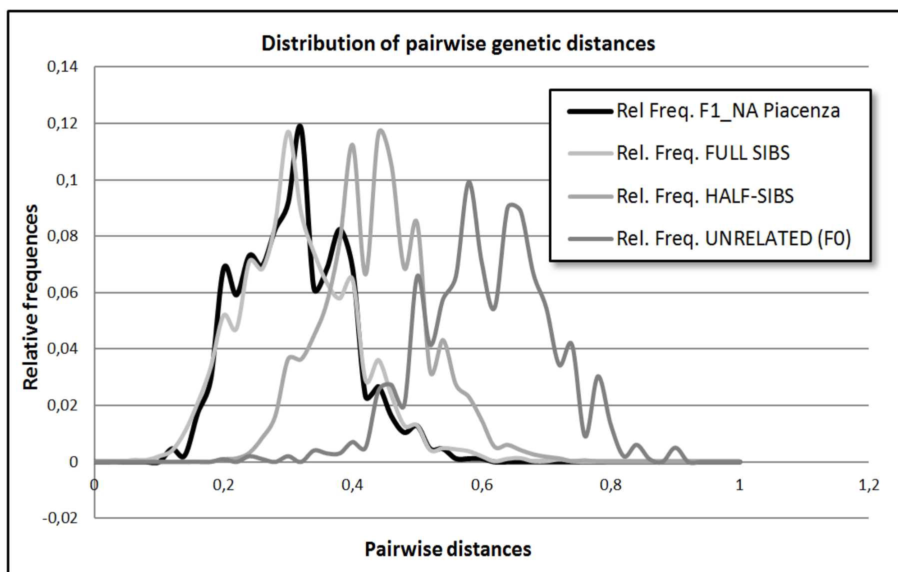


Figure 2. Distribution of observed pairwise genetic distances based on Band-Sharing information at 7 microsatellite loci for different degree of relatedness (full sibs, half sibs and unrelated) plotted with the distribution observed for not allocated individuals of Stock_1.

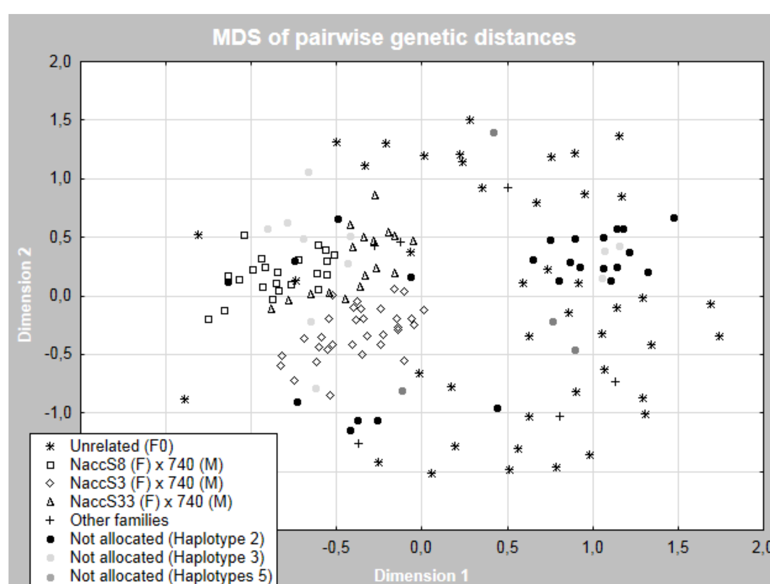


Figure 3. MDS obtained from pairwise genetic distances among all F1 individuals of Stock_2 and the parental sample (asterisks). Triangles, squares and diamonds represent the 3 most abundant families identified while other allocated individuals are symbolized as crosses. Circles, representing not allocated individuals, are filled according to the *A. naccarii* mitochondrial haplotype.
Stress value = 0.29.

Discussion

Genetic variability

The combination of mtDNA and microsatellite information resulted to be totally efficient in detecting the correct F0 parents of all the F1 progeny. In fact, in the two stocks only 4 (3%) out of 137 animals allocated resulted to be putative compatible with more than one parent pair.

Results clearly demonstrate that only a small fraction of genetic variability observed in the F0 stock was present in the F1 groups, either reared by the TRP as future captive brood-stock or directly released in the wild. This is confirmed by diversity indices, showing a drop in values of H_f and B_i (Table 5). This loss is the direct consequence of 15 years of random selection of breeders used to generate the F1 stocks. In fact, 81% of the allocated individuals of Stock_1 and 91% of the allocated individuals of Stock_2 were all assigned to just three single families, closely related in the case of Stock_2. The close relatedness among the larger families in Stock_2 has a direct effect on the heterogeneity indices, which result to be lower than in Stock_1, as a consequence of the lower number of parents involved.

There are two steps at which a strong bottleneck can occur; the first is represented by the limited number of founders used for reproductions. This usually depends on the number of available ovulating females and on the number of larvae that the hatchery do manage. In the case of the "Azienda Agricola *VIP*", in the 1990s (when the F1 stock were produced), only few animals were usually reproduced per year. A further bottleneck can occur if a limited number of the fingerlings produced in a given year is used for brood-stock foundation or for immediate restocking activities. If no attention is paid to pedigree composition there is a high probability of having stocks with high average relatedness.

In Stock_1 and Stock_2, the identification of F0 parents failed for 70 (53%) and 38 (33%) individuals, respectively. Non-allocated animals are probably offspring of F0 parents that died before the first fin clips collection performed for the F0 characterization. Since these animals may be the unique heirs of an important fraction of genetic variability, relatedness among these animals was estimated, which allowed to identify at least 4 familiar groups (1 in Stock_1 and 3 in Stock_2) to be included in any future management plan.

The alternative explanation, which is that non-allocated animals were F2, generated by crossing F1 breeders, can be reasonably excluded since, when the non-allocated animals were generated, no F1 individual was already mature. However, for the future, this possibility cannot be excluded since some F1 animals are being reproduced, even if mainly for meat and caviar production. In this view, the possibility to perform second generation allocations should be developed in a near future, possibly improving the band sharing method by including the available information on tetrasomic inheritance (Boscari *et al.* 2011).

Besides the low genetic variability, the possible genetic contamination due to the presence of alien species also represents an important issue in conservation programmes. Sturgeon species are not easy to distinguish, especially at early life stages, and the usual co-presence of different species and interspecific hybrids in sturgeon hatcheries makes accidental exchanges highly common. The detection of hybrids in Stock_1 is not surprising as these animals originally belonged to the stock reared by the Province of Piacenza, in which interspecific hybrids between *A. transmontanus* and *A. naccarii* were already detected by Congiu *et al.* (2011).

Management implication and future perspectives

Our study emphasizes how the *ex situ* management of highly imperiled species must rely on the careful characterization of residual genetic diversity and how careless conservation activities can have a strong detrimental impact. As seen in other Adriatic sturgeon captive stocks, reared in captivity as future brood-stocks, which resulted to be scarcely variable and not suitable for *ex situ* conservation (Congiu et al. 2011), the use of Stock_1 animals as breeders for production of fingerlings to be released in the wild is strongly discouraged. In fact, the random pairing of animals makes mating events between full sibs highly probable (35%) and the consequent release of the offspring not recommended.

The present study once again highlights the urgent need to design a specific rehabilitation program based on solid pedigree analysis and relatedness estimations for the Adriatic sturgeon. This program

must include the complete characterization of all F1 stocks of *A. naccarii* reared in Italian plants in order to describe the total amount of genetic variability available. For example, other adult animals, not analyzed in the present study, are reared by the TRP and there is the possibility that these animals are not related with the ones analyzed here. Therefore, genetic screening of these animals should represent a priority for the TRP with the aim of having a clear picture of the available genetic diversity in their ponds. Additionally, the *VIP* farm's owner has kept and retained in captivity several F1 individuals, obtained by crosses carried out over the last 20 years, and most of them have now reached the sexual maturity. This stock amounts to about 450 individuals and the information provided by hatchery records gives hope for an important source of genetic variability. The analysis of this stock definitely represents a priority task.

In synthesis, we strongly suggest a careful management of future breeding activities considering as priorities those individuals of wild origin that are still alive (Congiu *et al*. 2011) and optimizing the preservation of the available genetic diversity by means of a long term breeding plan that should take into account the following guidelines: maximization of the number of available familiar groups in the F1 brood-stocks, equalization of family sizes and priority to breeders with the lowest degree of relatedness. All these perspectives and advices are in line with the Ramsar Declaration on Global Sturgeon Conservation (2006) and with the FAO guidelines "*Sturgeon hatchery practices and management for release*"(2011) based on the FAO Code of Conduct for Responsible Fisheries (1995).

Acknowledgements

We are grateful to the Province of Piacenza for the financial support to the analysis of the Stock_1, Ticino River Park for providing samples of Stock_2 and University of Padova ex 60% grants for the financial support to the analysis of the Stock_2. We thank Alex Nunziata and Gianluca Lazzari for their help during the experimental activities.

References

Anders, P.J.; Drauch-Schreier, A.; Rodzen, J.; Powell, M.S.; Narum, S.; Crossman, J.A., 2011: A review of genetic evaluation tools for conservation and management of North American sturgeons: roles, benefits, and limitations. J. Appl. Ichthyol. **27** (Suppl. 2), 3-11.

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.; 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. Doi:10.1111/1755-0998.12203. *In press*

Chebanov, M.; Rosenthal, H.; Gessner, J.; Van Anrooy, R.; Doukakis, P.; Pourkazemi, M.; Williot, P., 2011: Sturgeon hatchery practices and management for release-Guidelines. FAO Fischeries and Aquaculture Technical Paper No 570. Ankara, FAO: 110 pp.

Christie, M.R.; Marine, M.L.; French, R.A.; Blouin, M.S., 2012: Genetic adaptation to captivity can occur in a single generation. PNAS **109,** 238-242.

Clement, M.; Posada, D.; Crandall, 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.; 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**, e18249.

Doukakis, P.; Babcock, E.A.; Pikitch, E.K.; Sharov, A.R.; Baimukhanov, M.; Erbulekov, S.; Bokova, Y.; Nimato, A., 2009: Management and recovery options for Ural river Beluga sturgeon. Conserv. Biol. **23**(3), 769-777.

Forlani, A.; Fontana, F.; Congiu, L., 2007: Isolation of microsatellite loci from the endemic and endangered Adriatic sturgeon *Acipenser naccarii.* Conserv. Genet. **9,** 461-463.

Hayes, B.; He, J.; Moen, T.; Bennewitz, J., 2006: Use of molecular markers to maximize diversity of founder populations for aquaculture breeding programs. Aquaculture **255,** 573-578.

Henderson-Arzapalo, A.; King, T.L., 2002: Novel microsatellite markers for Atlantic sturgeon (*Acipenser oxyrinchus*) population delineation and broodstock management. Mol. EWcol. Notes **2,** 437-439.

Kozfkay, C.C.; Campbell, M.R.; Heindel, J.A.; Baker, D.J.; Kline, P.; Powell, M.S.; Flagg, T., 2008: A genetic evaluation of relatedness for broodstock management of captive, endangered Snake River sockeye salmon, *Oncorhynchus nerka.* Conserv. Genet. **9,** 1421-1430.

Librado, nP.; Rozas, J., 2009:DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25,** 1451-1452.

Ludwig, A., 2006: A sturgeon view on conservation genetics. European J. of wildlife Res. **52,** 3-8.

Ludwig, A., (2008): Identification of Acipenseriformes species in trade. J. Appl. Ichthyol. **24,** 2-19.

May, B.; Krueger, C.C.; Kincaid, H.L., 1997: Genetic variation at microsatellite loci in sturgeon: primer sequence homology in *Acipenser* and *Scaphirhynchus.* Can. J. Fish Aquat. Sci. **5,** 1542-1547.

Philippart, J.C., 1995: Is captive breeding an effective solution for the preservation of endemic species? Biological Conserv. **72,** 281-295.

Ramsar Declaration of Global Sturgeon Conservation (2006). Proceedings of the 5th International Symposium on Sturgeon, 9-13 May 2005, Ramsar, Iran. J. Appl. Ichthyol. **22,** 5-11.

Russell, M.A.; Amato, G., 2004: *Ex situ* population management in the absence of pedigree information. Mol. Ecol. **13,** 2829-2840.

Templeton, A.R.; Crandall, K.A.; Sing, C.F., 1992: A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132**, 619–633.

Theodorou, K.; Couvet, D., 2010: Genetic management of captive population: the advantages of circular mating. Conserv. Genetic. **11,** 2289-2297.

Williams, S.E.; Hoffman E.A., 2009: Minimizing genetic adaptation in captive breeding programs: a review. Biological Conserv. **142,** 2388-2400.

Zane, L.; Patarnello, T.; Ludwig, A.; Fontana, F.; Congiu, L., 2002: Isolation and characterization of microsatellite in the Adriatic sturgeon (*Acipenser naccarii*). Mol. Ecol. Notes **2,** 586-589.

ANNEX D

Captive breeding program for the critically endangered Adriatic sturgeon (*Acipenser naccarii*).

Elisa Boscari[1], Isabelle Dupanloup[2], Riccardo Corradin[3] and Leonardo Congiu[1]

[1] Department of Biology, University of Padova, Via Ugo Bassi 58/b, 35121 Padova, Italy
[2] Institute of Ecology and Evolution, University of Bern, Bern, Switzerland
[3] Department of Statistic, University of Padova, Via Ugo Bassi 58/b, 35121 Padova, Italy

elisa.boscari@gmail.com

Summary

The Adriatic sturgeon (*Acipenser naccarii*) is a critically endangered endemism of the Adriatic region, with tetraploid genome and long life cycle. Its recovery totally depends on the management of few captive stocks that have now reached the sexual maturity and directly descend from same group of wild parents.

In the present study, an exhaustive genetic characterization was performed in order to reconstruct individual pedigrees. This information was used to plan a coordinated strategy for successful long-term species preservation through the designing of a breeding plan in which the different familiar groups are considered as breeding units, and in which the mating priorities are identified.

A two-step strategy is proposed, in which a first short-term breeding program, relying on the only 13 remnant individuals of certain wild origin, is followed by a second series of crosses among already mature F1 families, produced in captivity during last 20 years from wild parents. Simulations were run to evaluate the allele richness in F2 virtual generations under different scenarios. Simulations to assess the recommended number of individuals to breed, the costs and logistical aquaculture constraints were also performed. Results suggest that each family should be represented by 3 breeders when possible and that the selection of mating families based on genetic distance is to be preferred at least in short term programs. This work represents the first breeding plan for the Adriatic sturgeon and provides common guidelines for the coordination of the unlinked and too often short-term efforts made by different local administrations. The approach here proposed can be also transferred to the several other tetraploid sturgeon species on the brink of extinction.

Introduction

The continuously increasing field of endangered species management has seen the need to incorporate conservation genetics within recovery actions (Meffe 1986; Pertoldi *et al.* 2007). In fact, conservation management aiming at the preservation of genetic variability of wild populations can be an essential help to prevent the extinction of endangered species (Hedrick 2001; DeSalle & Amato 2004; Leus 2011; Witzenberger & Hochkirch 2011).

Given the continuous deterioration of the status of many species, now considered critically endangered, and the few reproductive populations available in nature, *in situ* conservation efforts are not sufficient to mitigate or prevent the effects of human and stochastic threats (Leus 2011). Thus, conservation activities must often rely on *ex situ* breeding programs (captive brood-stock management) and possibly on subsequent release of captive-generated animals (Philippart 1995; Russello & Amato 2004; Araki *et al.* 2007; Frankham *et al.* 2002; Leus 2011).

Brood-stock management generally involves two basic principles. The first one is the complete genetic characterization of all breeders available in captivity providing relevant information about its purity and genetic diversity. Traditionally, in most captive program the "*founder assumption*" was applied to the founder population of wild origin, by considering all the individuals as equally unrelated (Russello & Amato 2004; Congiu *et al.* 2011). Moreover, genetic characterizations are extremely useful for the reconstruction of pedigree information of captive progeny (Kozfkay *et al.* 2008; Congiu *et al.* 2011). Molecular markers are

generally at the basis of each strategy and microsatellites are the most widely used in animal conservation, allowing good performances in inferring relationships among individuals (DeSalle & Amato 2004; Russello & Amato 2004; Rodzen *et al.* 2004; Hayes *et al.* 2006; Witzenberger & Hochkirch 2011). This is crucial to identify the best strategy aimed at minimizing the loss of the residual genetic variability (Russello & Amato 2004).

The second principal involves the management of this genetic variability through a specific long-term breeding plan with the final aim of generating self-sustaining populations suitable for release (Fraser 2008; Kozfkay *et al.* 2008). With this regards, a group of captive individuals opportunely selected, named "*breeders unit*", is usually chosen as group of breeders for the production of the future generation (Goldestein *et al.* 2000). From this crucial choice depends the diversity that will be released in natural environment. Subsequently, a breeding scheme is planned determining the mating pairs, the timetable and the priorities if necessary (Myers *et al.* 2001; Russello & Amato 2004; Kozfkay *et al.* 2008).

Among animals, a recent report (March 2010) of the *International Union for Conservation of Nature* (IUCN) suggests sturgeons as the most threatened group of species (Anders *et al.* 2011; Congiu *et al.* 2011).

The tetraploid Adriatic sturgeon (*Acipenser naccarii*) is an important endemism of the Adriatic region (Bemis & Kynard 1997; Birstein & Bemis 1997), now nearly completely extinct in natural environment, in urgent need of effective recovery measures (Congiu *et al.* 2011; Boscari & Congiu 2014). The only captive stock of wild origin (F0 stock), of about 50 individuals, was collected since 1977 at the private V.I.P. plant (Orzinuovi, BS, Italy). During the last 30 years, several of them died of old age and only 13 F0 individuals are still alive. However, thanks to the successful reproduction in captivity (Bronzi *et al.* 1999), since 1988 several F1 individuals were produced.

The recovery of the Adriatic sturgeon presently depends on the management of these F1 captive stocks, all descending from the same parental F0 stock (Congiu *et al.* 2011). Now, these animals have reached the sexual maturity and are ready for reproduction.

Past genetic analysis performed on F1 stocks of *A. naccarii* retained as future brood-stocks by Ticino River Park, Province of Piacenza and Province of Treviso (Italy) revealed how the absence of a careful genetic management have caused the foundation of F1 stocks pointless as future captive brood-stocks (Congiu *et al.* 2011; Boscari & Congiu 2014).

In the present study the genetic characterization of an heterogeneous F1 stock (V.I.P. Stock), retained at the V.I.P. plant as result of 20 years of reproductions, was performed in order to reconstruct individual pedigrees.

This information was used to plan a coordinated strategy for successful long-term species preservation through the designing of a breeding plan in which the different familiar groups are considered as breeding units, and in which the mating priorities are identified.

The resulting breeding plan was also supported by simulations aimed to estimate the loss of genetic diversity under different scenarios as well as to evaluate the sustainability of the plan from economic and logistic point of view.

Materials and methods

Sample collection and DNA extraction

Individuals of *A. naccarii* analyzed in the present study belong to the larger captive stock of adults individual directly descendant from wild parents (F0) (Congiu *et al*. 2011).

All specimens (F0 and F1 animals) were provided by Azienda Agricola V.I.P., located at Orzinuovi (Brescia, Italy), thanks to the owner Giacinto Giovannini, which has retained several F1 individuals from each reproduction performed during the past 20 years in order to produce a new F1 brood-stock.

Genomic DNA of 445 F1 individuals was extracted from fin-clip (10-100 mg) using the DNA Easy Tissue Extraction Mini Kit (Euroclone) and stored at -4°C.

Mitochondrial analysis

The amplification of the mitochondrial control region (d-Loop) was performed for all individuals using the primer pair PRO1F-PHE1R, following the experimental conditions reported in Congiu *et al.* (2011). All PCR products were purified by enzymatic reaction with ExoSAP-it^{TM} (Usb) and

then sequenced at the external service BMR Genomics using ABI Prism3730XL automatic sequencer.
A multi-alignment was created using ClustalW in Mega5 and a haplotypes network was constructed using TCS v. 1.13 (Clement *et al*. 2000) following the statistical parsimony approach (Templeton *et al*. 1992).
The estimation of mitochondrial genetic diversity in comparison with the F0 stock was performed using haplotypes and nucleotide diversity estimated from number of segregating sites and from mean number of pairwise differences using DnaSP software v. 5 (Librado & Rozas 2009).

Microsatellite analysis

All F1 individuals were genotyped at 7 microsatellite loci (LS-39, AnacE4, AnacC11, AnacA6, An20, AoxD234 and AoxD64) selected among the 24 genotyped for the characterization of the parental stock (F0) (Congiu *et al*. 2011). In the case of multi-allocations (to more than one parent pair), the number of loci was increased up to 10 (An16, AoxD241 and AoxD161) until complete pedigree resolution.
Microsatellite loci were amplified following the condition reported in the original references (Table 1).
GeneMapper software version 1.95 (Soft Genetics LLS[R]) was used for the scoring.
The analysis of nuclear data was performed by a phenotypic approach based on Band-Sharing (BS), proposed by Congiu *et al.* (2011), opportunely modified as described in the following section (see next chapter).
Profiles of presence/absence (1/0) of alleles were used to estimate a Band Sharing coefficient (BS) as measure of similarity between individuals. Pairwise genetic distances were calculated (1-BS).
The estimation of genetic diversity was performed using allele richness, and two diversity indices (H_f and B_i), analogues to Gene diversity, expressly developed to record different aspects of familiar heterogeneity (Boscari & Congiu 2014). These indices were applied to individuals with known pedigree.

Pedigree reconstruction

Based on the method proposed in Congiu *et al*. (2011), a new tool for pedigree analysis named *"BreedingSturgeons"* was developed using c++ format. The program allows estimating an allocation compatibility index of each F1 individual to all possible parent pairs in the F0 stock.
Under the assumption of tetrasomic inheritance (Boscari *et al*. 2011), the compatibility index of a given F1 individual to a given parent pair was estimated as follows: at each locus, starting from the observed microsatellite profile (phenotype) all possible tetraploid genotypes are inferred for both the F1 and the two putative parents. All possible pairwise combinations of genotypes inferred for the two parents are screened to assess if alleles observed in the progeny are justified, with the constraint that the offspring is assumed to inherit two allele copies from each parent. These allele copies can be identical except from completely heterozygous parental genotypes.
The single locus compatibility index *ci*(x) is estimated as follows:

$$ci(x) = \sum[A(p1) + A(p2)]/4,$$

where, *ci*(x) is the compatibility of that animal with a parent pair at a single locus and it ranges between 0 and 1, A(p1) is the number of allele copies inheritable from parental 1 ($0 \leq A(p1) \leq 2$) and A(p2) is the ones inheritable from parental 2 ($0 \leq A(p2) \leq 2$). The same approach is extended to all loci genotyped to estimate a multi-locus compatibility index as average of single locus accuracy (*CI*(x)).
This procedure was applied to all possible parental pairs with a threshold value of compatibility opportunely fixed to allow segregation errors to 1 locus (4 alleles).
Individuals for which no parent pairs are over the threshold were considered as not allocated and they probably are sons of wild parent died before the genetic characterization of the F0 stock. Individuals allocated with a compatibility value between the threshold value and 1 were individually checked for the mitochondrial concordance with the putative mother and for the locus responsible of the imperfect compatibility. In fact, one of the loci (LS-39) is suspected to present a low frequency null allele. Individuals for which the incomplete allocation was caused by LS-39 and the mitochondrial haplotype correspond to the maternal one were considered as allocated.

Locus	Motif	Size (bp)	Species	Reference
LS-39	$(GTT)_{10}$	116-154	*A. fulvescens*	May *et al.* 1997
AnacE4	$(CA)_{20}$	326-354	*A. naccarii*	Forlani *et al.* 2007
AnacC11	$(TCTA)_{12}$	167-193	*A. naccarii*	Forlani *et al.* 2007
AnacA6	$(CA)_{15}$	289-313	*A. naccarii*	Forlani *et al.* 2007
An16*	$(ATCT)_{24}$	171-217	*A. naccarii*	Zane *et al.* 2002
An20	$(ATCT)_{10}(TG)_{5}$	159-213	*A. naccarii*	Zane *et al.* 2002
AoxD234	$(TAGA)_{17}$	215-275	*A. oxyrinchus*	Henderson-Arzapalo & King 2002
AoxD64	$(TAGA)_{16}$	216-252	*A. oxyrinchus*	Henderson-Arzapalo & King 2002
AoxD241*	$(TAGA)_{36}$	156-198	*A. oxyrinchus*	Henderson-Arzapalo & King 2002
AoxD161*	$(CTAT)_{15}$	111-155	*A. oxyrinchus*	Henderson-Arzapalo & King, 2002

Table 1. List of microsatellite loci used for characterization of the F1 V.I.P. Stock. Repeat motif, size range, species in which microsatellites were developed and reference are reported for each locus. *= supplementary loci genotyped on a limited number of individuals to increase the resolution of parental allocation.

The program also generates virtual offspring from a sample of putative parents (where males and females must be specified). This possibility was used to generate 1000 virtual F1 profiles from the F0 generation. The fraction of virtual F1 profiles allocated back only to the parental pair corresponds to the allocation power of the microsatellite panel. Since a correct reassignment rate of 92% in a virtual parental allocation was achieved with only these 7 loci, they have been selected to genotyped all F1 individuals.

In order to identify closely related individuals (full-sibs or half-sibs) among not allocated animals (for which parents were unknown), threshold values were estimated from the distribution of pairwise genetic distances between individuals with a known degree of relatedness, according to Congiu *et al.* (2011). The threshold values estimated for our panel of loci (0.58 for sibs and 0.66 for half-sibs) were applied to the distance matrix between F1 individuals and, putative groups of related individuals were identified.

Species and hybrids identification

In order to detect the presence of other species or interspecific hybrids within the *A. naccarii* stock, the analyses of the mitochondrial control region and the species-specific primers proposed by Boscari *et al.* (2014) were applied.

Breeding strategy

The strategy here proposed was not based on the choice of single breeders but the whole families were taken into account. The parent pair of each family was considered to be representative of the genetic diversity of corresponding group of fingerlings. This allowed to use a higher genetic information since all the F0 individuals are genotyped at 24 loci (Congiu *et al.* 2011). Accordingly, the genetic distance between two families is represented by the genetic distance between the two cumulative parental profiles.

In the case of families with unknown pedigree, the estimation of genetic distances from the cumulative profile of parents was not possible. In these cases, the parental profiles were inferred cumulating the phenotypes of their offspring using the 7 loci genotyped for parental allocation analysis. Genetic distances estimated on these partial profiles were included in the matrix after confirming the positive correlation (0,770 with 999 interactions performed) between distances estimated at 24 and 7 loci.

The strategy proposed for the breeding plan is similar to the minimal kinship (*mk)* strategy considering genetic distances instead of kinship which would require the allele frequencies.

Thus, the crosses were chosen following three criteria: (a) prioritize families with higher genetic value (b) exclude crosses between families with a shared parent (c) avoid repetition of already performed crosses.

The priority of the different families (named "priority families") for reproduction was estimated based on how much the parents of each family are already represented in the F1 "*breeders unit*". This

value was estimated by averaging the number of progeny that each parent have in the "*breeders unit*". In this way, the smallest families with under-represented parents are the first to be selected for reproduction.

With the aim of excluding crosses between half-sibs families, starting from the F0 profiles at 24 loci, all the possible combinations of virtual families (cumulative profiles) sharing one parent were created and the pairwise genetic distances estimated. The highest distance observed (0.35) was used as threshold value to exclude all crosses involving a shared parent. For the families whose parents were unknown the information at 7 loci was used yielding a value of 0.45.

Analogously to the *mk* strategy, in which individuals with lowest similarity are mated, in our case, each "priority family" should be crossed with the more distant family (named "mating family").

Simulations

The number of individuals that can be induced every year, the number of crosses that can be simultaneously performed, the availability of ponds for rearing the families separately until tagging, the maximum number of animals that can be hosted and several other logistical constraints are often fixed by the hatchery facilities and must be taken into account when planning a long-term breeding program. In turn, the above variables contribute to determine the costs of the project.

A R-script named "*CostsBreedingSturgeons*" (available upon request) was compiled to estimate the annual costs of the project and the number of animals that can likely be reared at the different life stages. The above logistical constraints, the biological features (*e.g.* mortality and growth rates at different life stages) and costs given by reproductions, food, tagging, manipulation, overheads *ecc.* were included with the aim of estimating the total financial support required.

Variables are strictly plant-dependent. We calibrated the different variables on the hatchery where the F1 V.I.P. Stock will be maintained in the coming years but the process can be easily applied to different conditions and sturgeon species.

A second algorithm, "*BreedingPlanSturgeons*" (also available upon request), was used to estimate the expected fraction of alleles of the parental generation that are successfully transmitted to the progeny under different breeding strategies. To this purpose the program "*BreedingSturgeons*", previously described and used for parental allocation, was applied to generate virtual F2 fingerlings starting from observed F1 parental phenotypes.

Simulation	N_i "priority family"	N_i "mating family	Selection method for the "mating family"	Tot. n of crosses performed	N_i simulated per family	Replicas
A	1	1	"Maximum Distance"	26	20	100
B	2	2	"Maximum Distance"	26	20	100
C	3	3	"Maximum Distance"	26	20	100
D	4	4	"Maximum Distance"	26	20	100

Table 2. Variables used for the four simulations performed to assess the optimal number of individuals to reproduce per family. For these simulations only the 26 families, for which the scoring at 24 microsatellite loci was available, were considered.

The optimal number of individuals per family that ensures the transmission of F1 generations alleles was estimated as follows. Based on the order of the "priority families", four different scenarios were simulated in which the "mating families" were chosen by "Maximum Distance", selecting 1, 2, 3 or 4 individuals per family, respectively (simulations A, B, C and D in Table 2). For each scenario 100 replicates were performed and the mean cumulative number of alleles successfully transmitted was counted.

Once fixed the optimal number of individuals per family, the two alternative strategies to select the "mating family" were simulated and compared: "Random" and "Maximum Distance". To this purpose, 100 replicates of F2 generations were performed and the mean cumulative number of alleles transmitted were counted.

In all the above simulations, the sex of animals of the "*breeders unit*" was not considered due to the lack of this information. Also the families for which less than 24 loci were genotyped were excluded from these simulations; thus, the families used for the simulations of the breeding plan were 26.

Results

Mitochondrial genetic variation

The F1 sample analyzed showed a decrease of haplotype diversity compared to the parental stock of wild origin (Table 3). Two of the 7 original haplotypes (Boscari *et al.* 2014) were lost (hapl-1 and hapl-4) due to the random selection of mating pairs while a third haplotype was very rare and was lost due to a poisoning event in which about half of the F1 animals died in July 2013 (Table 4).

Ten individuals were heteroplasmic with a double peak in the sequence chromatogram. Accordingly, the mother of these full sibs individuals was observed to be heteroplasmic at the same position (Boscari *et al.* 2014).

The F1 individual N0857 presented an unknown haplotype highly divergent from the *A. naccarii* ones and after Blast match was identified as white sturgeon (*A. transmontanus*) (Table 4).

Haplotype	N. indiv before poisoning	N. indiv. after poisoning
1	0	0
2	118	53
3	220	106
4	0	0
5	75	45
6	4	0
7	37	23
A. transmontanus	1	1

Table 4. Number of individuals with different haplotypes for the F1 stock before and after poisoning. The total number of individuals is higher than 445 because 10 of them presented a sequence heteroplasmy for haplotypes 5 and 7 and were counted twice.

Sample	N	S	S_i	H	H_d	Θ_w	Θ_π
F0	43*	13	3	7	0.783	0.004	0.006
F0 - present	13	9	0	4	0.744	0.004	0.006
F1 before poisoning	441**	10	0	5	0.665	0.002	0.002
F1 after poisoning	230	10	0	4	0.681	0.002	0.002

Table 3. Diversity indices for all samples. Number of individuals (N), number of segregating sites (S), number of singletons (S_i), number of haplotypes (H), haplotype diversity (H_d), nucleotide diversity estimated from number of segregating sites (Θ_w) and nucleotide diversity estimated from mean number of pairwise differences (Θ_π) were included in the table. *The heteroplasmic female was considered as two females with different haplotype. **Hybrids were discarded from the analysis.

Nuclear genetic variation and parental allocation analysis

The 7 loci selected for the characterization of this stock are the same used in Boscari & Congiu (2014) for the analysis of two populations of *A. naccari,* involved in the conservation activities in the Ticino River Park. These loci have a number of alleles per locus ranging from 8 to 14 (Table 5).

Table 6 summarizes results of parental allocation for all the F1 individuals (N = 445). Mitochondrial haplotypes used to solve some ambiguities in multiple parental allocations are also reported.

Out of 445 individuals, 382 were successfully assigned to a single parent pair and no multi allocations were observed, confirming a good power for this sub-set of loci. The remaining 63 individuals were not allocated to any of the possible F0 parent pairs; however, according to their genetic distances and to mitochondrial information, four groups of full-sibs with different mitochondrial haplotype were detected: Fam-unknownA (haplotype 3, 23 individuals), Fam-unknownB (haplotype 5, 4 individuals), Fam-unknownC (haplotype 2, 11 individuals) and Fam-unknownD (hamplotype 6, 4 individuals).

Thus, the stock is composed by 30 families with known pedigree and at least four not allocated families.

The F1 stock composition is synthesized by two diversity indices reported in Table 7, indicating a contribution of several F0 founders (16 out of 20 males and 14 out of 22 females) in the stock and a remarkable heterogeneity. As a comparison, the same indices estimated on the F1 brood-stock reared by the Province of Piacenza (Boscari &

Congiu 2014) and founded without pedigree information are reported.
Unfortunately, due to the sudden poisoning event as described above, only 233 animals survived. The abundance of each survived family is also reported in table 6. Out of 30 families with known pedigree 3 were lost. Concerning the 4 not allocated families, D was lost, A and C were reduced to 11 and 5 individuals, respectively and B was unchanged

Locus	Wild (N = 42)	Wild – present (N = 13)	F1 before poisoning (N = 441)	F1 after poisoning (N = 230)	F1-PC-Stock_1* (N = 133)
LS-39	8	8	8	8	7
AoxD64	9	8	9	9	9
AnacE4	10	7	12	12	8
AnacC11	11	7	10	10	8
AoxD234	14	12	14	14	13
An20	11	9	11	11	10
AnacA6	9	7	9	8	7
Allele richness	**10.29**	**8.29**	**10.57**	**10.29**	**8.86**

Table 5. Number of alleles per locus in the different stocks and Allele richness . *Values reported in Boscari & Congiu (2014) relative to another stock of *A. naccarii* retained as future captive brood-stock. Detected hybrids were excluded.

F1 V.I.P. Stock

N_i before poisoning	N_i after poisoning	F1 individuals	Parental pairs	Haplotype
11	2	N2882B, N0645, N1245, N2153, **N3432**, N4418, N6091, N7251, N9027, N9181, **N9860**	**Pelvienne (F) x NaccS18 (M)**	2
9	5	**N0187, N0911, N1671**, N5520, N6045, **N7547**, N7654, N7688, **N8595**	**NaccS8 (F) x Matto (M)**	2
14	14	**N0983, N1349, N3180, N4021, N4364, N6373, N6670, N6868, N7326, N7755, N7808, N8390, N9726, N9729**	**NaccS7 (F) x NaccS6 (M)**	3
30	10	N0964, N1939, N2738, N4150, N4453, N4868, **N7413, N7897**, N8029, N8101, N8361, N9493, N2490, **N5779, N0952, N1734, N6057,** N6931, **N7911**, N0303, **N0749**, N1094, **N2274**, N4263, N5268, N5468, N6096, N6842, **N9295**, N5734	**NaccS8 (F) x NaccS31 (M)**	2
10	4	N9837, N3918, **N4322, N6584, N7772, N9926,** N1400, N4143, N4395, N4504	**NaccS33 (F) x NaccS11 (M)**	3
12	8	N1167, N1174, **N1401**, N1515, **N2235, N2260**, N4109, **N5453, N5475, N6197, N8290, N9563**	**O2 (F) x Matto (M)**	5
32	23	**N0675, N1790, N5660**, N6727, **N1540, N2142**, N2718, **N5209**, N5739, **N0237, N1221, N1255**, N1267, **N1533**, **N2268, N3009, N3201**, N4821, **N5212, N5280**, N5630, **N5680, N5728, N5904**, N6219, **N6579, N6784**, N8147, **N8850, N9510, N9953**, N8234	**NaccS26 (F) x NaccS29 (M)**	5
22	15	N1780, **N1921, N4450, N6500, N9009, N9955,** N0268, **N1713, N3764, N3928, N4652**, N4850, **N4940, N5150**, N5175, **N5474, N5566**, N6059, N7549, **N7909, N9256**, N0786	**NaccS19 (F) x NaccS17 (M)**	5/7
48	27	**N0203, N0291, N0319, N0550,** N0691, **N1077,** N1426, **N1478**, N1561, N1768, **N2122, N2148**, N2538, N3055, N3920, N3924, **N4157, N4313, N4801**, N5919, **N5941, N5992, N6006, N7510, N7611,** N7670, **N7753, N8091, N8199**, N8688, N8836, **N9074**, N9521, **N23000**, N7781, **N0890, N1289, N4167**, N6024, **N0968, N1911, N2082**, N7895, N9066, N2573, N8771, N6181, N1764	**NaccS28 (F) x NaccS17 (M)**	3
38	13	N0371, **N1772**, N2037, **N2566, N2703**, N2780, N2999, N3226, N3729, **N3861**, N4139, N4975, **N5208**, N5382, N5554, N5854, N6343, **N6703, N6765**, N7264, N7336, **N7383**, N7554, **N7629,** N7726, **N8338**, N8713, **N8777**, N8797, N9298, **N9346**, N9613, N2129, **N3446**, N8939B, N3947, N7738, N8318	**NaccS16 (F) x NaccS23 (M)**	3

31	20	**N0148**, **N1827**, **N7230**, N7373, N7530A, **N8189**, N0046, **N0679**, N1157, N1386, **N1390**, **N2697**, N2742, **N3101**, N3458, **N4991**, **N5470**, **N5486**, **N6092**, **N6516**, N6753, N7099, **N7423**, **N7509**, N8158, **N8892**, **N9025**, N9230, **N9303**, **N9569**, **N9987**	**NaccS19 (F) x NaccS31 (M)**	5/7
13	2	N0965, N3813, N8203, N8815, **N9134**, N0808, **N3525**, N4356, N4398, N4670, N4920, N6509, N8312	**NaccS16 (F) x NaccS30 (M)**	3
17	12	**N2254**, N2955, **N5254**, N7250, N8549, **N9103**, **N0699**, **N1511**, **M2242**, **N2256**, **N2843**, **N5826**, **N5884**, **N6200**, **N7107**, N8129, N8854	**NaccS33 (F) x NaccS9 (M)**	3
15	8	**N1923**, **N4544**, N5945, **N6741**, N9962, **N0265**, N1563, **N3758**, **N4100**, N4214, **N5571**, N5743, **N6069**, N7798, N0807	**NaccS8 (F) x NaccS17 (M)**	2
10	2	N0461, N0473, N1131, N7374, N8939A, **N6644**, N7474, N2059, **N3264**, N6962	**NaccS12 (F) x NaccS27 (M)**	2
24	7	**N4503**, **N4591**, N7530B, N7577, **N8818**, N8507, **N0841**, **N1357**, N1984, N2542, **N3522**, N3596, **N3648**, N4145, N6064, N6249, N6414, N6592, N7114, N7556, N8146, N8187, N8426, N8576	**NaccS3 (F) x 740 (M)**	3
21	14	**N6730**, **N0625**, **N4411**, **N6511**, **N6884**, N9179, **N1934**, **N2757**, N0393, N4023, **N4282**, **N5183**, N5967, N6041, **N7112**, N7706, **N9048**, **N9255**, N9480, **N9750**, **N0934**	**NaccS8 (F) x 740 (M)**	2
3	3	**N2689**, **N4883**, **N9367**	**NaccS28 (F) x NaccS23 (M)**	3
4	1	N4650, **N5733**, N6550, N7113	**NaccS3 (F) x Matto (M)**	3
5	2	**N8113**, N0280, N3589, **N3668**, N9371	**NaccS33 (F) x 740 (M)**	3
2	-	N0648, N0549	**NaccS7 (F) x VerdeBis (M)**	3
2	1	**N9786**, N0049	**NaccS12 (F) x NaccS17 (M)**	2
2	2	**N4295**, **N4443**	**Indecisa (F) x Matto (M)**	2
1	1	**N5519**	**Indecisa (F) x NaccS11 (M)**	2
1	-	N8097	**NaccS12 (F) x Matto (M)**	2
1	1	**N6713**	**NaccS3 (F) x NaccS6 (M)**	3
1	-	N0542	**NaccS3 (F) x NaccS29 (M)**	3
1	1	**N9523**	**NaccS16 (F) x 740 (M)**	3
1	1	**N3383**	**NaccS16 (F) x NaccS29 (M)**	3
1	1	**N3447**	**NaccS24 (F) x NaccS29 (M)**	5
63	33	**N0369**, **N4038**, **N4743**, **N4964**, N7358, **N7715**, **N7730**, N9298B, N9590, N1055, **N1163**, N1716, N2882A, **N3547**, N3843, N4123, N4155, **N4511**, N4955, **N5386**, **N5407**, **N5469**, N6186, N7101, **N7543**, N7641, N9227, N0160, N0463, N0673, N0702, N0855, **N0857**, N1345, **N1673**, **N1997**, N2165, N2197, N2824, **N2854**, N3568, **N3655**, **N3847**, **N4125**, N4305, N4454, **N5981**, **N6242**, **N6769**, **N7084**, N7652, N7886, N7921, **N8201**, **N8793**, N9056, N9373, **N9507**, **N9824**, **N2062**, **N5914**, **N8123**, **N8991**	**Not allocated**	2-3-5-6

Table 6. Results of pedigree reconstruction of all F1 individuals by parental allocation to the F0 stock. The number of individuals, before and after poisoning, and mitochondrial haplotype are reported. Each individual was identified by microchip (still alive individuals are in bold). Among not allocated, individuals underlined are hybrids detected by the application of the method for species and hybrids identification as described in the following session.

Index	H_f	B_i
F0 (N = 43)	**0.95**	**0.99**
F0 – present (N = 13)	**0.92**	**0.96**
F1 (N = 382)	**0.80**	**0.94**
F1 - present (N = 200)	**0.79**	**0.96**
F1 – PC – Stock_1* (N = 60)	0.63	0.84

Table 7. Diversity indices estimated for parental samples (original stock and present stock) and F1 stocks (before and after poisoning). Data for another stock of *A. naccarii* founded in the past without pedigree information are also reported for comparison (Boscari & Congiu 2014).

Identification of species and hybrids contaminations

Four interspecific hybrids were detected. The individual N0857, with a white sturgeon mitochondrial haplotype, resulted to be a hybrid between a white sturgeon female (*A. transmontanus*) and an Adriatic sturgeon male (*A. naccarii*).

Other three individuals (N1997, N0463 and N9056) resulted hybrids despite the *A. naccarii* haplotype. One of them was always a hybrid with the *A. transmontanus* which was, in this case, the paternal species. The other two individuals resulted hybrids between an Adriatic sturgeon female (*A. naccarii*) and a Russian sturgeon male (*A. gueldenstaedtii*). These results were also confirmed by the presence at different loci of alleles never shown in the Adriatic sturgeon parental stock. These four individuals are been discarded from the breeding plan.

BREEDING PLAN

Short-term breeding scheme (F0xF0)

Considering that the 13 still alive F0 individuals are the only ones of sure wild origin, crosses among these animals were given priority. Moreover, 9 of these animals seem to have never been crossed based on the pedigree analysis.

We tried to have the contribution of each remnant F0 breeder in at least three families within the F1 generation. The already existing families representatives of each F0 parent were counted and the corresponding numbers are reported in Figure 1 as ScoreA. Doing this, we disregarded the families with less than 10 animals. The ScoreB represents the number of crosses per F0 that should be planned in order to generate at least three families per breeder in the F1 generation (ScoreA+B).

Mating pairs were selected among the more distant males and females basing on the pairwise genetic distances at 24 loci (Fig. 1) trying to minimize the number of crosses and maximize the number of parents represented. Following this strategy, 18 crosses were selected as priority (Fig. 1), among which only one was already present within the F1 stock; however, being composed by few animals it is nonetheless recommendable.

Long-term breeding scheme (F1xF1) and selection of candidate breeders

A total of 32 groups of full sibs (28 with known and 4 with unknown ancestors) of *A. naccarii* composed the "*breeders unit*". Thirty families are reared at the V.I.P. Plant and two at the Ticino River Park (Table 6).

Figure 2 reports the pairwise distance matrix among all families included in the "*breeders unit*" in which are shown crosses to be avoided as well as recommended crosses between each "priority family" and the corresponding "mating family", selected upon distance.

ScoreA	ScoreB	Score A+B	Alive F0	Sex	Nac4	Nac7	Nac14	Nac20	Nac25	Nac28	Nac5	Nac6	Nac13	Nac17	Nac23	Nac30	Nac31
0	3	3	Nac4	F	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	3	Nac7	F	0,43802	0	0	0	0	0	0	0	0	0	0	0	0
0	4	4	Nac14	F	0,59259	0,4955	0	0	0	0	0	0	0	0	0	0	0
0	3	3	Nac20	F	0,53968	0,53488	0,58621	0	0	0	0	0	0	0	0	0	0
0	3	3	Nac25	F	0,57627	0,57025	0,5	0,50794	0	0	0	0	0	0	0	0	0
1	3	4	Nac28	F	0,47619	0,48837	0,62069	0,50746	0,57143	0	0	0	0	0	0	0	0
0	3	3	Nac5	M	0,48718	0,58333	0,60748	0,488	0,60684	0,472	0	0	0	0	0	0	0
1	2	3	Nac6	M	0,54955	0,5614	0,68317	0,54622	0,65766	0,29412	0,52727	0	0	0	0	0	0
1	2	3	Nac13	M	0,55	0,5122	0,6	0,46875	0,48333	0,45313	0,54622	0,48673	0	0	0	0	0
2	2	4	Nac17	M	0,54955	0,54386	0,44554	0,63025	0,54955	0,52941	0,54545	0,59615	0,61062	0	0	0	0
1	3	4	Nac23	M	0,56667	0,46341	0,49091	0,54688	0,46667	0,5	0,47899	0,61062	0,45902	0,50442	0	0	0
0	3	3	Nac30	M	0,52542	0,50413	0,44444	0,50794	0,49153	0,49206	0,4359	0,58559	0,55	0,47748	0,43333	0	0
2	3	5	Nac31	M	0,51724	0,59664	0,60377	0,5	0,53448	0,56452	0,25217	0,59633	0,61017	0,54128	0,52542	0,53448	0

Figure 1. Distance matrix at 24 microsatellite loci between still alive F0. Only crosses between males and females are selected. Cells highlighted in grey indicate crosses selected for the short-term breeding plan on the basis of the ScoreA and the ScoreB.

Breeding plan simulations

Results of sustainability simulations, made by the script "*CostsBreedingSturgeons*", were conducted taking into account the logistical constraints of the Cassolnovo hatchery, where the breeding units will be reasonably reared, and are reported in Figure 3 as an output divided in three matrices.

The first graph (Fig. 3a) represents the matrix of families and shows how many crosses could be performed per year. Lines correspond to the years of the breeding plan, while columns represent the evolution of the relative abundance of cohorts in different years.

Normally, assuming to induce five males and five females, is reasonable to expect at least one reproduction. Each female yields about 30.000 eggs and, assuming a hatching rate of 50%, 15.000 larvae per cross are expected. In order to save pedigree information, fingerlings produced by different family combinations should be kept separated for the first two years (until tagging). Considering that only 6 ponds are available, three crosses per year are feasible (Fig. 3a) corresponding to about 45.000 larvae (column [1] in Fig. 3a).

The number of fingerlings at the second year (column [2] in Fig. 3a) is determined by the expected mortality.

At the beginning of the third year, after tagging, part of the animals should be retained in order to guarantee the long term preservation of the captive brood-stock. This number (column [3] in Fig. 3a) was estimated in order to have an average of 20 individuals per family at the 10th year of age, taking into account a mortality rate of 10% per year. The remnant animals, reported in the second graph (matrix of releases, Fig. 3b), can be released in the wild.

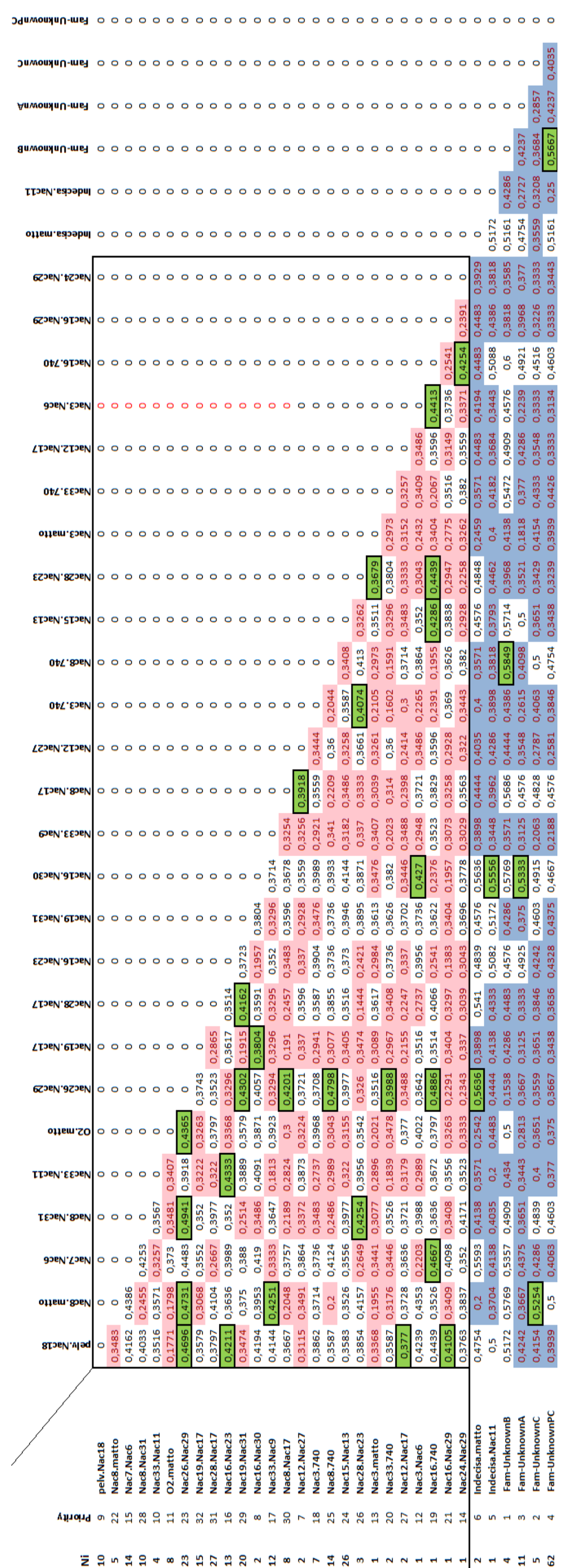


Figure 2 (right side). Pairwise distances among the families of the "*breeders unit*" estimated by comparing the cumulative profiles of the parent pairs. The black frame includes distances estimated on 24 loci and excludes the ones based on 7 loci. Red and blue cells represent crosses to be avoided between possibly related families identified with distance values lower than the thresholds of 0,35 (at 24 loci) and 0,45 (at 7 loci), respectively. Selected crosses are highlighted in green.

The third graph (matrix of costs; Fig. 3c) resumes the costs of different phases (columns) of the breeding plan and estimates the financial support required per year (last column). In this case, the cost of maintaining the adult breeders (Shared tank) was set to 0 but can be opportunely modified.

Results of the simulations of the breeding plan performed by the script "*BreedingPlanSturgeons*" are shown in Figures 4a and 4b.

Figure 4a shows that there are no significant differences between simulations C and D in which 3 and 4 individuals were involved, respectively. This result indicates that, breeding one single individual per family, the fraction of allele transmitted is significantly lower than using two individuals. The maximum inheritance efficiency is reached with a minimum of 3 individuals per family. In the long run, at the end of the breeding program, all the alleles are successfully transmitted if more than one breeder per family are mated.

Fig. 4b shows the comparison between the above simulation C, with 3 individuals mated per family, and a simulation in which each "mating family" was selected by the "Random" method rather than the "Maximum Distance". During the first stage of the breeding plan the "Maximum Distance" appear to be more efficient in transmitting alleles, while in the long run the two strategies do not differ significantly.

For all simulations the order in which the "priority families" were chosen are fixed by their priority index.

Figure 3. Example of results obtained by a run of the "*CostsBreedingSturgeons*" R-Script. The run was performed assuming 6 tanks available for the first two years of rearing and one common pond of 1200 mq for the following years.

a) Matrix of crosses

		N. crosses per year	Cohorts									
			[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Years	[1]	3	44610	0	0	0	0	0	0	0	0	0
	[2]	3	44652	5979	0	0	0	0	0	0	0	0
	[3]	3	44874	5846	141	0	0	0	0	0	0	0
	[4]	3	44682	5993	141	123	0	0	0	0	0	0
	[5]	3	45528	5966	141	124	107	0	0	0	0	0
	[6]	3	44712	5981	141	122	113	100	0	0	0	0
	[7]	3	45453	5980	141	126	112	102	92	0	0	0
	[8]	3	44838	6061	141	121	114	102	94	84	0	0
	[9]	3	44754	6043	141	132	112	101	93	82	75	0
	[10]	3	44487	5859	141	123	120	102	95	85	77	69

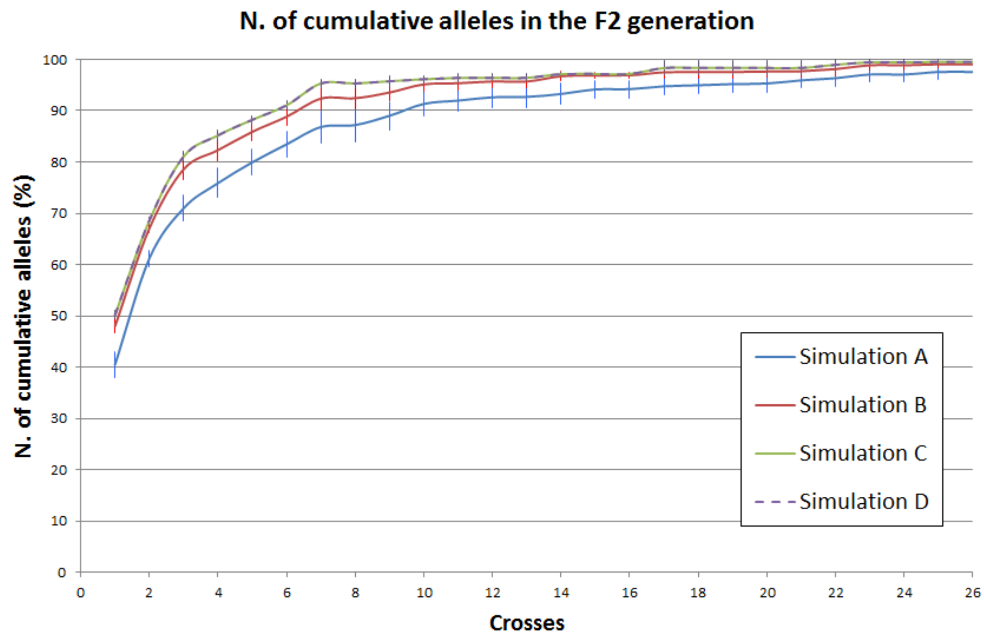
b) Matrix of releases

		N. of individuals three years old to release per year									
		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Years	[1]	0	0	0	0	0	0	0	0	0	0
	[2]	0	0	0	0	0	0	0	0	0	0
	[3]	0	0	1840	0	0	0	0	0	0	0
	[4]	0	0	0	1844	0	0	0	0	0	0
	[5]	0	0	0	0	1860	0	0	0	0	0
	[6]	0	0	0	0	0	1874	0	0	0	0
	[7]	0	0	0	0	0	0	1870	0	0	0
	[8]	0	0	0	0	0	0	0	1880	0	0
	[9]	0	0	0	0	0	0	0	0	1825	0
	[10]	0	0	0	0	0	0	0	0	0	1896

c) Matrix of costs

		Costs					
		Induction	1st year	2nd year	Chip	Shared tank	Financial support
Years	[1]	5000	18000	0	0	0	23000
	[2]	5000	18000	12000	0	0	35000
	[3]	5000	18000	12000	9905	0	44905
	[4]	5000	18000	12000	9925	0	44925
	[5]	5000	18000	12000	10005	0	45005
	[6]	5000	18000	12000	10075	0	45075
	[7]	5000	18000	12000	10055	0	45055
	[8]	5000	18000	12000	10105	0	45105
	[9]	5000	18000	12000	9830	0	44830
	[10]	5000	18000	12000	10185	0	45185

a)



b)

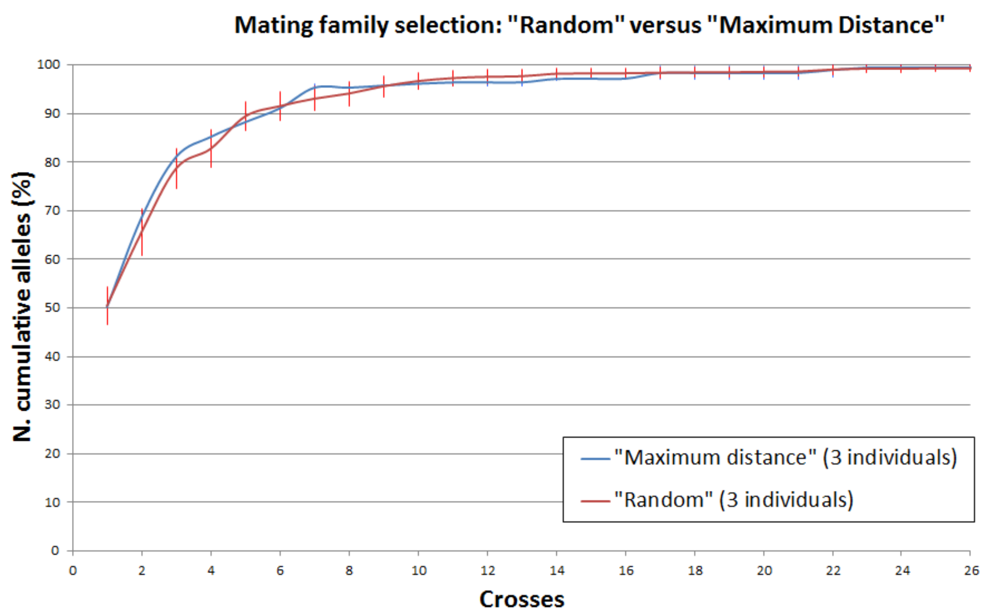


Figure 4. Results of the simulations performed by the R-script "*BreedingPlanSturgeons*" to test: a) the optimal number of breeders per family and, b) the strategy for the choice of the "mating family". The cumulative percentages of transmitted alleles are averaged on 100 replicates and reported with the corresponding standard deviations.

Discussion

The breeding program for the Adriatic sturgeon here proposed met the two principals generally required in captive breeding: the complete genetic characterization for pedigree reconstruction and the planning of a breeding scheme on the basis of an opportunely constituted "*breeders unit*".

This is the first study providing detailed information on the real status of this endangered species in Italy through the characterization of the major source of F1 animals which have reached the sexual maturity and are ready for reproductions in captivity. These information allowed, for the first time, the planning of a specific long-term breeding program paying attention to all aspects related to aquaculture constraints in order to reach its expected complete feasibility.

Results obtained by parentage analysis based on the modified Band-Sharing approach are comparable with other studies with microsatellite loci (Congiu *et al.* 2011; Boscari & Congiu 2014) with the improvement that all ambiguities were completely solved, in fact, no animals were multi-allocated. This also confirmed that the 7 loci selected for the analysis present a good resolution power for parental allocation procedures.

Results of the parental allocation confirmed that the V.I.P. Stock is the major source of variability for this species. Several families were identified and a great part of the F0 parents are represented within the stock. In regards to the relationships among animals, the probability to select for reproduction two related individuals by chance is very low (5,6%), compared to the 35% estimated for the *A. naccarii* Stock_1, reared by the Ticino River Park (Boscari & Congiu 2014). This is well described by the lower values of diversity indices H_f and B_i indicating a high degree of genetic homogeneity and a low number of families in the Stock_1. However, these animals, that taken alone are not recommendable as ex situ brood-stock, include families absent or under-represented in the V.I.P. Stock. Anticipating the desirable collaboration between subjects involved in the Adriatic sturgeon conservation through the exchange of animals between plants, in the present study, all the interesting animals were included regardless of their location.

Another important aspect emerged by the characterization of the F1 stock is the detection of interspecific hybrids with *A. transmontanus* or *A. gueldenstaedtii*, thus highlighting the importance of a careful assessment of species purity before any conservation action.

The recruitment of new breeders from the wild is not possible for the Adriatic sturgeon but, given the presence of some still alive individuals of wild origin, we propose a priority short-term breeding program to be attempted with these F0 animals. Most of them were never mated and a their successful reproduction would represent a relevant contribution to the future diversity of the captive stock of this species. However, the absence of reproduction events in captivity of these animals might be due to a low reproductive value. This would probably preclude the possibility to successfully reproduce these individuals also in the future. The fact that most survived animals were never reproduced also suggests that longevity of individuals used for artificial reproductions might have been negatively impacted either by physiological costs for reproduction itself or by hormonal induction.

Considering the low chances of successful reproduction of these animals and their limited number, most efforts were devoted to developing a long-term breeding strategy involving families of F1 generation. Dealing with sturgeons, in fact, the difficulties to have a given animal available for reproduction when needed makes a breeding program based on single individuals unfeasible. For this reason, the program here proposed considers the entire families as breeding entities, assuming that at least some of the individuals of each family are ready for reproduction every year.

The program starts with the exclusion of combinations between families possibly sharing one parent based on a threshold value of genetic distance. Doing this, also family combinations with no shared parents but with a genetic distance lower than 0,35 are excluded. These probably are combinations of families whose F0 parents, even if different, were more similar than average. This allow to minimize possible errors due to the "*founder assumption*" in which wild founders of a captive population are assumed to be equally unrelated.

In the establishment of F2 brood-stock we decided to plan the retention of 20 adult individuals for each new family produced. Taking into account the

rates of mortality during the different years and considering that in our model we plan to produce 3 different family combinations per year (as imposed by logistical constraints of the hatchery we considered in the model), this means that at least 141 individuals should be retained (Fig. 3a) at the beginning of the third year of age, immediately after tagging. The rest of the animals can be used for restocking purposes (Fig. 3b).

The choice of keeping separated for two years the animals produced by different family combinations, until the moment of tagging, will allow to track the pedigree of every single animal. In turn, it provides useful information for monitoring the releasing program or for managing the following generation of captive brood-stock. In the model presented in Figure 3, one individual per family is crossed and, the relatively low number of releasable animals per year is given by the initial number of eggs and by mortality rates acting at different developmental stages. A simple way to increase this number, keeping separated different family combinations, is to increase the number of breeders of the same family, thus allowing to pool their fingerlings. This would also increase the efficiency of allele transmission guaranteeing that all the F0 alleles of each family are inherited by F2 generation as confirmed by simulations. The suggested minimum number of 3 individuals per family is often hard to be reached for several reasons among which the small size of some families or the discontinuous availability to reproduction of breeders (mainly true for females). In this latter case, if the "priority family" is not available, one possibility is to postpone the cross to the following reproductive seasons and to slightly modify the priority order. If, on the other hand, the unavailable individuals belong to the "mating family", one possibility is to save the priority order and to select an alternative "mating family" choosing the next more distant.

Given the high economical and logistical costs of maintaining each F2 family in separated ponds, any expedient to overcome this problem would represent an improvement. One possibility could be to pool all the animals obtained from different family combinations of the same year and, if needed, to perform a parental analyses in the future. The parental allocation would be simplified by the low number of crosses performed in the same year (traceable by the tag) and consequently by the restricted number of possible parents.

The two alternative approaches for selection of the "mating family" ("Random" versus "Maximum Distance") were tested through simulations in order to asses which one allows a better representation of the parental generation. Results show that on the long run the two approaches tend to give similar results. On the other hand, after a limited number of crosses, the percentage of alleles successfully transmitted in the "Maximum Distance" approach tends to be higher. Moreover, the "Random" selection approach clearly show a higher standard deviation on 100 replicates confirming its lower reliability in the case of short-term programs . This is expected because under "Maximum Distance", families with low allele sharing are preferentially combined and the variety of transmittable alleles tends to be higher than under "Random" approach.

In our case, with a low number of crosses feasible per year and with the uncertainty of a long term financial support, the distance-based approach is to be preferred.

Four families with unknown ancestor were also included in the "*breeders unit*" to ensure the genetic contribution of their parents that were died before the first F0 characterization. These not allocated animals were genotyped at 7 loci only, which was enough to infer the full sib relationships but not for a reliable estimation of the generic distances from the other families. For this reason, these families were given a priority order and one crosses was planned, but were excluded from the families selectable as mating pair based on genetic distances. Consequently, for these families only one cross was planned.

Conclusion

The present work represents the first breeding plan for the Adriatic sturgeon based on a detailed genetic characterization and can be considered as a reference guideline for all conservation actions based on controlled reproductions of this species and possibly of other tetraploid sturgeons.

Presently, several local administrations are active in the attempt of recovering the Adriatic sturgeon in its historical natural range. Unfortunately, these actions are often represented by short-term programs and are completely disjointed. Providing

a common program that can connect the different actions, keeping save their identity, represents the first step towards the only possible responsible way represented by the concerted management.
The plan here proposed is in line with the general consideration of the captive breeding as a temporary solution with the final aim to retain in captivity only small brood-stock, representatives of the genetic variability of the species. Hence, the future perspectives are the optimization of the release practices to reach self-sustaining populations in semi-wild and thus wild environment. Additionally, for what concern the Adriatic sturgeon, important actions at environmental level are urgently needed such as the restoration of river connectivity and of spawning sites.
The above described breeding program was planned without taking into account the several unpredictable troubles that can arise when dealing with sturgeons, such as the temporary unavailability to reproduction. For this reason we suggest to consider this breeding plan as a main guideline to be tentatively followed allowing fine tuning along the way.

Acknowledgement

We thank Giacinto Giovannini and sons, owners of the Azienda Agricola V.I.P., for providing all F1 samples. We are grateful to the "Parco Regionale Veneto del Delta del Po" for the financial support to the analysis of the V.I.P. Stock. Finally, a special thanks to Giorgio Bertorelle, Andrea Benazzo and Paolo Bronzi for their active contribution.
We thank Lara Bellini for its help during the experimental activities.

Bibliography

Anders PJ, Drauch-Schreier A, Rodzen J, Powell MS, Narum S, Crossman JA (2011). A review of genetic evaluation tools for conservation and management of North American sturgeons: roles, benefits, and limitations. *Journal of Applied Ichthyology*, 27(Suppl.2): 3-11.

Araki H, Cooper B and Blouin MS (2007). Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. Science, 318: 100-103.

Bemis W and Kynard B (1997). Sturgeon rivers: an introduction to acipenseriforme biogeography and life history. Environmental Biology and Fishes, 48: 167-183.

Birstein VJ and Bemis WE (1997). How many species are there within the genus Acipenser? In: Birstein VJ, Waldman JR, Bemis WE (Eds.), Sturgeon Biodiversity and Conservation. Kluwer Academic Publishers, Dordrecht, pp. 157-163.

Boscari E, Barbisan F, Congiu L (2011). Inheritance pattern of microsatellite loci in the polyploidy Adriatic sturgeon (Acipenser naccarii). Aquaculture, 321: 223-229.

Boscari E and Congiu L (2014). The need for genetic support in restocking activities and *ex situ* conservation programs: the case of the Adriatic sturgeon in the Ticino River Park. *Proceedings of Jourbal of Applied Ichthyology. Accepted*

Boscari E, Barmintseva A, Pujolar JM, 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, doi: 10.1111/1755-0998.12203, in press.

Bronzi P, Rosenthal H, Arlati G, Williot P (1999). A brief overview on the status and prospects of sturgeon farming in Western and Central Europe. Journal of Applied Ichthyology, 15: 224-227.

Clement M, Posada D, Candrall KA (2000). TCS: a computer program to estimate gene genealogies. Molecula Ecology, 9: 1657-1659.

Congiu L, Pujolar JM, 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: e18249.

DeSalle R and Amato G (2004). The expansion of conservation genetics. Nature Reviews Genetics, 5(9): 702-712.

Forlani A, Fontana F, Congiu L (2007). Isolation of microsatellite loci from the endemic and endangered sturgeon *Acipenser naccarii*. Conservation Genetics, 9: 461-463.

Frankham R, Ballou JD and Briscoe DA (2002). Introduction to conservation genetics. Cambridge university press, Cambridge, UK.

Fraser DJ (2008). How well can captive breeding programs conserve biodiversity? A review of salmonids. Evolutionary Applications, 1: 535-586.

Goldstein PZ, DeSalle R, Amato G, Vogler AP (2000). Conservation genetics at the species boundary. Conservation Biology, 14(1): 120-131.

Hayes B, He J, Moen T, Bennewitz J (2006). Use of molecular markers to maximize diversity of founder populations for aquaculture breeding programs. Aquaculture, 255:573-578.

Hedrick PW (2001). Conservation genetics:where are we now? Trends in ecology and evolution, 16(11): 629-636.

Henderson-Arzapalo A and King TL (2002). Novel microsatellite markers for Atlantic sturgeon (*Acipenser oxyrinchus*) population delineation and broodstock management. Molecolar Ecology Notes, 2: 437-439.

Kozfkay CC, Campbell MR, Heindel JA, Baker DJ, Kline P, Powell MS, Flagg T (2008). A genetic evaluation of relatedness for broodstock management of captive, endangered Snake River sockeye salmon, Oncorhynchus nerka. Conservation Genetics, 9: 1421-1430.

Leus K (2011). Captive breeding and conservation. Zoology in the Middle East, 54: 151-158.

Librado NP and Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25: 1451-1452.

May B, Krueger CC, Kincaid HL (1997). Genetic variation at microsatellite loci in sturgeon: primer sequence homology in *Acipenser* and *Scaphirynchus.* Canadian Journal of Fisheries and Aquatic Science, 5: 1542-1547.

Meffe GK (1986). Conservation genetics and the management of endangered fishes. *Fisheries,* 11: 14-23.

Myers JM, Heggelund PO, Hudson G, Iwamoto RN (2001). Genetics and broodstock management of coho salmon. Aquaculture, 197: 43-62.

Pertoldi C, Bijlsma R, Loeschcke V (2007). Conservation genetics in a globally changing environments: present problems, paradoxes and future challenges. Biodiversity and conservation, 16: 4147-4163.

Philippart JC (1995). Is captive breeding an effective solution for the preservation of endemic species? Biological Conservation, 72: 281-295.

Rodzen JA, Famula TR, May B (2004). Estimation of parentage and relatedness in the polyploid white sturgeon (Acipenser trasnsmontanus) using a dominant marker approach for duplicated microsatellite loci. Aquaculture, 232: 165-182.

Russello MA and Amato G (2004). Ex situ population management in the absence of pedigree information. Molecular ecology, 13: 2829-2840.

Templeton AR, Crandall KA, Sing CF (1992). A cladistics analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics, 132: 619-633.

Witzenberger KA and Hochkirch A (2011). Ex situ conservation genetics: a review of molecular studies on the genetic consequencies of captive breeding programmes for endangered animal species. Biodiversity and Conservation, 20: 1843-1861.

Zane L, Patarnello T, Ludwig A, Fontana F, Congiu L (2002). Isolation and characterization of microsatellite in the Adriatic sturgeon (*Acipenser naccarii*). Molecular Ecology Notes, 2: 586-58

APPENDIX I

Vidotto et al. BMC Genomics 2013, 14:407
http://www.biomedcentral.com/1471-2164/14/407

BMC Genomics

RESEARCH ARTICLE Open Access

Transcriptome sequencing and *de novo* annotation of the critically endangered Adriatic sturgeon

Michele Vidotto[1*], Alessandro Grapputo[1], Elisa Boscari[1], Federica Barbisan[1], Alessandro Coppe[1], Gilberto Grandi[2], Abhishek Kumar[1,3] and Leonardo Congiu[1]

Abstract

Background: Sturgeons are a group of Condrostean fish with very high evolutionary, economical and conservation interest. The eggs of these living fossils represent one of the most high prized foods of animal origin. The intense fishing pressure on wild stocks to harvest caviar has caused in the last decades a dramatic decline of their distribution and abundance leading the International Union for Conservation of Nature to list them as the more endangered group of species. As a direct consequence, world-wide efforts have been made to develop sturgeon aquaculture programmes for caviar production. In this context, the characterization of the genes involved in sex determination could provide relevant information for the selective farming of the more profitable females.

Results: The 454 sequencing of two cDNA libraries from the gonads and brain of one male and one female full-sib *A. naccarii,* yielded 182,066 and 167,776 reads respectively, which, after strict quality control, were iterative assembled into more than 55,000 high quality ESTs. The average per-base coverage reached by assembling the two libraries was 4X. The multi-step annotation process resulted in 16% successfully annotated sequences with GO terms. We screened the transcriptome for 32 sex-related genes and highlighted 7 genes that are potentially specifically expressed, 5 in male and 2 in females, at the first life stage at which sex is histologically identifiable. In addition we identified 21,791 putative EST-linked SNPs and 5,295 SSRs.

Conclusions: This study represents the first large massive release of sturgeon transcriptome information that we organized into the public database *AnaccariiBase*, which is freely available at http://compgen.bio.unipd.it/anaccariibase/. This transcriptomic data represents an important source of information for further studies on sturgeon species. The hundreds of putative EST-linked molecular makers discovered in this study will be invaluable for sturgeon reintroduction and breeding programs.

Background

Sturgeons (order: Acipenseriformes, infraclass: Chondrostei) are a very ancient fish group distributed in the Palearctic hemisphere with about 25 species, most of which are considered to be on the brink of extinction [1]. Their conservation importance has led to the inclusion of all the species in the red list of the International Union for the Conservation of Nature (IUCN) and to commercial restrictions under the Convention for International Trading of Endangered Species (CITES). These fish are also interesting from a biological standpoint, presenting peculiarities that make the characterization of their transcriptome worthwhile. Often referred to as living fossils, sturgeons very ancient separation from teleosts occurred over 250 Mya [2], placing them in a key phylogenetic position for evolutionary studies on vertebrates. A second aspect of interest is related to the ploidy of sturgeons. Different species are characterized by different degrees of ploidy which are the result of multiple and independent duplication events [3-5].

Sturgeon species can be divided into two main groups based on their number of chromosomes: the first having approximately 120 and the second approximately 240 chromosomes. The level of ploidy to be ascribed to each

* Correspondence: michele.vidotto@gmail.com
[1]Department of Biology, University of Padova, Via G. Colombo 3 I-35131, Padova, Italy
Full list of author information is available at the end of the article

BioMed Central

© 2013 Vidotto et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

chromosome number is still being debated. Some authors consider species of the two groups to be diploid and tetraploid respectively [6] while others attribute the tetra/octoploid condition to the same groups [7]. In this paper we characterised the transcriptome of the Adriatic sturgeon (*Acipenser naccarii*) which belong to the 240 chromosomes group [8]. A detailed transcriptome investigation of a polyploid sturgeon species will become crucial to assess the functional reduction of ploidy once the genome will be sequenced or once either the trancriptome of a 120 chromosomes species will be available. Following a polyploidization event, in fact, the redundant genetic material experiences a functional reduction which can be monitored by analysing the transcribed part of the genome [9].

The high economic value of these animals adds another aspect to the characterisation of the transcriptome of a sturgeon species. Sturgeon eggs, sold as caviar, are one of the most valuable products in international food trade [10] and their very high monetary value is the main cause of the extremely-endangered status of most sturgeon species. In response to the rapid decline of natural populations, aquaculture production of caviar is rapidly increasing. One of the main problems for aquaculture caviar producers is that 50% of the animals are profitless males which need to be discarded from production as quickly as possible to minimise expenditure and maximise space. However, sex discrimination in sturgeon farming for caviar production can only be performed by ultrasound analysis after 4 or 5 years. The rearing of males can, thus, represent up to 30% of total farming costs [11]. A genetic identification of the sexes at an early life stage based on PCR techniques could, therefore, contribute to lowering the costs of caviar production in aquaculture and have knock-on effects in both farming and conservation. Aquaculture activity would significantly benefit from this possibility and poaching on natural populations would consequently be reduced.

There are good indications that sex is genetically determined in sturgeon [11,12] however, genomic screening performed with the aim of identifying a sex marker has not, as yet, yielded satisfactory results [11,13]. Knowledge of which genes are involved in sex differentiation in sturgeons is limited and analyses at the transcriptome level of the expressed genes at the first stage at which sex can be histologically determined could contribute to expanding the knowledge base.

In a March 2010 press release, IUCN identified sturgeons as the world's most endangered group of animals with 85% of the species being at risk of extinction. For the very low numbers of wild breeders, future restoration efforts must rely on *ex situ* conservation strategies through the setup of long-term breeding programs. The availability of a high number of genetic markers to guarantee adequate genetic support to releasing activities, through parental allocation and traceability of the hatchery of origin, become important in this context. Moreover, the availability of EST-linked markers yielded by transcriptome characterisation may provide a suitable tool for the identification of footprints of selective pressures in the released stocks, due to natural or anthropogenic stress.

The present paper reports the first characterisation of the Adriatic sturgeon transcriptome obtained by 454 titanium sequencing. We provide the results of the comparison of one male and one female library. Finally we characterized microsatellite and SNPs loci to be employed for conservation purposes.

The results of this characterization have been organized in a public database which represents to our knowledge the first large amount of information of a sturgeon transcriptome.

Results and discussion

Cleaning and assembly

Two one-quarter picotiter plates of a 454 FLX sequencing run generated 154,882 and 176,703 reads from the *A. naccarii* male (cDNA3) and female (cDNA4) respectively. FastQC [14] overview of raw sequences showed that mean per-base quality remains above 24 for the first 350 bp and, thereafter, drops rapidly towards the end of the reads (data not shown). The cleaning process was passed by 99% of the reads from each library, yielding a total of 110.25 Mbp of cleaned sequences with an average length of 336 bp and mean Phred quality of 28. The main features of the sequences that passed the preprocessing step are summarized in Table 1 while their length distribution is plotted in Additional file 1. The mean GC content calculated for the whole dataset was 37.92%. GC content across sequence length follows a normal distribution thus discarding the hypothesis that systematic bias was present (data not shown). As expected, more than 50% of the total sequences (121,467 sequences) were 400 bp or longer.

Table 1 Statistics of reads preprocessing for *A. naccarii* libraries

Category	Male (cDNA3)	Female (cDNA4)	Total
Total number of raw reads	154,882	176,703	331,585
Total number of cleaned reads	153,215	175,198	328,413
Percentage of cleaned reads	99.00	99.00	99.00
Median length (bp)	376	354	365
Average GC content in percentage	37.77	38.06	37.92
Total length of cleaned reads (Mb)	52.91	57.34	110.25
Average phred quality of cleaned reads	28	28	28

The first round of MIRA assembled 256,738 reads (77.43% of the total cleaned reads) into 44,232 contigs and 16,593 singletons. The first assembly resulted in 27.62 Mbp of total consensus, composed of 60,825 sequences with an average length of 454.14 bp, average Phred quality of 39, a mean GC content of 38.47% and an average coverage of 4.22 reads. More details about the generated contigs and singletons are reported in Table 2. In the second round MIRA reassembled 6,242 contigs (14%) and 3,504 singletons (21%) from the previous assembly into 4,203 metacontigs, with an average coverage of 2.32 sequence/metacontig (Table 3).

Finally the two assembly runs were merged giving a total of 55,282 sequences, 42,193 contigs plus metacontigs (21.87 Mbp) and 13,089 singletons (3.91 Mbp). This resulted in a 9.11% sequence reduction compared to the first assembly as clearly illustrated by Figure 1. Overall, the sequences of this final dataset were characterized by a mean length of 466 bp, an average Phred quality of 40 and a mean coverage of 4.64 reads. GC content remained the same as in the first assembly (details relating to contigs and singletons are shown in Table 2). Changes in length and quality distribution of contigs from the first to the second round assembly are shown in Additional file 2 and Additional file 3 respectively.

We performed the iterative assembly process being aware that some degree of assembly accuracy is lost. In fact, by forcing MIRA to resolve ambiguous positions by choosing a consensus, the probability of losing rare transcriptional variants is increased. However, two assembly cycles were performed for two reasons: 1) we were interested in having a general overview of genes expressed in *A. naccarii* by minimizing redundancy, and 2) information on rare variants can be traced back, realigning all the original reads on the corresponding contigs. After assembly, all reads of origin were aligned against belonging contigs and metacontigs, obtaining a multiple alignment for each of them. The distribution of the average coverage observed in the contigs and metacontigs from the first and final assemblies are reported in Additional file 4. Pair-wise relationships between sequence length, number of reads per contig and average sequence quality after the two assemblies are shown in Additional file 5. All contigs and cleaned reads are provided within the AnaccariiBase database, available at the web page: http://compgen.bio.unipd.it/anaccariibase/. From here on, we will no longer make any distinction between contigs and metacontigs and both will be indicated simply as contigs.

Table 2 Contigs and singletons summary statistics for first round and final assembly by MIRA of *A. naccarii* transcriptome

	Round 1	Final assembly
Reads assembled (#)	256,738	256,738
Reads assembled (%)	77.43	77.43
Total contigs (#)	***44,232***	***42,193***
Contigs (%)	72.72	76.32
Total contigs length (Mb)	22.64	21.87
Average contigs length (bp)	511.89	518.29
Average contigs GC content (%)	38.49	38.83
Average contigs quality (phred)	43	43
Average contigs coverage (bp/position)	3.2	4.09
Total singletons (#)	***16,593***	***13,089***
Singletons (%)	27.28	23.68
Total singleton length (Mb)	4.98	3.91
Average singleton length (bp)	300.19	298.55
Average singleton GC content (%)	38.39	38.46
Average singleton quality (phred)	28	28

Table 3 Metacontigs summary statistics after the second round assembly by MIRA

	Round 2
Total metacontigs (#)	4,203
Reassembled contigs (#)	6,242
Percentage of reassembled contigs	14.11
Reassembled singletons (#)	3,504
Percentage of reassembled singletons	21.12
Total consensus metacontig (Mb)	2.95
Metacontig average length (bp)	700.94
Percentage of metacontig average GC content	38.83
Metacontig average consensus quality (phred)	45
Metacontig average coverage (bp/position)	1.66

Functional annotations

De novo annotation of *A. naccarii* transcriptome was performed with multi-step procedure starting from similarity search against gender specific nucleotide sequences, main protein and nucleotide databases, full transcribed and protein sequences from other fishes in Ensembl database.

BLAST against sequences available from the genus *Acipenser*

The comparison of *A. naccarii* sequences with 6,088 ESTs for the genus *Acipenser* already available revealed 8,804 *A. naccarii* contigs (15.93%) matching 2,047 different subjects (33.62%). The limited percentage of matching sequences can probably be ascribed to the different tissues of origin: gonad and brain in the Adriatic sturgeon, and mainly pituitary gland, skin and spleen in the reference database.

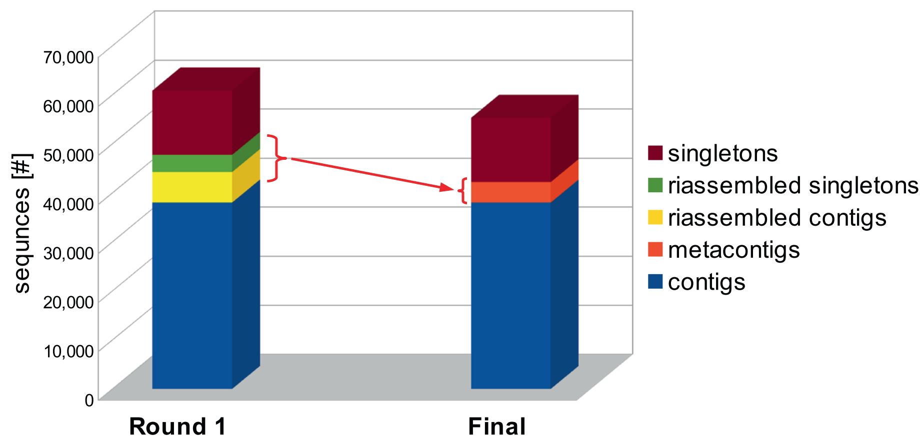


Figure 1 Redundancy reduction after two assembly rounds with MIRA for *A. naccarii* data. Graphical representation of the contigs and singletons built in the first assembly round, which were re-assembled as metacontigs in the second round, and then joined to get the final assembly.

BLASTX against the main protein sequence databases

The comparison of contigs and singletons to the NCBI non-redundant protein database (nr) using BLASTX, came out with 9,850 contigs and 2,339 singletons (22.05% of total sequences) matching 9,433 different known or predicted proteins. The taxonomic classification of hits from the nr database, by species, is represented in Figure 2.

BLASTX search in Swiss-Prot section of the UniProtKB database, identified 11,088 transcripts (20.06%) with significant matches against 7,111 different well-annotated proteins.

BLASTN against the main nucleotide database

The BLASTN search against the NCBI nucleotide database (nt) identified significant similarity for 10,195

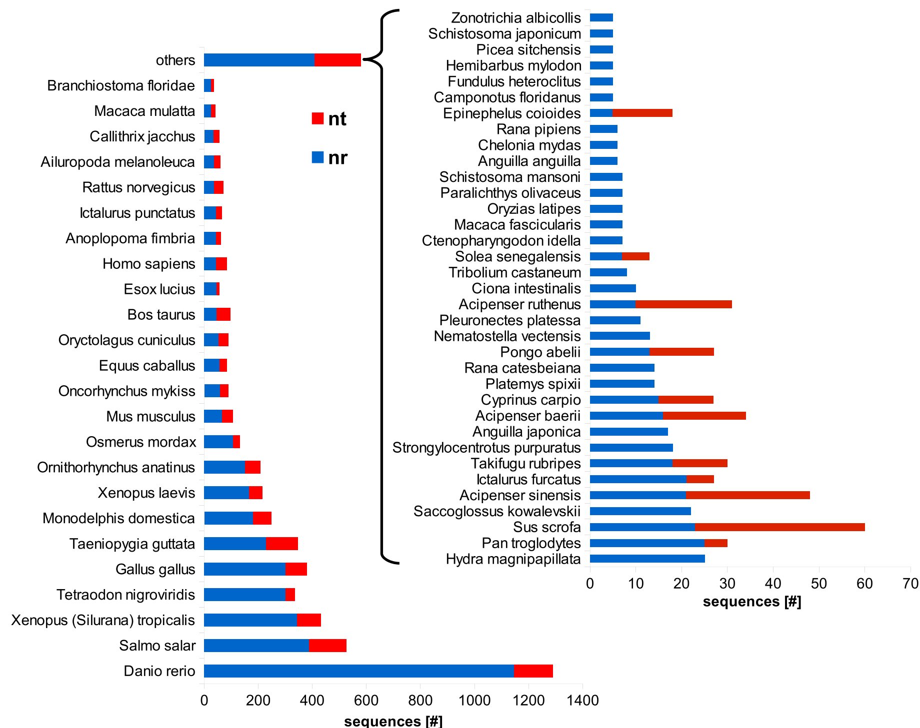


Figure 2 Taxonomic classification of *A. naccarii* contig annotations. Assignment of annotations obtained from BLASTX and BLASTN comparisons (e-value 1e-03) of contigs against NCBI nr and nt databases to different species was performed with MEGAN 4, based on the absolute best BLAST hits. The bar chart shows contigs annotated with the 24 more-represented species in annotations from nr. The contribution of annotations from nt, for the same species, is marked in red. "Others" includes the 34 species less represented in nr annotations.

transcripts (18.44%) with 4,509 different subjects. Among sequences with a significant match against nt, 5,366 had not previously been matched against nr and Swiss-Prot databases. Considering all the BLAST searches performed so far, a total of 17,734 ESTs obtained at least one hit, representing 32% of the Adriatic sturgeon transcriptome.

Evaluation of the unannotated fraction

A total of 43,093 non-redundant transcripts remained unannotated after the BLAST search against the nr database. ORF prediction showed that 41,935 of these sequences (97.31%) contain a putative open reading frame.

Evolutionary comparison with other fishes

The non-redundant contigs of the two *A. naccarii* libraries were compared to Ensembl release 66 complete cDNA sets for the species of the RS-list.

TBLASTX and BLASTX best hit results are collected in Tables 4 and 5 respectively. The fraction of *A. naccarii* transcripts that identify putatively orthologous genes in other fish and humans reflects the phylogenetic distance between *A. naccarii* and other species. The two non-teleost species on the RS-list, the Sea Lamprey and the Coelacanth, share a higher fraction of genes 33.49% and 30.57% respectively as also confirmed at the protein level (31.97% and 29.24%). A possible explanation for this finding is that these two species separated from the ancestor of teleosts before the Whole Genome Duplication (WGD) known to have interested the teleost clade [15]. Part of the newly-formed genetic material is known to have persisted after duplication, possibly evolving new functions and thus becoming unrecognisable by sturgeon ESTs. This process of gene diversification results in a reduction of the percentage of detectable genes in teleosts. However, a careful analysis of the matching transcripts and proteins should be performed to confirm this hypothesis.

In any case, the number of putatively orthologous genes matched by *A. naccarii* transcripts in other species is expected to be influenced not only by the genetic similarity among species but also by different parameters such as the accuracy of the genome characterisation in the different species used for the comparison and their evolutionary history in which, for example, different mutation rates may play an important role. In fact, different lines of Osteichthyes are known to have very different evolutionary rates as a result of different factors such as metabolic features or generation times [16]. These differences may deeply affect the number of genes that can be recognized as orthologous among species. Zebrafish seems to share fewer genes (22.89% through transcripts and 22.42% through proteins) with *A. naccarii* than do other teleosts but, the fraction of the *A. naccarii* matching ESTs is comparable to other species. The conclusion is that the Danio genome seems to have a higher number of genes. However this could have a different explanation: first, the number of genes is actually higher according to the high level of genes retention after the WGD hypothesized for this species [17] second, more simply, this result is biased by the more complete genome characterization for this model organism.

Evaluation of the non-coding RNA component

NcRNA are implicated in every step of gene expression. To discover and annotate potential non-coding RNAs in our transcriptome (miRNA, rRNA, MtrRNA, snoRNA, lncRNA), we searched for genes corresponding to non-coding RNA from genomes of the fish species described above, using BLASTN. Alignment results are collected in Table 6. The highest number of alignments was found against miRNA sequences from the 4 teleosts, in particular in Medaka, whose 9 miRNA were found to be homologous in sturgeon. Mitochondrial and ribosomal RNA were next in abundance. Surprisingly, 11 rRNA pseudogenes from humans were found to be homologous in *A. naccarii.* The alignment method used here can underestimate the number of ncRNA detected as different types of ncRNA have different degrees of sequence conservation between species, with miRNA and snoRNA usually well-conserved while longer-functional ncRNA are not [18]. Moreover, lncRNA elements tend to maintain a consensus secondary structure through

Table 4 TBLASTX best hit (e-value < 1e-03) of *A. naccarii*

	Lamprey	Coelacanth	Danio	Stickleback	Medaka	Fugu	Tetraodon	Human
reference cDNAs (#)	11,476	21,958	48,636	27,628	24,662	48,003	23,265	180,654
reference genes (#)	10,449	19,174	27,948	20,839	19,687	18,685	19,749	47,266
***A. naccarii* ESTs with hit on reference cDNAs (#)**	***7,431***	***12,428***	***13,068***	***11,528***	***11,270***	***11,143***	***10,886***	***12,740***
A. naccarii ESTs with hit on reference cDNAs (%)	13.44	22.48	23.64	20.85	20.39	20.16	19.69	23.05
reference genes identified (#)	3,499	5,862	6,396	5,710	5,565	5,447	5,429	6,116
reference genes identified (%)	***33.49***	***30.57***	***22.89***	***27.4***	***28.27***	***29.15***	***27.49***	***12.94***

Sequences from known-, novel- and pseudo-gene predictions, from Ensembl realise 66, were collected for the following species: *Petromyzon marinus, Latimeria chalumnae, Danio rerio, Gasterosteus aculeatus, Oryzias latipes, Takifugu rubripes, Tetraodon nigroviridis* and *Homo sapiens*. *A. naccarii* transcriptome sequences were searched against each database. For each sequence, the best hit was annotated.

Table 5 BLASTX best hit (e-value < 1e-03) of *A. naccarii*

	Lamprey	Coelacanth	Danio	Stickleback	Medaka	Fugu	Tetraodon	Human
reference proteins (#)	11,429	21,817	41,693	27,576	24,661	47,841	23,118	97,041
reference genes (#)	10,402	19,033	26,160	20,787	19,686	18,523	19,602	21,860
A. naccarii ESTs with hit on reference proteins (#)	***7,052***	***11,264***	***11,530***	***11,010***	***10,816***	***10,766***	***10,359***	***11,102***
A. naccarii ESTs with hit on reference proteins (%)	12.76	20.38	20.86	19.92	19.57	19.47	18.74	20.08
reference genes identified (#)	3,326	5,565	5,865	5,526	5,403	5,346	5,283	5,478
reference genes identified (%)	***31.97***	***29.24***	***22.42***	***26.58***	***27.45***	***28.86***	***26.95***	***25.06***

Best hits from the alignment of *A. naccarii* transcriptome sequences against all translations from known-, novel- and pseudo-gene predictions in Ensembl realise 66 for the different species considered in this work.

compensatory base mutations and, therefore, are difficult to detect by sequence alignments alone [19].

GO annotation

We started the GO annotation from the BLASTX results against nr. GO terms were retrieved from the association to best-hit for 10,036 (18.15%) of the overall 55,282 *A. naccarii* sequences. Protein domains and motif information were retrieved by InterProScan via Blast2GO and corresponding annotations were merged with already existent GO terms. A total of 29,671 contigs provided significant InterProScan information, with only 3,326 of them resulting in GO annotation. After merging, 6,344 unique GO terms (3,811 for biological process, 758 for cellular component, 1,775 for molecular function), were successfully transferred to 8,784 contigs (16%). As expected, the evidence code distribution shows an over-representation of electronic annotations (IEA), although other non-automatic codes, such as Inferred from Direct Assay (IDA) and inferred by mutant phenotype (IMP), were also well represented (see bar-plot in Additional file 6).

A. naccarii ESTs were classified by GO-slims within the biological process, molecular function and cellular component domains and a Direct Acyclic Graph (DAG) of the ontologies was generated. Figure 3 shows the number of putative ESTs annotated with high-level GO terms by cutting the DAG graph at level 3 for each of the 3 domains. We also performed enzyme code (EC) annotation through Blast2GO for sequences with GO annotations and retrieved KEGG maps for the metabolic pathways in which they participate. In total 3,634 ESTs were annotated with 448 ECs that identify unique enzymes, participating in 116 different pathways (see Additional file 7). The most populated pathways are "Purine metabolism" (map 00230) with 33 enzymes involved, "Arginine and proline metabolism" (map 00330) with 25 enzymes and Glycolysis/Gluconeogenesis (map 00010) with 22 enzymes.

Estimation of sequencing completeness

To evaluate the coverage of the cDNA libraries by reads, a rarefaction analysis similar to that described in [20] was employed. The extrapolation from the hyperbolic model, fitted onto the average points, obtained by the 10 replications of sampling and reference transcript identification, using reads from the male library only, showed that 7,293 different transcripts were potentially identifiable in the Danio cDNA set (asymptote "*a*" of the model function). The 6,043 different transcripts actually identified using all reads represents 83% of the theoretical maximum. The angular coefficient calculated at final read count was 0.157. Using reads from the female library only, the number of transcripts actually identified was 5,989, which, compared to the 7,176 maximum transcripts identifiable at infinite sequencing, represents 83% of the total. The slope at the final read count was 0.145. Finally, putting together reads from both libraries, the model-based extrapolation denoted 8,262 different transcripts potentially identifiable, and the 7,286 actually identified represents 88%. The three extrapolated curves are shown in Additional file 8. As expected, the slope at maximum read count was 0.140.

Table 6 BLASTN best hit (e-val < 1e-03) of *A. naccarii* transcriptome against non-coding RNA genes from Ensembl database

	Lamprey	Coelacanth	Danio	Stickleback	Medaka	Fugu	Tetraodon	Human
reference ncRNA (#)	2,628	2,918	4,431	1,617	735	703	813	9,399
A. naccarii ESTs with hit (#)	4	17	28	29	19	21	31	62
A. naccarii ESTs with hit (%)	***0.01***	***0.03***	***0.11***	***0.05***	***0.05***	***0.03***	***0.04***	***0.06***
reference ncRNA identified (#)	4	7	11	11	11	7	13	23
reference ncRNA identified (%)	***0.15***	***0.24***	***0.25***	***0.68***	***1.5***	***1***	***1.6***	***0.24***

All non-coding RNA genes and pseudogenes in Ensembl realise 66 for the different species were searched against *A. naccarii* transcriptome.

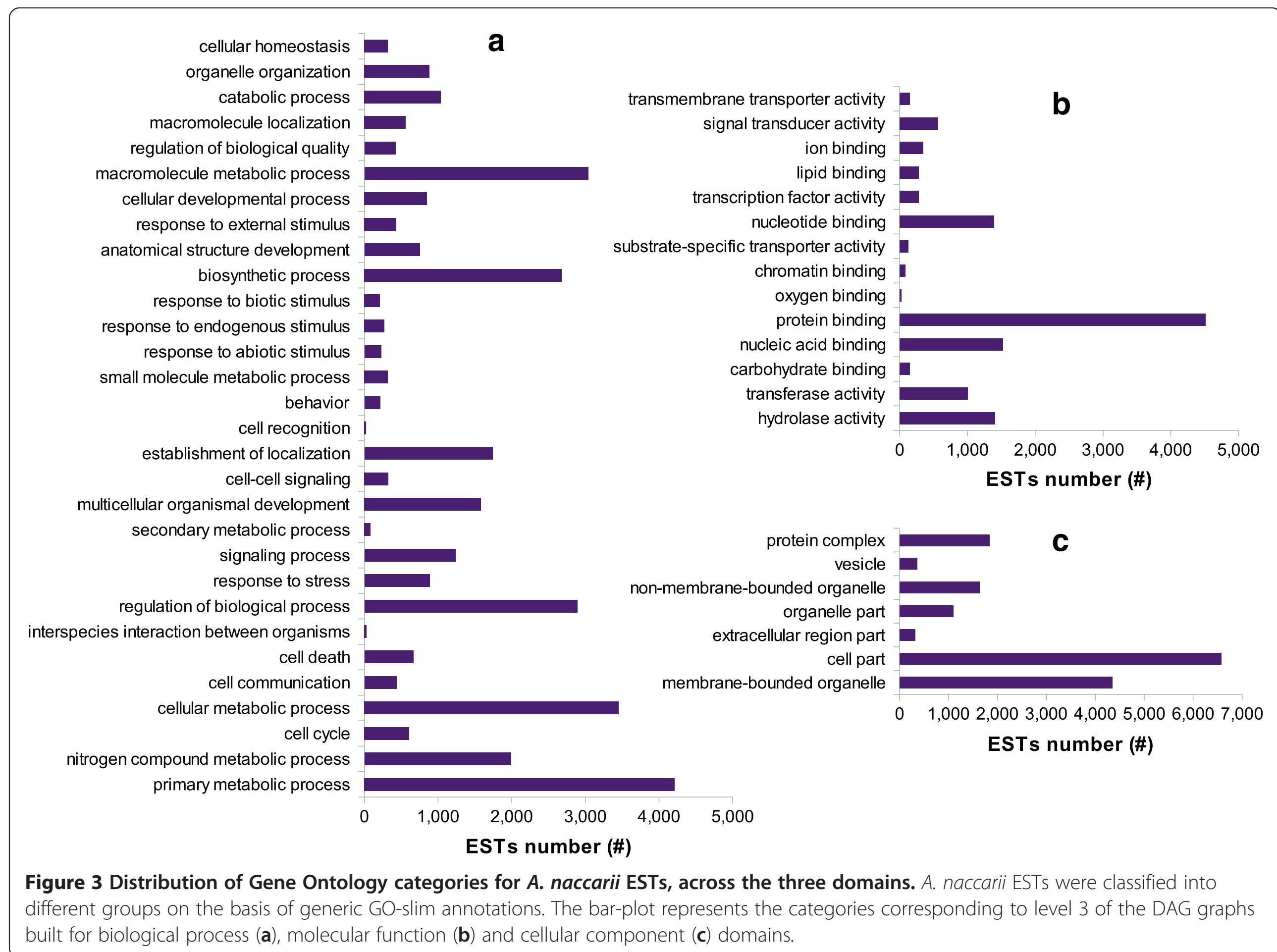


Figure 3 Distribution of Gene Ontology categories for *A. naccarii* ESTs, across the three domains. *A. naccarii* ESTs were classified into different groups on the basis of generic GO-slim annotations. The bar-plot represents the categories corresponding to level 3 of the DAG graphs built for biological process (**a**), molecular function (**b**) and cellular component (**c**) domains.

Further analysis, exemplified in Additional file 9, showed that by changing the reference cDNA datasets, the absolute value of the potentially identified transcripts and those actually identified changes, but the ratio between these quantities remains nearly constant. Therefore, the latter ratio is a robust value indicating the fraction of the cDNA libraries really sequenced.

Estimation of transcriptome completeness

To estimate the total number of A. *naccarii* transcripts potentially present in the two tissues (gonads and brain), we adapted the capture-recapture method widely used in ecology to estimate animal population sizes [21]. This method requires a precise estimate of the fractions of ESTs that can be considered common between the male and female libraries (see method). Since, before joint assembly, each read was labelled with the library of provenance, final contigs were classified according to the origin of their reads as being cDNA3-specific, cDNA4-specific or common [22]. First, we separated 17,399 cDNA3-specific contigs from the male library (31% of the total) and 17,523 cDNA4-specific contigs from the female library (32% of the total). The direct subtraction between the two groups of library-specific contigs isolated 394 contigs showing mutual alignments from each fraction. The indirect subtraction identified 41 cDNA3-specific and 38 cDNA4-specific contigs, that aligned on 85 common subjects. Finally, using NCBI nr as the common database, we identified an additional 13 cDNA3-specific and 12 cDNA4-specific contigs which map onto the same 10 protein sequences. After all subtractions, 16,951 cDNA3-specific and 17,079 cDNA4-specific contigs remained, that may represent potentially sex-distinctive transcripts.

With the Rcapture R package we estimated the transcripts population size to be 68,904 with a standard error of 210. This means that we have probably sequenced about 80% of the total transcripts in the two tissues of *A. naccarii*. Additional redundancy could still be present in the common contig fraction (despite attempts to reduce it in the assembly phase), which would have increased the estimated total number of transcripts in the tissues of origin. By adopting the capture-recapture approach we are aware that the resulting percentages probably represent an

over-estimate of the real fraction of captured transcripts. In fact, the procedures used for the identification of the transcripts common to both libraries may cause some bias due, for example, to a non-correct identification of gene families or to different variants of duplicated genes, especially when dealing with a tetraploid species like *A. naccarii*. Moreover, the assumption that the two fish share the same transcriptomes might be bold because the different genders of the two animals could be responsible not only for different genes directly involved in sex determination but also, for example, for differences in the developmental rate. Nevertheless, we think that the approach here proposed, even if indicative when applied to a single comparison, might be very useful for comparative analyses of multiple libraries with the purpose of estimating the relative completeness, especially if obtained by the same sample.

The transcriptome completeness was also evaluated, as noted above, by searching for constitutively-expressed mitochondrial genes. Of the 12 polypeptide coding genes of the white sturgeon *Acipenser transmontanus* mitochondrial genome, 11 were found in our assembly. Only the gene for ATPase subunit 8 was missing. Contigs that aligned with these genes showed between 93 and 100% identity.

Search for sex-determining genes

We evaluated the presence of 32 candidate genes known to be involved in sex determination and sexual development in vertebrates by queering the transcriptome with 3 collections of orthologs and paralogs for those genes (Ensembl Compara, Homologene, Acipenser-specific genes, see methods). The first collection represents the largest variety of annotated homologous (orthologs and paralogs), from sequenced genomes, categorised in Ensembl Compara. The second collection is represented by clusters of more specific orthologs, downloaded from NCBI HomoloGene [23]. The third group of sequences is a collection of complete or partial CDSs from other sturgeon species of the genus Acipenser available in NCBI GenBank. The use of large collections of putative orthologs and paralogs maximizes the possibility of detecting homologues. In contrast, restricted collections of reliable homologues allow a higher confidence on the match they find. If a contig is confirmed as the best subject for a given gene in searches of all trees, then we can be more confident about its identity.

Significant matches were found for 22 of the 32 genes investigated. The alignments of matching contigs were manually inspected to exclude false-positive matches exclusively due to the presence of widespread protein domains. A complete list of the best matching contigs considered to have a reliable similarity against the 22 genes recognised is contained in Additional file 10. A similar transcriptomic screening for genes involved in sex differentiation was performed on the lake sturgeon (*A. fulvescens*) [24]. The authors report positive matches for 12 genes (SOX2, SOX4, SOX17, SOX21, SOX9, DMRT1, RSPO1, WT1, WNT4, FOXL2, TRA-1, FEM1), all but one included in our search list, the exception being TRA-1. All genes were also detected in *A. naccarii* with the exceptions of DMRT1 and WNT4. Positive matches with SOX genes (SOX2, SOX4, SOX21) were discarded after manual inspection, because the same contigs also matched other SOX genes with higher scores. This multiple matching is due to the fact that genes of the SOX family often share the conserved High Mobility Group box domain and assignment based on this domain makes for a less-reliable identification.

The absence of the DMRT1 gene from both the *A. naccarii* libraries is especially interesting and might be due to the incomplete coverage. A second possibility is that this gene is not expressed at the stage at which our samples were collected. In fact, the animals analysed for this project were six months old and were at an early stage of gonad differentiation. This is, to our knowledge, the first stage at which sturgeons, which cannot be sexed visually, have unambiguous evidence of gonad differentiation through fine histological investigation [25]. The lake sturgeons analysed by Hale and colleagues [24] were estimated to be 13 or 14 years old. All characterisations of DMRT1 genes from other sturgeon species have been performed on mature or sub-mature animals [24,26]. Finally, a low expression of this gene is displayed in the Siberian sturgeon with no evident gonad differentiation [27]. Thus, the absence of DMRT1 in the transcriptome of the very young *A. naccarii* analysed would suggest that this gene is expressed at a later stage of development in this species (and probably in all sturgeons). DMRT1 in known to play an important role as an activator of the genetic cascade of sex differentiation in some other fish, such as Medaka [28]. Even if most of the genes involved in sex determination are known to act in a dosage-dependent manner [29], under the hypothesis that sex differentiation in sturgeon is genetically determined, one could expect that, at the origin of the genetic cascades leading to the different genders, a sex-linked genomic polymorphism occurs. For this reason, special attention was given to the contigs observed to be library-specific. Among the 22 genes detected, only 5 (WT1, LHX1, CYP19A1 (aromatase), FHL3, FEM1A) and 2 (AR, EMX2), were found to be specific to male (cDNA3) or female (cDNA4) libraries. These genes represent, in our opinion, interesting candidate transcripts for experimental validation by PCR amplification. The remaining fifteen genes were detected by contigs belonging to the common fraction.

Discovery of variants

At 90% Bayesian probability, we were able to identify 23,084 SNPs and 59,150 INDELs. After having filtered out variants beside simple sequence repeats, 21,791 SNPs (94.04%) and 57,996 INDELs (98.05%) were retained from 6,283 and 8,678 contigs respectively. Between contig-containing variants, the average SNP per contig was 3.5, while the mean INDELs per contig was 6.7. The mean frequency across all contigs was 1 SNP every 1. Kbp, and 1 every 377 bp for the INDELs. We identified 14,433 transitions (Ts) and 7,358 transversion (Tv), thus confirming that transitions are more common than transversions in our dataset [30].

We then classified SNPs that fell in predicted coding regions according to the type of mutations: non-synonymous (Ka) or synonymous mutations (Ks). Of the overall contig-containing SNPs, we were able to identify a putative ORF for 2,482 of them on the basis of the best match against nr database, while for 3,786 an ORF was predicted. Of the overall SNPs found in coding regions, 2,750 represented non-synonymous mutations while 1,056 were synonymous. We found that 1,280 contigs (2.32% of all contigs) had Ka/Ks > 1 thus indicating genes putatively under diversifying selection within our samples. On average, we found 0.73 non-synonymous and 0.28 synonymous SNPs per contig in coding regions; this means one non-synonymous mutation every 9 Kbp of coding portion, and 1 synonymous mutation every 20.7 Kbp. Distribution of SNPs and INDELs across contigs together with distributions of Ka/Ks are shown in Figure 4. We also scanned the entire EST set for Sample Sequence Repeats (SSRs, also known as microsatellites), and we identified 5,295 SSRs present in simple formation, within a total of 4,670 (8%) contigs. In particular, we found 1,891 dinucleotides, 2,377 trinucleotides, 1,001 tetranucleotides, 100 pentanucleotides and 45 hexanucleotides. The graph in Figure 5 shows the frequency of repeat types found accordingly to unit size. Of the overall contig-containing SSRs, 4,639 also contain a putative ORF. In total 1,779 SSRs are predicted within ORFs (33% of all identified SSRs). This availability of a relevant number of EST-linked microsatellites and SNPs represents a precious prerequisite for sturgeon conservation genetics by providing the possibility to monitor the effect of selection on captive and released stocks.

AnaccariiBase: a free genomic resource for *A. naccarii*

Freely available at: http://compgen.bio.unipd.it/anaccariibase/, AnaccariiBase contains *A. naccarii* transcriptome information and results of bioinformatics analysis, organised in different layers. The database is focused on contig sequences and annotations, and can be searched through

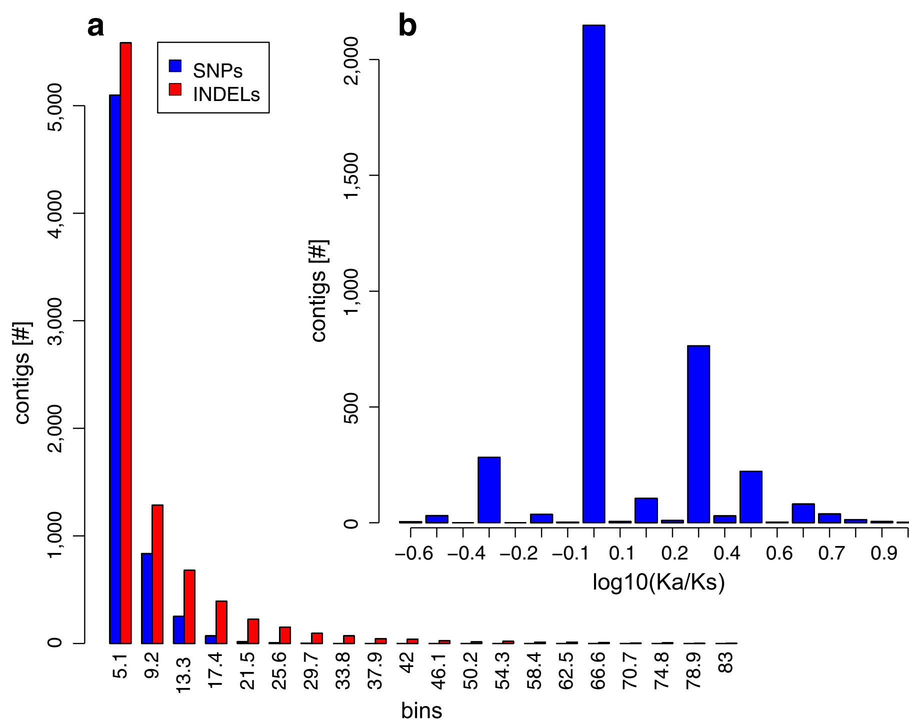


Figure 4 Distribution of SNPs and INDELs across *A. naccarii* contigs. a) Bar-plot of the distribution of SNPs and INDELs in contig-containing variants. Most contigs contain up to 5 variants. **b**) Bar-plot of log10 distribution of Ka/Ks for contig-containing SNPs that lie in predicted ORFs. Contigs with Ka/Ks > 1 (log10 > 0) are proposed to be under diversifying selection in *A. naccarii*.

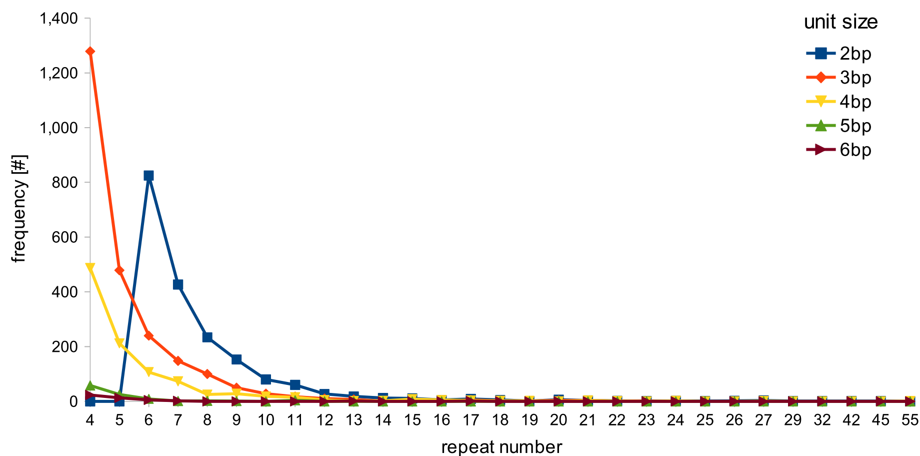


Figure 5 Frequency of classified SSR repeat types in *A. naccarii*. The graph shows the frequency of each repeat motif classified, considering the sum of the frequencies for complementary sequences (for example, the sum of frequencies for the dinucleotides AC and its complementary GT), for the 5,295 total SSRs identified in 4,670 contigs.

contig ID and key-words. Moreover, it allows the user to conduct a local BLAST search on the fly against contigs to identify one or more transcript significantly similar to a given query sequence. Furthermore the system provides a customizable data retrieval tool to download large amounts of data. The information layers are detailed hereafter. (1) Contig information. For each contig, an ID is given together with the FASTA sequence and an informative description, which is defined by the Blast2GO natural language text mining functionality, related to the BLAST hits. The best hit is used when a Blast2GO description is unavailable. (2) Assembly. The list of the reads assembled into each contig is accessible to the user, together with their sequences. (3) Gene Ontology. GO terms associated to each transcript are given for Biological Process, Molecular Function, and Cellular Component domains, with hyper-link to the GO database. (4) BLAST results. Pre-calculated BLAST results of contigs against the main nucleotide and protein databases, are shown in the classic BLAST output format. Results are hyperlinked to the external databases, and include alignment descriptions and details about the pairwise alignments of each contig with the corresponding BLAST hits.

Conclusions

The present study provides the first insight into the transcriptome of the Adriatic sturgeon, a critically endangered species endemic to the Adriatic Sea. More generally, this is also the first large release of transcriptomic information for a sturgeon species, shared through a dedicated and searchable database. With over 55,000 high quality sequences, the information reported represents a significant advance in sturgeon genetics. The apparently limited fraction of successfully annotated sequences with GO terms (16%) might be due to the very ancient separation (about 250 MYBP) of sturgeons from any other species for which a relevant genomic information is available. Additionally, following the sturgeon-specific Whole Genome Duplications [4] part of the redundant genetic information probably underwent a functional divergence that may have further decreased the fraction of successfully annotable ESTs. Beside the evolutionary interest of a database obtained from a member of the Chondrostean, certainly applied genetics studies on sturgeons will benefit from this resource. The present paper also report the results of an investigation on genes related to sex differentiation. Out of the 32 genes investigated 7 were detected in only one of the two libraries suggesting a possible differential expression between genders at this early stage of gonad differentiation. This result might be affected by the limited coverage of our sequencing and should be considered as a starting point for further investigations. Interestingly, DMRT1, a master gene for the sex determination known to be expressed in both sexes in different sturgeon species was not detected, suggesting that, differently from other fish species, DMRT1 is expressed in sturgeons only in latter stages of maturity. Finally, the availability of thousands of EST-linked microsatellites makes possible the establishment of a genome-wide genetic markers panel useful to monitor the effect of different selective pressures and to monitor the effects of restocking practices. Restocking of most sturgeon species depends on *ex situ* conservation because of the dramatic decline of natural populations [1]. In synthesis, the data provided in the present study and shared through a dedicated website represents the first substantial release of information on a sturgeon transcriptome and will hopefully constitute a useful contribution to sturgeon genetics, aquaculture, and conservation.

Methods

Preparation of samples, construction of cDNA libraries and sequencing

Two 6-month-old individuals were collected from the "Azienda Agricola VIP" farm (Orzinuovi, Brescia, Italy). Their sex was determined by histological analysis of the gonads as being one male and one female. Animals were anaesthetised with chloretone and painlessly killed. No analyses or experiments were conducted with live animals. The University of Padova ethic board CEASA (Comitato Etico di Ateneo per la Sperimentazione Animale) exempted this study from review as an extra moenia activity. Biopsies were performed from gonads and brain for RNA purification and a part of the gonads was used for sex determination according to the procedures described in Grandi and Chicca [25]. An RNeasy mini-column kit (QIAGEN) was used to extract total RNA from 30 mg of each tissue from each individual. Total RNA was checked for integrity, purity and size distribution. RNA samples from each individual were pooled and stored in three volumes of 96% ethanol and 0.1 volume of sodium acetate to obtain 5 μg of pooled RNA in a final volume of 120 μl. Pooled RNA was sent to Evrogen (Moscow, Russia; www.evrogen.com). The SMART (Switching Mechanism At 5′ end of RNA Template) kit was used to retrotranscribe total poly-adenilated RNA. First-strand cDNA synthesis was performed with SMART Oligo II oligonucleotide (5`-AAGCAGTGGTATCAACGCAGAGTACGCrGrGrG-3′) and CDS-GSU primer (5′-AAGCAGTGGTATCAACGCAGAGTACCTGGAG-d(T)20-VN-3′) using 0.3 μg of total RNA. Double-strand cDNA was obtained from 1 μl of the first-strand reaction (diluted 5 times with TE buffer) by PCR with SMART PCR primer (5′-AAGCAGTGGTATCAACGCAGAGT-3′). Amplified cDNA PCR product was purified using QIAquick PCR purification Kit (QIAGEN, CA). The two SMART prepared libraries were then normalised using the duplex-specific nuclease (DSN) method [31]. Normalisation included PCR amplification of the normalised fraction.

In order to gain more material, 30 ng of normalised cDNA were used for 100 μl PCR and 7 cycles of PCR amplification with SMART PCR primer were performed as suggested by Evrogen (http://www.evrogen.com/kit-user-manuals/Trimmer-2.pdf). Moreover, in order to decrease the possibility to introduce biases due to PCR, the above amplification was independently replicated 20 times and the products pooled before sequencing. Adapters were trimmed using GsuI (Fermentas) following the standard protocol and cDNA purification was performed with Agencourt AMPure XP (BECKMAN COULTER). BMR Genomics, University of Padua, Italy (http://www.bmr-genomics.it), prepared and sequenced 454 protocol libraries. Approximately 15 μg of normalized cDNA from each library were sequenced in a 1/4 picotiter plate on a Genome Sequencer FLX instrument using GS FLX Titanium series reagents.

Cleaning and assembly

The raw reads from every library were extracted from 454 SFF files through the open source alternative sff_extract 0.2.10. Summary control of raw reads' quality was done with FastQC 0.10.0. Sequences were cleaned using the est_process module driven by preprocessest.pl into the est2assembly 1.13 package [32] that perform sequencing adaptor removal, low complexity region masking, quality trimming, and poly A/T detection and removal. After preprocessing, *A. naccarii* male and female reads were tagged accordingly to the library of origin and jointly assembled, thus allowing contigs to be classified for reads content as being composed by males only, by females only or by both sexes. Sequences of the two libraries were jointly assembled by MIRA 3.2.1 [33]. The obtained contigs and singletons were further re-assembled by performing a second round to decrease the redundancy caused by the heuristic nature of the assembly process [34]. In the first run (*de novo* assembly), all cleaned reads were used as input and processed with the following parameters: -job = denovo, est, accurate, 454; -LR:fo = no, -SB:lsd = no, -CL:cpat = 0:qc = 0, -ED:ace = 1, -OUT:sssip = yes, -CO:fnicpst = yes, -LR:mxti = no, -AS:mrpc = 1. In the second run the following parameters were used: -job = denovo, est, accurate, 454, -notraceinfo, -LR:fo = no, -CL:cpat = 0:qc = 0, -ED:ace = 1, -OUT:sssip = yes, -CO:fnicpst = yes, -LR:mxti = no, -AS:mrpc = 1. After the second assembly step, the native reads of each contig (from the first round) or metacontig were traced back through SSAHA2 2.5.4 [35] with parameters (–rtype 454 -output sam). For each of the above contigs or metacontigs, the alignment of the corresponding reads resulted in a SAM file allowing the calculation of the coverage using SAMtools [36] and BEDtools [37].

Functional annotations

BLASTX 2.2.25+ [38] similarity searches for the entire transcriptome were conducted locally against the NCBI non-redundant (nr) database (downloaded 2010/10/19) as the first *de novo* annotation step. The Swiss-Prot part of the UniProt database (downloaded on 2012/02/24) was also queried. Local TBLASTX 2.2.25+ similarity searches were conducted locally against (1) 6,088 EST sequences of the genus *Acipenser* downloaded from NCBI-taxonomy (2011/04/18), mainly obtained from *A. sinensis* [39] and *A. transmontanus* [40] (2) a super-set of all transcripts resulting from Ensembl (release-66) including known-, novel- and pseudo-gene predictions for the following list of reference species (hereafter RS-list):

Petromyzon marinus, Latimeria chalumnae, Danio rerio, Gasterosteus aculeatus, Oryzias latipes, Takifugu rubripes, Tetraodon nigroviridis and *Homo sapiens.* In order to identify non-coding sequences, BLASTN 2.2.25+ similarity searches were conducted locally against whole non-coding RNA gene and pseudogene sequences from Ensembl release-66 for the species of the RS-list. A BLASTN similarity search was also performed against the NCBI nucleotide sequence (nt) database (downloaded on 2012/02/24). BLASTX, BLASTN and TBLASTX searches were carried out using default parameters.

Given the high evolutionary distance among the species compared, alignments with an e-value < 1e-03 were considered significant and a maximum of 20 hits were taken into account for each query. The taxonomic classification of annotations was performed by MEGAN 4 [41] based on the absolute best BLAST hits. Contigs with multiple best BLAST hits were excluded from the count. The mapping of GO annotations to contigs was achieved with Blast2GO 2.4.7 [42]. Annotations were conducted only for contigs with significant BLASTX hits below e-value 1e-06, with 55 as the annotation cut-off and 5 as the GO weight. No HSP-hit coverage cut-off was used. InterProScan annotation was also conducted via Blast2GO. Obtained information for domains was included to improve global annotations.

Estimation of sequencing completeness

To test how completely our physical cDNA libraries were sequenced, we adopted the method described in Franssen et al. [20], based on saturation curve calculation. From the total cleaned reads pool, increasing subsets of reads were randomly selected and, for each read, the corresponding contig in which it was assembled was traced back. Detected contigs were blasted against a reference cDNA set using TBLASTX with the e-value cut-off at 1e-03. The best matching subject was recorded for each contig. The sampling was repeated 20 times with a constant increase in sample size, reaching the totality of cleaned reads in the last run, thus identifying, in the end, 20 pools of different reference cDNAs. The number of matching reference cDNAs at each cycle was plotted against the corresponding reads sample size and a hyperbolic model $y = ax/(b + x)$ was fitted to the points by non-linear regression to assess the parameters "a" and "b" with "a" representing the upper limit of the model function, i.e., the maximum theoretical number of reference transcripts identifiable by the initial cDNA libraries if these had been exhaustively sequenced. Moreover, the slope of the hyperbolic curve at maximum sample size gives an evaluation of how quickly the asymptotes "a" will be reached, thus indicating the decreasing potential to detect additional transcripts. We built saturation curves by sampling cleaned reads from: 1) male only, 2) female only and 3) joint libraries. In all cases, we mapped reads back to the final assembly contigs. The whole cDNA super-set from *Danio rerio* in Ensembl release-66 was chosen as the reference. However, our analysis demonstrated that the fraction of detected reference transcripts, with respect to the maximum estimated, and the slope of the curve at maximum sample size do not substantially change using different cDNA sets as a reference (see Additional file 9).

Estimation of transcriptome completeness

We inferred the total transcripts population size in the two *A. naccarii* samples by estimating the number of transcripts shared by the two independent libraries taking into account that the two animals analysed had the same age and the same history. By neglecting the differences due to sex-specific transcripts, the two sequence libraries were handled as two sampling replicates from the same transcripts' population. The fraction of the transcript from the first library that is also represented in the second one is a direct estimate of the completeness of the second library and vice-versa. The same approach has already been applied to estimate the number of human genes [43].

Since each read was labelled with the library of origin before joint assembly, final contigs were classified as being "male_library-specific", "female_library-specific" or "common". The common one is the fraction of contigs composed of reads of both libraries and then represented by transcripts considered to be shared by the two libraries. We performed a direct subtraction, i.e. a bidirectional BLASTN, between the libraries to identify the contigs that were not library-specific. Library-specific contigs that align for more than 80% of their length, with e-values below 1e-50 were moved into the common fraction. We also performed an indirect subtraction to take into account contigs representing partially-overlapping or non-overlapping portions of the same long transcript, which had not been assembled together due to the lack of a sufficient link. Both groups of library-specific contigs were searched for similarities, using TBLASTX, against cDNA sets resulting from Ensembl release-66 for the RS-list species. All cDNA sequences provided by NCBI-Taxonomy Browser for the genus Acipenser (2011-04-18) were also screened. Protein sequences available for other species were assessed by searching the NCBI nr database (2010/11/02) with BLASTX. The library-specific contigs matching the same subjects, with 1e-06 as the e-value threshold and > 80% query coverage were moved into the common fraction. We exploited the Rcapture R package [44] to estimate the total transcripts population sizes because it allows the association of a standard error to the obtained estimation. Furthermore, we assessed the completeness of the *A. naccarii* transcriptome by screening

for the presence of the 13 polypeptide coding genes [45,46] from the complete mitochondrial genome (mt) of *Acipenser transmontanus* (GenBank accession no.: AB042837) using BLASTN with a 1e-10 e-value threshold.

Search for sex-determining genes

We obtained sequences from genes known to be involved in sex determination and sexual development in vertebrates from different species and used them to search our assembled contigs by similarity in order to investigate the content of library-specific contigs, isolated by *in silico* subtraction in more detail. The genes and gene families considered were: WT1, LHX1, CYP19A1, FHL3, FEM1A, AR, EMX2, DAX1, SOX9, SOX17, SOX1, SOX11, SOX6, SOX14, FOXL2, RSPO, SF1, FGFR2, FGF9, GATA4, LHX9, ATRX, SOX2, SOX4, SOX21, WNT4, SRY, STRA8, FIGLA, AMH, VTG2, DMRT1 [47-49]. We obtained sequences in 3 different ways: 1) Ensembl database annotated orthologous and paralogous of the above genes were identified, starting from the well-annotated Zebrafish genome in Ensembl 66, by querying each common name. For each gene, we identified all orthologous and paralogous within Ensembl Compara version 66. Then, for each ortholog and paralog, all alternative transcripts were identified and the corresponding protein sequence downloaded. 2) Clusters of homologs (paralogs and orthologs) of candidate genes were identified within NCBI HomoloGene Release 66 and corresponding protein sequences were downloaded. 3) Nucleotide sequences for genes FOXL2, DMRT1, and SOX used as references in a previous scientific study aimed at gender identification in the Shovelnose sturgeon (*Scaphirhynchus platorynchus*) [26] together with corresponding sequences from other sturgeons of the genus Acipenser were downloaded from NCBI Genbank (15/10/2012). Each group of paralog and ortholog protein and nucleotide variant representing a gene was searched for similarity in our transcriptome assembly using TBLASTN and BLASTN respectively. Alignments with an e-value > 1e-03 and fewer than 50 positive matching nucleotide/aminoacid positions in the BLAST alignment were discarded. Each different contig that presented a match was extracted for each gene. For each contig (subject) matched by more than one homologue (query), the homologue with the highest alignment bit score was selected. Results obtained by the three approaches were compared for each gene and the more-likely contig was selected based on the following criteria: 1) BLAST alignment bit-score with the query; 2) per-base mean coverage (singletons were discarded); 3) nucleotide alignments between candidates to ensure they actually represented distinct sequences (using MAFFT v6.935b [50]); 4) alignments between contig translations and corresponding protein queries (using MAFFT); 5) presence of one or more distinctive and important functional domains encoded by the target gene within the translated and aligned fraction of contigs (by searching in Pfam-A version 26.0 [51]); 6) the ratio between the length of the translated-aligned fraction and the total contig length; 7) consistency of annotations obtained by blast2GO via alignment against all protein sequences included in the NCBI non-redundant database.

Discovery of variants

Since mean contig coverage is generally low (< 5X) and the transcriptome comes from different individuals, we adopted a method based on a probabilistic framework, which allows the estimation of uncertainty regarding variants calling, in order to identify SNPs and short INDELs [52]. We used Freebayes 0.9.4 [53] which employs Bayesian formulation to calculate the probability that multiple different alleles are present between the reference and the aligned reads. Freebayes is also able to call variants from polyploid pooled samples. SAM alignments calculated for each contig in the assembly phase were input into Freebayes with the following parameters: probability cut-off of 0.9, 5 as the minimum coverage required to process a site, and each SNP must be supported by at least 2 reads. As it has been shown that improved base-call accuracy can lead to a significant reduction in false-positive SNP calls, base alignment quality (BAQ) adjustment was applied to the input alignments through SAMTOOLS calmd 0.1.18 [36]. Homogenisation of the potential insertion and deletion distribution through reads-independent left realignment to improve the INDELs call was obtained by the bamleftalign tool included in the Freebayes package. It is known that variant calling near repetitive DNA sequences are prone to error, especially in 454 technology where over-calls or under-calls of repetitive stretch, are the most common errors [54]. We then filtered out all variants that were beside 4 repetitions of any sample sequence repeats (including homopolymeric regions). For each contig containing SNP, we calculated the number of transitions (Ts) and transversions (Tv). The mutation resulting from each SNP was characterized in terms of synonymous (Ks) or non-synonymous (Ka), and location (inside the ORF or in the 5′ or 3′ UTR regions). For each contig-containing SNP with a BLAST hit against the nr database, the ORF was deduced from the alignment against the best HSP. For contig-containing SNPs without a hit, the ORFs predicted by the ORFpredictor were used. We then calculated the ratio $(Ka + 1)/(Ks + 1)$ where, 1 was added to enable the calculation of the ratio even when $Ks = 0$. To find both perfect and imperfect microsatellite repeats (SSRs) for di-, tri-, tetra-, penta- and hexa-nucleotides in unit-size, within our contig sequences, we adopted the MISA tool version 1.0 [55], with min_repeat specifications of 6, 4, 4, 4, and 4 respectively. These thresholds are in agreement with the minimum lengths

recommended for repetitions outlined in [56], to allow the polymerase slippage events, which makes the identified microsatellites, potentially polymorphic. We set at 0 the maximal number of nucleotides that interrupt compound microsatellites. Moreover we distinguished SSRs within ORFs predicted by BLAST comparisons or in alternative by ORFpredictor.

Additional files

Additional file 1: Distribution of cleaned-read lengths for the *A. naccarii* male (cDNA3), female (cDNA4) and the joined libraries. Bin intervals are shown along the x-axis.

Additional file 2: Distribution of contig- and singleton- lengths for *A. naccarii* first round and final assemblies. While the average quality of singletons remains between 15 and 40, the average quality of assembled contigs rises to 88.

Additional file 3: Distribution of contigs' and singletons' average quality for first round and final assemblies. The figure shows how the number of singletons and contigs resulting from the first assembly (largest contig 2,732, N50 contig size 489, N90 contig size 324, N95 contig size 258), is reduced in the final set.

Additional file 4: Mean contigs coverage distribution for the first and final assemblies. Percentage of contigs falling in the different coverage intervals are referred to the final assembly. The average coverage of the contigs is quite low. As shown on the graph where about 61% of contigs have average coverage of only up to 3 per base.

Additional file 5: Pair-wise relationships between main properties characterising total contigs obtained by the first and final assemblies. Pair-wise relationships between lengths and qualities (A, D), lengths and number of reads per contig (B, E), qualities and mean reads per contig (C, F), in the set of 60,825 contigs from the first assembly and 55,282 contigs from the (second) reassembly of the *A. naccarii* transcriptome.

Additional file 6: Evidence code distribution of the annotation of *A. naccarii* transcriptome. Only the evidence codes assigned to at least one sequence are reported. EXP: Inferred from Experiment, IDA: Inferred from Direct Assay, IPI: Inferred from Physical Interaction, IMP: Inferred from Mutant Phenotype, IGI: Inferred from Genetic Interaction, IEP: Inferred from Expression, Pattern, ISS: Inferred from Sequence or Structural Similarity, ISO: Inferred from Sequence Orthology, ISA: Inferred from Sequence Alignment, ISM: Inferred from Sequence Model, RCA: inferred from Reviewed Computational Analysis, TAS: Traceable Author Statement, NAS: Non-traceable Author Statement, IC: Inferred by Curator, ND: No biological Data available.

Additional file 7: KEGG pathways found in the *A. naccarii* transcriptome. Enzyme Codes were mapped on the sequences with GO annotations through Blast2GO, then metabolic pathway map numbers in which enzymes carry out their function were retrieved, thus identifying 833 different enzymes participating in 116 different pathways.

Additional file 8: Saturation curve for male, female and both *A. naccarii* cDNA libraries. Read subsets of increasing sample size were randomly extracted from total pool in the library. For each subset, contigs in which reads where assembled were identified. Each contigs pool was used to identify *Danio rerio* cDNAs (TBLASTX 2.2.25+ e-value 1e-03). Re-sampling and identification process was repeated 10 times for each sample size. A mean value and a confidence interval for the number of identified *Danio* cDNAs was calculated for each sample size. Hyperbolic model $y = (ax)/(b + x)$ was fitted on points given by sample size versus average cDNAs hit so that model parameters "a" and "b" were estimated. The legend shows estimated parameter values obtained by fitting the hyperbolic model on the data. As can be seen, the curves from the single libraries retain the same trend and the difference is mostly due to the different number of reads in each library.

Additional file 9: Saturation curves plot for joint *A. naccarii* cDNA libraries against cDNA sets from other fishes. Were constructed several saturation curves, starting from total reads from the two libraries and using different sets of cDNA as a references. cDNA sets used are derived from all transcripts from Ensembl 66 for species in RS-list. The estimated parameters of the curves are reported in the legend.

Additional file 10: Sex related genes found in the Adriatic sturgeon transcriptome. The table lists the contigs of the *A. naccarii* transcriptome that best represent 22 of the 32 genes known to be involved in sex determination and sexual development of vertebrates, used for the screening. For each recognized gene are shown: the gene symbol; the gene extended name; the queries from the different sources that gave the best alignments with the defined thresholds, the NCBI HomoloGene cluster ID the query belongs to (for queries from this source); the Ensembl ID (for queries from this database), the GenBank Accession (possibly for all queries); the contig that best represents the putative *A. naccarii* orthologous (subject); the assembly fraction the contig belongs to (cDNA3, cDNA4, or common); its mean per-base coverage; the bit score of the alignment between the contig and its query; the contig translated-aligned fractions on the query; the putative Pfam domains contained within the translated and aligned fractions; finally the contig annotation given by Blast2GO.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
MV conceived the study, performed the bioinformatics analysis and wrote the manuscript. AG: participated to the study design and to the preparation of the manuscript. EB and FB performed the laboratory parts of the study. AC: built the web-database. GG: performed the sex determination by histological analysis of the gonads AK: participated in the design of the study. LC: conceived and coordinated the study and wrote the manuscript. All authors read and approved the final paper.

Acknowledgements
We are most thankful to Sergio Giovannini for providing the samples used in this study and to prof. Lorenzo Zane from Department of Biology, University of Padova for insightful discussion. A special thanks to Dr. Paolo Martini and Dr. Gabriele Sales from the same Department, for bioinformatics advice. We also thanks CARIPARO for providing the Ph.D. fellowship and University of Padova - Progetto di Ateneo (grant CPDA087543) for providing funds.

Author details
[1]Department of Biology, University of Padova, Via G. Colombo 3 I-35131, Padova, Italy. [2]Department of Biology, University of Ferrara, Via L.Borsari 46 I-44100, Ferrara, Italy. [3]Abteilung für Botanische Genetik und Molekularbiologie Botanisches, Institut und Botanischer Garten, Christian-Albrechts-Universität, Kiel, Germany.

Received: 21 January 2013 Accepted: 4 June 2013
Published: 18 June 2013

References
1. Congiu L, Pujolar JM, 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.
2. Bemis WE, Findeis EK, Grande L: **An overview of Acipenseriformes.** *Environ Biol Fishes* 1997, **48**:25–71.
3. Ludwig A, Belfiore NM, Pitra C, Svirsky V, Jenneckens I: **Genome duplication events and functional reduction of ploidy levels in sturgeon (Acipenser, Huso and Scaphirhynchus).** *Genetics* 2001, **158**:1203–1215.
4. Fontana F, Congiu L, Mudrak VA, Quattro JM, Smith TIJ, Ware K, Doroshov SI: **Evidence of hexaploid karyotype in shortnose sturgeon.** *Genome* 2008, **51**:113–119.
5. Boscari E, Barbisan F, Congiu L: **Inheritance pattern of microsatellite loci in the polyploid Adriatic sturgeon (Acipenser naccarii).** *Aquaculture* 2011, **321**:223–229.

6. Fontana F, Lanfredi M, Kirschbaum F, Garrido-Ramos MA, Robles F, Forlani A, Congiu L: **Comparison of karyotypes of Acipenser oxyrinchus and A. sturio by chromosome banding and fluorescent in situ hybridization.** *Genetica* 2008, **132**:281–286.
7. Birstein VJ, Vasiliev VP: **Tetraploid-octoploid relationships and karyological evolution in the order Acipenseriformes (Pisces) karyotypes, nucleoli, and nucleolus-organizer regions in four acipenserid species.** *Genetica* 1987, **72**:3–12.
8. 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.
9. Peng H, Zhang J, Wu X: **The ploidy effects in plant gene expression: Progress, problems and prospects.** *Sci China C Life Sci* 2008, **51**:295–301.
10. Ludwig A: **Identification of Acipenseriformes species in trade.** *J Appl Ichthyol* 2008, **24**:2–19.
11. Wuertz S, Gaillard S, Barbisan F, Carle S, Congiu L, Forlani A, Aubert J, Kirschbaum F, Tosi E, Zane L, Grillasca J-P: **Extensive screening of sturgeon genomes by random screening techniques revealed no sex-specific marker.** *Aquaculture* 2006, **258**:685–688.
12. Van Eenennaam AL, Van Eenennaam JP, Medrano JF, Doroshov SI: **Evidence of female heterogametic genetic sex determination in white sturgeon.** *J Hered* 1999, **90**:231–233.
13. Keyvanshokooh S, Pourkazemi M, Kalbassi MR: **The RAPD technique failed to identify sex-specific sequences in beluga (Huso huso).** *J Appl Ichthyol* 2007, **23**:1–2.
14. FastQC: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
15. Sato Y, Nishida M: **Teleost fish with specific genome duplication as unique models of vertebrate evolution.** *Environ Biol Fishes* 2010, **88**:169–188.
16. Krieger J, Fuerst PA: **Evidence for a slowed rate of molecular evolution in the order acipenseriformes.** *Mol Biol Evol* 2002, **19**:891–897.
17. Woods IG: **The zebrafish gene map defines ancestral vertebrate chromosomes.** *Genome Res* 2005, **15**:1307–1314.
18. Pang KC, Frith MC, Mattick JS: **Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function.** *Trends Genet* 2006, **22**:1–5.
19. Nawrocki EP, Kolbe DL, Eddy SR: **Infernal 1.0: inference of RNA alignments.** *Bioinformatics* 2009, **25**:1335–1337.
20. Franssen SU, Shrestha RP, Bräutigam A, Bornberg-Bauer E, Weber APM: **Comprehensive transcriptome analysis of the highly complex Pisum sativum genome using next generation sequencing.** *BMC Genomics* 2011, **12**:227.
21. Chao A: **Estimating Population Size for Sparse Data in Capture-Recapture Experiments.** *Biometrics* 1989, **45**:427–438.
22. Schwartz TS, Tae H, Yang Y, Mockaitis K, Van Hemert JL, Proulx SR, Choi J-H, Bronikowski AM: **A garter snake transcriptome: pyrosequencing, de novo assembly, and sex-specific differences.** *BMC Genomics* 2010, **11**:694.
23. Altenhoff AM, Dessimoz C: **Phylogenetic and functional assessment of orthologs inference projects and methods.** *PLoS Comput Biol* 2009, **5**:e1000262.
24. Hale MC, Jackson JR, DeWoody JA: **Discovery and evaluation of candidate sex-determining genes and xenobiotics in the gonads of lake sturgeon (Acipenser fulvescens).** *Genetica* 2010, **138**:745–756.
25. Grandi G, Chicca M: **Histological and ultrastructural investigation of early gonad development and sex differentiation in Adriatic sturgeon (Acipenser naccarii, Acipenseriformes, Chondrostei).** *J Morphol* 2008, **269**:1238–1262.
26. Amberg JJ, Goforth R, Stefanavage T, Sepúlveda MS: **Sexually dimorphic gene expression in the gonad and liver of shovelnose sturgeon (Scaphirhynchus platorynchus).** *Fish Physiol Biochem* 2009, **36**:923–932.
27. Berbejillo J, Martinez-Bengochea A, Bedó G, Vizziano-Cantonnet D: **Expression of dmrt1 and sox9 during gonadal development in the Siberian sturgeon (Acipenser baerii).** *Fish Physiol Biochem* 2013, **39**:91–94.
28. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, Jindo T, Kobayashi D, Shimada A, Toyoda A, Kuroki Y, Fujiyama A, Sasaki T, Shimizu A, Asakawa S, Shimizu N, Hashimoto S, Yang J, Lee Y, Matsushima K, Sugano S, Sakaizumi M, Narita T, Ohishi K, Haga S, Ohta F, Nomoto H, Nogata K, Morishita T, Endo T, Shin-I T, Takeda H, Morishita S, Kohara Y: **The medaka draft genome and insights into vertebrate genome evolution.** *Nature* 2007, **447**:714–719.
29. Ferguson-Smith M: **The Evolution of Sex Chromosomes and Sex Determination in Vertebrates and the Key Role of *DMRT1*.** *Sex Dev* 2007, **1**:2–11.
30. Kimura M: **A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.** *J Mol Evol* 1980, **16**:111–120.
31. Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA, Shagin DA: **Simple cDNA normalization using kamchatka crab duplex-specific nuclease.** *Nucleic Acids Res* 2004, **32**:e37.
32. Papanicolaou A, Stierli R, Ffrench-Constant RH, Heckel DG: **Next generation transcriptomes for next generation genomes using est2assembly.** *BMC Bioinforma*, **10**:447–447.
33. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S: **Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs.** *Genome Res* 2004, **14**:1147–1159.
34. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH: **Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing.** *Mol Ecol* 2008, **17**:1636–1647.
35. Ning Z, Cox AJ, Mullikin JC: **SSAHA: a fast search method for large DNA databases.** *Genome Res* 2001, **11**:1725–1729.
36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**:2078–2079.
37. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26**:841–842.
38. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications.** *BMC Bioinforma* 2009, **10**:421.
39. Cao H, Leng X, Li C, Wei Q, Gui J, Cheng H, Zhou R: **EST dataset of pituitary and identification of somatolactin and novel genes in Chinese sturgeon, Acipenser sinensis.** *Mol Biol Rep* 2012, **39**:4647–4653.
40. Lazzari B, Mariani V, Malinverni R, Caprera A, Giuffra E: **A comparative gene index for the white sturgeon Acipenser transmontanus.** *Mar Genomics* 2008, **1**:15–21.
41. Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC: **Integrative analysis of environmental sequences using MEGAN4.** *Genome Res* 2011, **21**:1552–1560.
42. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite.** *Nucleic Acids Res* 2008, **36**:3420–3435.
43. Ewing B, Green P: **Analysis of expressed sequence tags indicates 35,000 human genes.** *Nat Genet* 2000, **25**:232–234.
44. Baillargeon S, Rivest L-P: **Rcapture: Loglinear Models for Capture-Recapture in R.** *J. Stat. Softw* 2007, **19**:1–31.
45. Bobrowicz AJ, Lightowlers RN, Chrzanowska-Lightowlers Z: **Polyadenylation and degradation of mRNA in mammalian mitochondria: a missing link?** *Biochem Soc Trans* 2008, **36**:517–519.
46. Rorbach J, Minczuk M: **The post-transcriptional life of mammalian mitochondrial RNA.** *Biochem J* 2012, **444**:357–373.
47. Pala I, Schartl M, Thorsteinsdóttir S, Coelho MM: **Sex Determination in the Squalius alburnoides Complex: An Initial Characterization of Sex Cascade Elements in the Context of a Hybrid Polyploid Genome.** *PLoS One* 2009, **4**:e6401.
48. Shirak A, Seroussi E, Cnaani A, Howe AE, Domokhovsky R, Zilberman N, Kocher TD, Hulata G, Ron M: **Amh and Dmrta2 Genes Map to Tilapia (Oreochromis spp.) Linkage Group 23 Within Quantitative Trait Locus Regions for Sex Determination.** *Genetics* 2006, **174**:1573–1581.
49. McClelland K, Bowles J, Koopman P: **Male sex determination: insights into molecular mechanisms.** *Asian J Androl* 2012, **14**:164–171.
50. Katoh K, Frith MC: **Adding unaligned sequences into an existing alignment using MAFFT and LAST.** *Bioinformatics* 2012, **28**:3144–3146.
51. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, Bateman A: **The Pfam protein families database.** *Nucleic Acids Res* 2009, **38**:D211–D222.
52. Nielsen R, Paul JS, Albrechtsen A, Song YS: **Genotype and SNP calling from next-generation sequencing data.** *Nat Rev Genet* 2011, **12**:443–451.

53. Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** *Arxiv E-Prints* 2012, **1207**:3907.
54. Brockman W, Alvarez P, Young S, Garber M, Giannoukos G, Lee WL, Russ C, Lander ES, Nusbaum C, Jaffe DB: **Quality scores and SNP detection in sequencing-by-synthesis systems.** *Genome Res* 2008, **18**:763–770.
55. *MISA - MIcroSAtellite identification tool.* http://pgrc.ipk-gatersleben.de/misa/.
56. Mudunuri SB, Rao AA, Pallamsetty S, Nagarajaram HA: **Comparative analysis of microsatellite detecting software: a significant variation in results and influence of parameters**. In *Proc. Int. Symp. Biocomput.* New York, NY, USA: ACM; 2010:38:1–38:7.

doi:10.1186/1471-2164-14-407
Cite this article as: Vidotto *et al.*: Transcriptome sequencing and *de novo* annotation of the critically endangered Adriatic sturgeon. *BMC Genomics* 2013 **14**:407.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

APPENDIX II

Molecular Phylogenetics and Evolution 66 (2013) 223–232

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

*Tana*1, a new putatively active *Tc*1-like transposable element in the genome of sturgeons

José Martin Pujolar [a], Laura Astolfi [b], Elisa Boscari [a], Michele Vidotto [a], Federica Barbisan [a], Alice Bruson [c], Leonardo Congiu [a,*]

[a] Department of Biology, University of Padova, I-35131 Padova, Italy
[b] Department of Neuroscience, University of Padova, Padova, Italy
[c] Department of Pediatrics, Clinical Genetics Unit, University of Padova, Padova, Italy

ARTICLE INFO

Article history:
Received 24 July 2012
Revised 21 September 2012
Accepted 23 September 2012
Available online 29 September 2012

Keywords:
Acipenseriformes
Sturgeons
Transposons
Vertical transmission
Horizontal transfer

ABSTRACT

We report the discovery of a new putatively active *Tc*1-like transposable element (*Tana*1) in the genome of sturgeons, an ancient group of fish considered as living fossils. The complete sequence of *Tana*1 was first characterized in the 454-sequenced transcriptome of the Adriatic sturgeon (*Acipenser naccarii*) and then isolated from the genome of the same species and from 12 additional sturgeons including three genera of the Acipenseridae (*Acipenser*, *Huso*, *Scaphirhynchus*). The element has a total length of 1588 bp and presents inverted repeats of 210 bp, one of which partially overlapping the 3′ region of the transposase gene. The spacing of the DDE motif within the catalytic domain in *Tana*1 is unique (DD38E) and indicates that *Tana*1 can be considered as the first representative of a new *Tc*1 subfamily. The integrity of the native form (with no premature termination codons within the transposase), the presence of all expected functional domains and its occurrence in the sturgeon transcriptome suggest a current or recent activity of *Tana*1. The presence of *Tana*1 in the genome of the 13 sturgeon species in our study points to an ancient origin of the element that existed before the split of the group 170 million years ago. The dissemination of *Tana*1 across sturgeon genomes could be interpreted by postulating vertical transmission from an ancestral *Tana*1 with a particularly slow evolutionary rate Horizontal transmission might have also played a role in the dissemination of *Tana*1 as evidenced by the presence of a complete copy in the genome of Atlantic salmon. Vertical and horizontal transmission are not mutually exclusive and may have concurred in shaping the evolution of *Tana*1.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Transposable elements (TEs) are DNA sequences that can move or transpose themselves from one position to another within the genome. TEs are widely distributed in bacteria, plants and animals, and make up a substantial fraction of the genomes in which they reside (Kidwell and Lisch, 2000; Bowen and Jordan, 2002; Miskey et al., 2005). The human genome consist of about 50% transposons (Feschotte et al., 2002), while nearly 85% of the maize genome is composed of hundreds of families of transposable elements (Schnable et al., 2009). It is well established that variation in the amount of transposons occupying the genome considerably accounts for the large differences in genome size across eukaryotes (Kidwell, 2002; Wessler, 2006).

It has been hypothesized that the insertion of transposable elements into new genome locations has contributed to shape genome architecture and led to the emergence of genetic innovation (Feschotte and Pritham, 2007). TEs can produce a wide array of changes in the genome of their hosts, from generation of new alleles through insertion into genes to restructuration of genomes through chromosomal rearrangements (Wessler, 2006). Bursts in transposon activity can also lead to genome size expansion (Slotkin and Martienssen, 2007). In addition, many TEs originate in one organism but can move around between species and invade new genomes, since these elements are not dependent on host factors to mediate their mobility (Shan et al., 2005).

While TEs have the faculty to mutate genes, alter gene functions and generate new genes, hence enhancing the evolutionary potential of their hosts, from a strictly theoretical point of view, TEs are considered as parasitic or junk DNA (Doolittle and Sapienza, 1980; Orgel and Crick, 1980), which has been a topic of extensive debate (Kidwell and Lisch, 2000, 2001). At the host organismal level, negative selection commonly results from TE-induced deleterious mutations, while neutrality of TEs is mostly associated with inactive TEs (Kidwell and Lisch, 2000). Both host and TEs seem to have evolved strategies to minimize the deleterious impact of transposition. In this sense, TEs tend to be located

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

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2012.09.025

in heterochromatin regions, which are generally gene poor and have a low level of transcription, and where the deleterious impact of transposition is less important (Ikeda et al., 2007; Slotkin and Martienssen, 2007). Host organisms have also developed defence mechanisms against high transposon activity rates including DNA methylation and RNA silencing (reviewed in Slotkin and Martienssen, 2007).

Transposable elements are classified according to their structural organization and mechanism of transposition. Class I elements (retrotransposons) replicate through reverse transcription of an RNA intermediate. Class II elements (DNA transposons) move directly by a cut-and-paste mechanism operated by a single transposase enzyme encoded by the element. The transposase excises the transposon after recognizing the terminal inverted repeats (TIRs) flanking the transposon and then integrates the transposon in a new target site. The *Tc*1/mariner superfamily of transposable elements is the most widespread class of DNA transposons in nature, being found in fungi, plants and animals, including nematodes, insects, fish, amphibians and humans (Plasterk et al., 1999). *Tc*1/mariner elements are about 1300–2400 bp in length and contain a transposase gene flanked by terminal inverted repeats (Plasterk et al., 1999). Within this superfamily, different families can be identified based on the distance between the last two residues in the conserved catalytic motif DDE (Shao and Tu, 2001; Yuan and Wessler, 2011). While *Tc*1/mariner elements have been demonstrated to be active in nematodes and arthropods, all of the copies isolated until recently in vertebrates are defective in that they contain frameshifts, insertion/deletions and stop codons within the transposase gene, including the *Sleeping Beauty* transposon in salmonids, an ancestral element presumed to be active >10–15 million years ago that was synthetically reconstructed by eliminating the inactivating mutations (Ivics et al., 1996). However, the recent study of Clark et al. (2009) demonstrated that *Passport*, a transposon isolated from the European plaice (*Pleuronectes platessa*), is the first vertebrate *Tc*1-like element both natively intact and functionally active.

The present paper reports the discovery of *Tana*1, a new putatively active *Tc*1-like element found in the genome of sturgeons. First, we isolated and characterized the element in the Adriatic sturgeon (*Acipenser naccarii*), which we also identified in 12 additional sturgeon species of the same Acipenseriformes order. The study of transposons in sturgeons, an archaic group of chondrostean fish often referred as living fossils that comprises a total of 25–27 species distributed in the Northern Hemisphere (Rochard et al., 1991), is particularly relevant from an evolutionary perspective. In addition to their ancestral radiation (around 170 MYA; Krieger et al., 2008), the sturgeon genome shows some interesting peculiarities: (1) sturgeons possess a large number of chromosomes and different levels of ploidy across species, being karyotypically divided into two main groups with approximately 120 and 240 chromosomes, respectively. The 120-chromosome condition is the ancestral one, while 240 chromosome-karyotypes were reached through at least three independent duplication events during sturgeon evolution (Ludwig et al., 2001). The two groups are believed to be functionally diploid and tetraploid, respectively (Fontana et al., 2008; Boscari et al., 2011); (2) the genome of sturgeons is highly fragmented and presents a large number of microchromosomes of <20 MB in size (Fontana et al., 1999); (3) sturgeons show a high frequency of interspecific hybridization with viable progeny that is also fertile when the parental species have the same degree of ploidy (Birstein, 2002). Ludwig et al. (2001) suggested that together with gene silencing and chromosomal rearrangements, transposons might have played an important role in the formation of the Acipenseridae genome.

2. Material and methods

2.1. Sampling and DNA extraction

Genomic DNA was extracted from caudal fin clips (10–100 mg) of 13 sturgeon (Acipenseridae) species, using the DNAeasy Blood and Tissue Extraction kit (Qiagen). The Adriatic sturgeon (*Acipenser naccarii*) was used for the identification and characterization of the *Tana*1 element. The presence/absence of *Tana*1 was investigated in the remaining 12 species, including (i) a total of 10 *Acipenser* species: *A. baerii*, *A. brevirostrum*, *A. fulvescens*, *A. gueldenstaedtii*, *A. mikadoi*, *A. nudiventris*, *A. ruthenus*, *A. stellatus*, *A. sturio*, *A. transmontanus*; (ii) the beluga sturgeon *Huso huso*, which phylogenetic studies have shown to be embedded within the *Acipenser* group (Birstein and DeSalle, 1998; Krieger et al., 2008) (iii) *Scaphirhynchus albus*, which belongs to a separate genus of sturgeons within the Acipenseridae.

2.2. Tana1 identification in A. naccarii

With the aim of detecting repeated sequences to be used as probes for in situ hybridizations, genomic DNA of a single *A. naccarii* individual was cleaved with MboI (Amersham Biosciences) and separated by electrophoresis in a 0.7% low melting agarose gel. A single band visible after staining with ethidium bromide was isolated from the agarose gel, ligated into a pUC18 plasmid (pUC18) by Ready-to-go™ pUC18 BamHI/BAP + Ligase kit (Amersham Biosciences) and transformed in competent *Escherichia coli* TOP10 cells (Invitrogen). Recombinant clones were identified by amplification using universal vector primers M13 Forward and Reverse. All amplifications were performed using a Gene Amp PCR System 2700 thermocycler (Perkin Elmer) in a total volume of 20 μl. PCR reactions consisted of 10–50 ng of template DNA, 0.25 μM of each primer, 1× buffer (GE Healthcare), 0.1 mM dNTPs and 0.5 U Taq DNA polymerase (GE Healthcare). PCR cycling conditions were as follows: 2 min at 94 °C, 25 cycles of 30 s at 94 °C, 45 s at 55 °C and 45 s at 72 °C, and final extension for 5 min at 72 °C.

A subset of recombinant clones was sequenced. Nucleotide sequences were aligned using nucleotide and protein databases by *blastn* and *blastx* suites, respectively (http://www.ncbi.nlm.nih.gov). A partial *Tana*1 sequence was obtained, as identified by its high amino acid similarity with putative transposases of *Tc*1-like elements.

The partial *Tana*1 sequence obtained after MBoI digestion and sequencing was completed by screening a normalized cDNA library of *A. naccarii* transcriptome obtained using the 454 Titanium sequencing platform. The sequence alignment tool SSAHA2 v2.5.5 (Ning et al., 2001) was used to align all trimmed reads with the reference *Tana*1 sequence. Mapped reads were extracted using SAMtools v0.1.18 (Li et al., 2009) after filtering for reliable alignments (setting 1 as mapping quality threshold). Reads were assembled together with the reference sequence using GS *De Novo* Assembler (454 Life Sciences Corp).

The longest contig was used to prolong the original reference sequence. The extended sequence was blasted against itself using the *two sequences* option of *blastn* in order to identify the inverted repeats that signalize the start and ending of the element and obtain its complete sequence. Finally, the putative complete sequence of the *Tana*1 transposon was again matched against the full transcriptome library using SSAHA2.

2.3. Direct amplification from A. naccarii and other sturgeon species

A single primer matching the extremities of the TIRs of the transposon (*TotRepeat*: 5′-TACAGTACTGTGCAAAAGTTTTAGG-CAGG-3′) was designed to amplify the entire *Tana*1 element from

A. naccarii genomic DNA. PCR conditions were the same as above but with an annealing temperature of 60 °C. Recombinant colonies were screened by sequencing after amplifications with universal M13 primers until a putatively complete element without stop codons was found. PCR products were cloned into JM109 competent cells using the P-GEM-T Easy vectors (Promega) following the recommendations of the manufacturer.

In order to classify the element, a multiple alignment of *Tc*1/*mariner* superfamily transposases was used to construct a Neighbor Joining Tree showing the relationship of *Tana*1 with other *Tc*1/*mariner* elements, including the *Tc*1, *mariner*, *Ma T*, IS630 and ITmD37E families. The alignment used to generate the tree is available from EMBL at ALIGN_000448 (Claudianos et al., 2002).

The primer *TotRepeat* also amplified the entire *Tana*1 element in the rest of sturgeon species in our study. All sequences are deposited in GenBank (accession number: JX889425–889438). Amplified PCR products were cloned and a single colony of the expected molecular size was sequenced for each species, since the scope was to investigate the presence/absence of the element in other sturgeons, not to search for complete copies of the element. Obtained sequences were aligned using Clustal_X (Thompson et al., 1997). Pairwise genetic distances were calculated between all *Tana*1 elements in MEGA 5 (Tamura et al., 2011) and graphically represented by Neighbor Joining. The Basic Local Alignment Search Tool (BLAST) was used to find sequences similar to *Tana*1 in GenBank. BLAST similarity searches were conducted locally against different NCBI (National Centre for Biotechnology Information) databases using BLASTN and BLASTX. Distances were calculated between *Tana*1 and the most similar amino acid sequences found and a dendrogram was constructed using Neighbor-Joining.

Finally, the element was also amplified from *A. naccarii* cDNA in order to confirm the presence of the complete element in the transcriptome. For these amplifications, the same cDNA used to construct the libraries sequenced by 454 approach was used as template. Primers and PCR conditions were the same as above. Amplified fragments of the expected size were cloned and sequenced as previously described.

2.4. Quantification of Tana1 copy number in A. naccarii

Number of copies of the *Tana*1 transposon in *A. naccarii* was estimated using the Comparative Quantitation method (CQ; Rotor Gene 5.0 software, Corbett Research), a Real-Time PCR analysis that does not require a standard curve. The method combines two absolute quantification reactions: one for the target-specific gene and the other for the reference gene, the latter being used to determine fold-differences in copy number of the target gene. Relative concentrations of DNA during the exponential phase of the PCR reaction are determined by plotting fluorescence against number of cycles on a logarithmic sale. Relative amounts of DNA can be calculated from the take-off point of each sample, which is the cycle at which the fluorescence signal of a sample emerges above the background. The quantification of sample relative to a designated control (calibrator) can be calculated according to the formula: $(E)^{Ct1-Ct2}$, where (E) is the amplification efficiency and Ct1 and Ct2 are the take-off points of the calibrator and the sample, respectively. In a relative quantification analysis, the amount of DNA of the target gene is divided by the amount of DNA from the reference gene measured in the same sample for normalization.

The quantification of *Tana*1 was done in duplicate, using as target two different regions of the transposon. Region 1 is 120 bp long and is located between the TIR and the transposase. Region 2 is 130 bp and is located in the transposase coding region (Fig. 1). Primers used were *StuFor870* (5′-GTAAACTGTCTTCAGCAACCT-3′) and *StuFor985* (5′-TATGTAGGTTGAAACACAATCATT-3′) for Region 1 and *StuFor62* (5′-CAGGACTCCTTGTTCTTCTTTA-3′) and *StuRev197*

Fig. 1. Representation of the structure of the *Tana*1 element in *A. naccarii* including location of all primer pairs (1 = *Totrepeat*; 2 = *StuFor985*; 3 = *StuFor870*; 4 = *StuFor197*; 5 = *StuFor62*). The transposase is marked in grey. ATG = start codon. TIR = terminal inverted repeats.

(5′-CTTATCCATCATGCAATACCAT-3′) for Region 2. As reference, we used one microsatellite marker (AfuG63; Welsh et al., 2003), which is known to be of similar size to the transposon regions used (118–144 bp long in *A. naccarii*; Congiu et al., 2011). As calibrator, we used a mixture of the amplicons (purified and quantified) of all primer pairs used (both region 1 and 2 in the transposon plus the AfuG63 microsatellite). The calibrator was used at a final concentration of 10 pg/μl. The efficiency of the quantification assay was validated using serial dilutions of the calibrator at two different final concentrations (1 and 0.1 pg/μl).

Real-time PCR was performed with a fluorometric thermal cycler Rotor-Gene 6000 (Corbett Robotics) in a total volume of 25 μl. PCR reactions contained 1× PCR buffer, 3 mM MgCl2, 0.2 mM dNTPs, 0.3 μM each primer, 1.5 U Taq (Takara) and Eva-Green 1× qPCR Mix (Biotium). Optimum annealing temperature was determined using a Shuttle PCR (two temperature PCR). PCR conditions consisted of one cycle of 3 min at 95 °C followed by 40 cycles of 10 s at 95 °C and 20 s at 59 °C. Each sample was quantified in triplicate and each experiment was repeated three times.

2.5. Cytogenetic localization of Tana1 in A. naccarii

For cytogenetic analysis, chromosomes of *A. naccarii* were prepared from cell tissue cultures (Fontana et al., 1997). A fluorescent in situ hybridization (FISH) was conducted to explore the chromosome distribution of *Tana*1 following the methodology described in Lanfredi et al. (2001). The complete *Tana*1 was biotin-labeled by PCR and used as DNA probe. PCR labeling was performed by amplification in 50 mM KCl; 10 mM Tris–HCl (pH 9), 1.5 mM $MgCl_2$, dATP 200 μM, dCTP 200 μM, dGTP 200 μM, dTTP 175 μM; Biotin-16-dUTP 25 μM (Roche), 15 pM of *TotRepeat* primer, 1 U Taq DNA polymerase (Amersham Biosciences) and 100 ng genomic DNA. PCR reactions were performed in the same conditions previously described. Immunological detection was carried out at 37 °C for 1 h with fluoroscein-labeled anti-biotin antibody set (Boehringer). The chromosomes were counterstained with propidium iodide (SIGMA). Observations were made with a Leitz Orthoplan epifluorescence microscope with the appropriate filter combinations. Photographs were taken using Kodak Ektachrome 400 ASA color print film.

3. Results

3.1. Identification of Tana1 in A. naccarii

After digestion with MboI, a single band of about 950 bp was distinguishable on agarose gel. After cloning, a subset of recombinant clones was sequenced, resulting in a 971 bp sequence showing a significant similarity with *Tc*1-like transposase domains. This sequence was used as reference for the recognition of *Tana*1

in the transcriptome of *A. naccarii*. After alignment, 82 reads mapped along the reference *Tana*1 sequence, covering 100% of its length with no ambiguous positions. Average coverage was 21.97 reads per position, ranging from 4 to 46 reads. The assembly of the mapped reads together with the reference sequence produced three contigs. Contigs had an average length of 843 bp and a maximum length of 1634 bp.

The initial *Tana*1 fragment of 971 bp was extended using the longest contig of 1634 bp (constituted by the reference sequence plus 57 reads). At this point, the terminal inverted repeats (TIRs) at the end of the extended sequence were identified using *blastn*, obtaining the complete sequence of the transposon of 1588 bp in length. Finally, the putative complete transposon was aligned against the full transcriptome library. The reference was completely covered (100% in length).

3.2. Characterization of Tana1 in A. naccarii

The complete nucleotide sequence of *Tana*1 was 1588 bp long in *A. naccarii* (Fig. 1). The transposase was flanked by terminal inverted repeats that are 210 bp long. While the first TIR was separated from the 5′ extremity of the transposase gene by a 212 bp non-coding region, the second TIR partially overlapped with the coding region so that the stop codon of the transposase was found within the second TIR region. The sequence of the terminal base pairs of the TIR included the TACAGT motif, which is typical in all TIR sequences.

The coding region for the transposase was 1023 bp (341 amino acid) long. Several motifs that are characteristic of *Tc*1-like transposons (Plasterk et al., 1999) were identified in the *Tana*1 sequence (Fig. 2). Firstly, two helix-turn-helix (HTH) motifs within the bipartite DNA-binding domains on the N-terminal of the transposase, each motif consisting of threeα-helices. Secondly, a GRPR sequence located between the two HTH motifs. Thirdly, a nuclear localization signal (NLS) that partially overlaps with the third α-helix of the second HTH motif. Finally, the characteristic DDE motif within the catalytic domain, with 38 amino acids between the second aspartic acid and the glutamic acid. *Tana*1 also presented the "signature string" of the *Tc*1-*mariner* superfamily within the catalytic domain (Yuan and Wessler, 2011), an H motif downstream of the second aspartic and a P motif 10 amino acids upstream of the glutamic acid.

A 96% genetic identity was found when comparing the transposase coding region of the two *A. naccarii* sequences obtained, with a total of 14 changes out of 341 aminoacids. Both sequences were complete and did not present stop codons or insertions/deletions. The DDE signature was variable, with the third residue in one sequence being a glutamic (instead of lysine), resulting in a DDK motif.

In order to classify the element, the *Tana*1 amino acid sequence was compared with other *Tc*1-*mariner* superfamily elements (*Tc*1, *mariner*, *Ma T*, IS630 and ITmD37E families). A dendrogram constructed using Neighbor-Joining showed the *Tana*1 sequences clustering together within the *Tc*1 family clade with a high bootstrap value (Fig. 3).

3.3. Amplification of Tana1 from A. naccarii cDNA

Since the entire sequence of *Tana*1 found in the transcriptome was obtained by assembling many short reads, this cannot be considered as an indication that the complete element is transcribed. For this reason, a PCR amplification assay directly from cDNA was performed. Four complete sequences of the transposon directly amplified from cDNA were obtained, which showed no premature termination codons within the coding regions of their transposases. Comparison of genomic DNA and cDNA sequences of the transposase coding region showed an overall 97.5% sequence identity.

```
Acipenser sturio        MARLSTATRHKVVILHQQGLSQAEISRQTGVSRCAVQALLKKHKETGSIEDRRRSGRPRK 60
Acipenser naccarii-1    MARLSTATRHKVVILHQQGLSQAEISRQTGVSRCAVQALLKKHKETGNAEDRRRSGRPRK 60
Acipenser naccarii-2    MARLSTATRHKVVILHQQGLSQAEISRQTGVSRCAVQALLKKHKETGNVEDRRRSGRPRK 60
Salmo salar             MARLSTATRHKVVILNQPGLSQADISRQTGVSRCAVQALLKKHKKTGNVEERRCSGWPRR 60
                        *************** * ***** ******************** **  * ** ** **

Acipenser sturio        LTAADERHIMLTSLRNRKMSSSAISSELAENSGTLVHPSTVRRSLVRSGLHGRLAAKKPY 120
Acipenser naccarii-1    LTAADERHIMLTSLRNRKMSSSAISSELAENSGTLVHPSTVRRSLVRSGLHGRLAAKKPY 120
Acipenser naccarii-2    LTAADERHFMLSSLRNRKMSISAISSELAENSGTQVHPSTVRRSLVRSGLHGRLAAKKPY 120
Salmo salar             ITAADERHIMLTSLRNRKMSSSAISSELAENSGTLVHPSTVRKDLVRSGLRGRLVAKKPY 120
                         ******* ** ******** ************* *******  ****** *** *****

Acipenser sturio        LRRGNKAKRLNYARKHSNWSAEKWQQVLWTDESKFEIFGCSRRQFVRRRAGERYTNECLQ 180
Acipenser naccarii-1    LRRGNKAKRLNYARKHRNWGAEKWQQVLWTDESKFEIFGCSRRQFVRRRAGERYTNECLQ 180
Acipenser naccarii-2    LQRGNKAKRLNYARKHRNWGAEKWQQVLWTDESKFEIFGCSRRQFVRRRAGERYTNECLQ 180
Salmo salar             LQRGNKAKRLNYARKHRKWGSEKWQQVLWTDESKVLIFGCSRRQFVHRRAGELYTNECLQ 180
                        * *************** ** *************  ********** ***** *******

Acipenser sturio        ATVKHGGGSLQVWGCISVNGVGDLVRINGLLNAEKYRQILIHHAIPSGRHLIGPKFILQH 240
Acipenser naccarii-1    ATVKHGGGSLQVWGCISANGVGDLVRINGLLNAEKHRQILIHHAIPSGRHLIGPKFILQH 240
Acipenser naccarii-2    ATVKHGGGSLQVWGCISANGVGDLVRMNGLLNAEKYRQILIHHAVPSGRHLIGPKCILQH 240
Salmo salar             ATVKHGGGSLQVWGCISANGVGDLVRINGLLNAEKYRQILFHHAIPSGRHLIGPKCILPH 240
                        ************************** ******** **** *** ********** ** *

Acipenser sturio        DNDPKHTAKVIKNYLQRKEEQGVLEVMVWPPQSPDLNIIESVWDYMKREKQLRLPKSTEE 300
Acipenser naccarii-1    DNDPKHTAKVIKNYLQRKEEQGVLEVMVWPPQSPDLNIIESVWDYMKREKQLRLPKSTEE 300
Acipenser naccarii-2    DNDPKHTAKVIKNYLQRKEEQGVLEVMVWPPQSPDLNIIKSFWDYMKREKQLRLPKSTEE 300
Salmo salar             DNDPKHTAKVIKNYLQRKEEQRVLEVMVWPPQSPDLNIIESVWDYMKREKHLRLPKSTEE 300
                        ******************** ****************** * ****** ********

Acipenser sturio        LWLFLLDVWANLPAEFLQKLCASVPGRIDAVLKAKGGHTKY 341
Acipenser naccarii-1    LWLVLQDVWANLPAEFLQKLCASVPRRIDAVLKAKGGHTKY 341
Acipenser naccarii-2    LWLVLQDVWTNLPAEFLQKLCPSVPRRIDAVLKAKGGHTKY 341
Salmo salar             LWLVLQDVWATLPAEFLQKRCASVPIRIDAVLKAKGGHTKY 341
                        *** * ***  ******** * *** ***************
```

Fig. 2. Alignment of the amino acid sequence of the transposase of the *Tana*1 element in *A. naccarii*, *A. sturio* (most distant species to *A. naccarii* within the Acipenseridae; Krieger et al., 2008) and Atlantic salmon *Salmo salar* (GenBank AGKD01009409). Included in the sequence are the two Helix-Turn-Helix motif (HTH) motifs consisting of three α-helices each (dark grey), the GPPR sequence (light grey), the nuclear localization signal NLS (underlined) and the DDE motif (red) within the catalytic domain. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

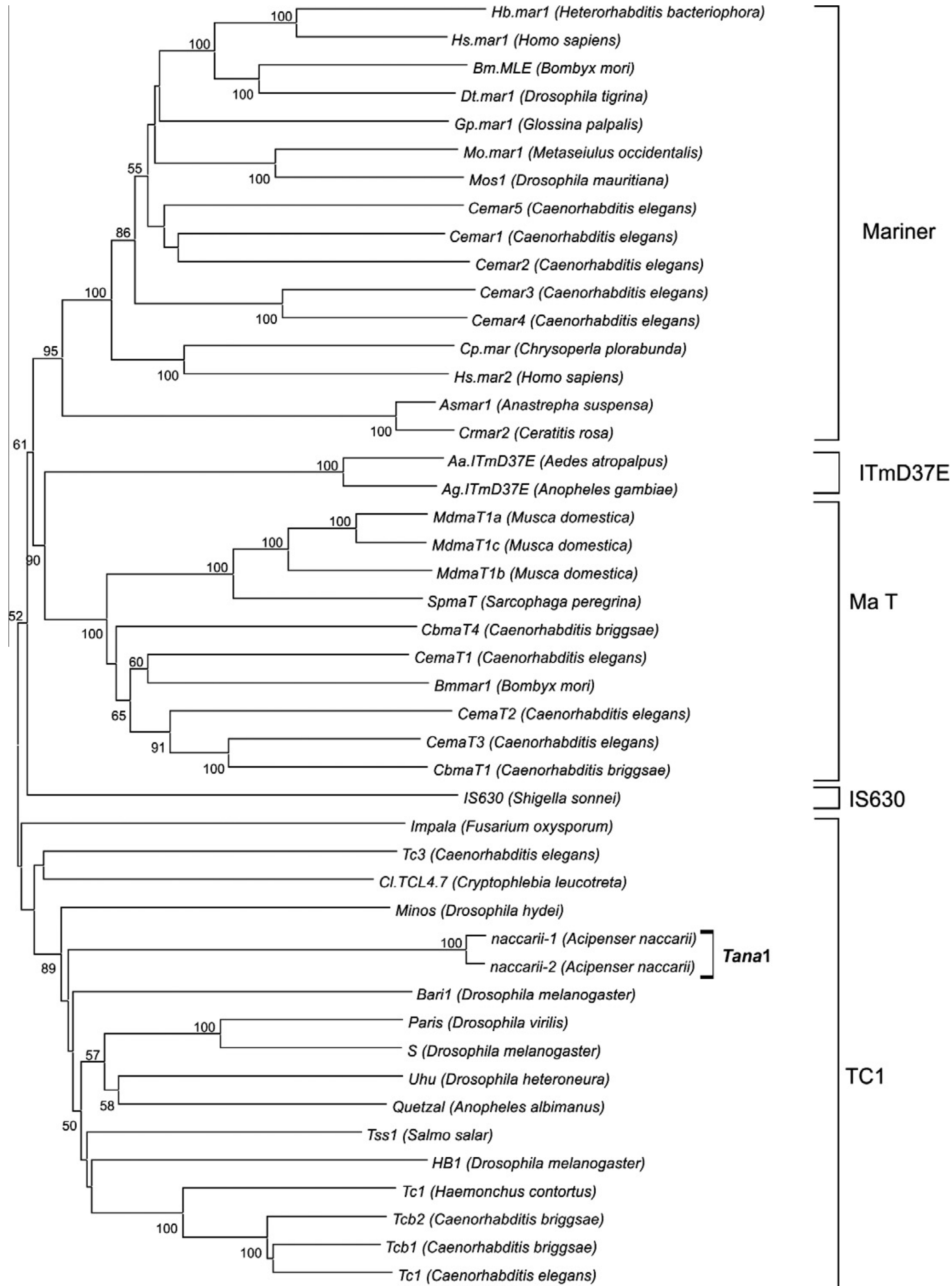


Fig. 3. Neighbor Joining tree showing the relationship of the *Tana*1 element with other members of the *Tc*1/*mariner* superfamily, including the following families: *Tc*1, *mariner*, *Ma T*, IS630 and ITmD37E. The alignment used to generate the tree is available from EMBL at ALIGN_000448. Bootstrap values >50% (1000 replications) are shown above the nodes.

3.4. Presence of Tana1 in other sturgeon species

The *Tana*1 element was detected in all sturgeon species tested in our study, including three genera of the Acipenseridae: *Acipenser* (*A. baerii*, *A. brevirostrum*, *A. fulvescens*, *A. gueldenstaedtii*, *A. mikadoi*, *A. nudiventris*, *A. ruthenus*, *A. stellatus*, *A. sturio and A. transmontanus*), *Huso* (*H. huso*) and *Scaphirhynchus* (*S. albus*).

Total length of the element varied between 1475 and 1594 bp (Table 2). As in *A. naccarii*, the copies isolated from all species presented long inverted repeats of 210 bp, with the exception of the ones cloned from *A. baerii*, *A. brevirostrum* and *A. fulvescens*, in which short deletions or insertions were found. In all species, the second inverted repeat partially overlapped with the coding region. Total length of the transposase ranged from 910 to 1030 bp. All transposases presented all the *Tc*1-like transposon motifs observed in the *A. naccarii Tana*1, with the typical DDE catalytic domain found in all species expect *A. baerii* (DDK), *A. mikadoi* (DDK) and *A. ruthenus* (DDG). The only exception was the sequence isolated from *A. gueldenstaedtii*, in which the third motif of the catalytic domain was missing. In the copies isolated from the rest of

Table 1
Estimation of number of copies of the *Tana*1 element in *A. naccarii* using the Comparative Quantitation method, including the proportion of number of copies between target (*Tana*1) and reference (microsatellite AfuG63).

Tana1 region	Target/reference (control)	Target/reference (samples)	Mean
Region 1 (120 bp)	0.93 ± 0.11	413 ± 87	382 ± 28
Region 2 (130 bp)	1.05 ± 0.08	351 ± 41	

species, the spacing between the second and third motif of the catalytic domain was 38 amino acids. As in the case of *A. naccarii*, the sequences from *A. nudiventris*, *A. stellatus* and *A. sturio* were intact and complete, while the rest of sequences presented stop codons or insertions/deletions causing sequence frameshifts.

Sequence identity between the different *Tana*1 elements ranged between 91.5% and 99.4%, with similar values found between all species pairs. A dendrogram constructed using Neighbor Joining (Fig. 4) showed that *Tana*1 elements did not cluster according to sturgeon phylogeny (for a revision see Krieger et al., 2008) and the phylogenetic trees of *Tana*1 and hosts were discordant. This was evidenced by one *A. naccarii* sequence clustering with *A. sturio*, despite the latter being the most distant species to *A. naccarii* within the Acipenseridae (Krieger et al., 2008).

3.5. Presence of Tana1 in other species

A BLAST search using *A. naccarii Tana*1 as query showed a high genetic similarity (91%) with a contig of Atlantic salmon *Salmo salar* obtained by shotgun sequencing (GenBank AGKD01009409). In addition to other Atlantic salmon contigs, *Tana*1 also matched with sea bass *Dicentrarchus labrax* (86% identity) and Nile tilapia *Oreochromis niloticus* (75% identity) contigs. A 66% identity match was found with a rainbow trout *Oncorhynchus mykiss* EST (Expressed Sequence Tag) sequence. Accordingly, the Atlantic salmon sequence appeared as the most similar to the Acipenseridae *Tana*1 cluster in a dendrogram constructed using Neighbor Joining (Fig 4). While all matches found in GenBank showed a significant

Table 2
Description of the *Tana*1 element in all sturgeon species including total length of the element, length of the first and second inverted repeats, length of the transposase, presence of start codon (ATG), stop codon and insertions/deletions causing sequence frameshift and catalytic domain.

Species	Total bp	First TIR	Second TIR	Transposase bp	ATG	Stop codon	Frame shift	DDE
A. naccarii-1	1588	210	210	1023	Yes	No	No	DDE
A. naccarii-2	1588	210	210	1023	Yes	No	No	DDK
A. baerii	1570	202	210	1013	Yes	No	Yes	DDK
A. brevirostrum	1594	210	219	1030	Yes	Yes	Yes	DDE
A. fulvescens	1595	210	217	1030	Yes	No	Yes	DDE
A. gueldenstaedtii	1475	210	210	910	No	Yes	Yes	DD_
A. mikadoi	1588	210	210	1021	Yes	Yes	Yes	DDK
A. nudiventris	1588	210	210	1023	Yes	No	No	DDE
A. ruthenus	1573	210	210	1008	Yes	Yes	No	DDG
A. stellatus	1586	210	210	1023	Yes	No	No	DDE
A. sturio	1588	210	210	1023	Yes	No	No	DDE
A. transmontanus	1586	210	210	1021	Yes	No	Yes	DDE
H. huso	1559	210	210	997	Yes	No	Yes	DDE
S. albus	1579	210	210	1014	Yes	Yes	Yes	DDE

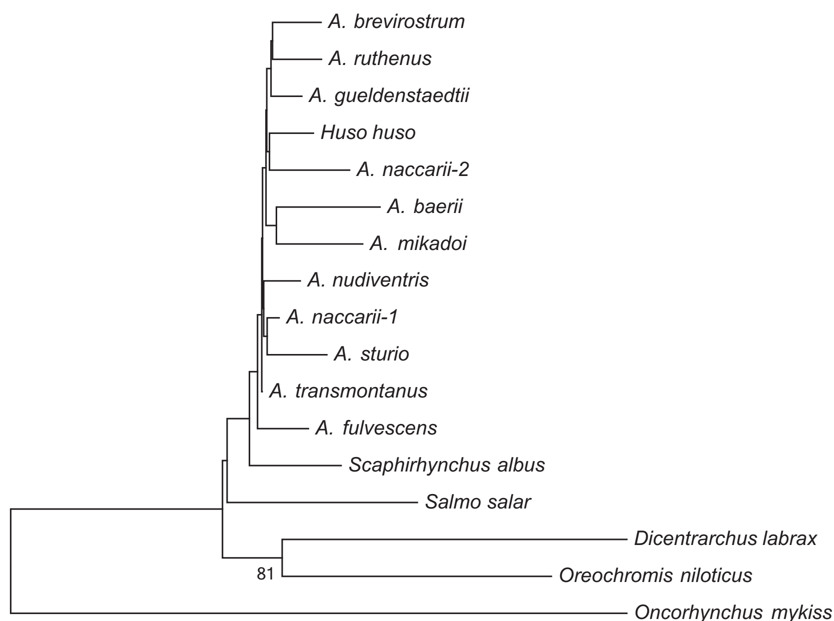


Fig. 4. Neighbor Joining tree showing the relationship of all sturgeon *Tana*1 elements. Also inclused are the most similar sequences found in GenBank including Atlantic salmon *Salmo salar* (AGKD01009409), sea basss *Dicentrarchus labrax* (CABK01013436), Nile tilapia *Oreochromis niloticus* (AERX01033129) and rainbow trout *Oncorhynchus mykiss* (FP324215).

similarity with *Tc*1-like transposase elements, all were defective and presented stop codons and deletions/insertions, with the exception of the Atlantic salmon sequence AGKD01009409. This sequence also included inverted repeats flanking the transposase, which were absent in the rest of sequences. The structure of the element found in Atlantic salmon was highly similar to the structure of *Tana1*, with two long inverted repeats of 210 bp, a 212 bp non-coding region and a partial overlap between the second TIR and the transposase. The only difference was the presence of three insertions in the first TIR in Atlantic salmon, with a total length of 232 bp. Finally, Atlantic salmon also presented the characteristic DDE motif within the catalytic domain, with 38 amino acids between the second aspartic acid and the glutamic acid.

3.6. Estimation of Tana1 copy number

The Comparative Quantitation method was used to estimate the number of copies of the *Tana*1 element in *A. naccarii*. The copy number ratio between target (the transposon) and reference (microsatellite AfuG63) was 413 ± 87 for region 1 and 351 ± 41 for region 2 (Table 1), which indicates that the transposon has roughly about 400 times more copies than the microsatellite, with a mean of 382 ± 28. While most microsatellite loci in *A. naccarii* have four copies per genome, AfuG63 showed a maximum of two alleles in over 50 animals genotyped (Congiu et al., 2011), which suggests that AfuG63 is a diploidized locus as hypothesized by Welsh et al. (2003) in lake sturgeon *A. fulvescens* and green sturgeon *A. medirostris*. Accordingly, we calculated the number of copies of *Tana*1 to range between 764 (382×2) in case of AfuG63 being a diploid locus and 1528 (382×4) in case of being tetraploid.

3.7. Cytogenetic localization of Tana1

In all the 10 metaphase plates of *A. naccarii*, the hybridization signal with the *Tana*1 transposon probe was clearly visible (Fig. 5). The hybridization spots were wide spread in a large number of chromosomes, which suggests an extensive distribution of the *Tana*1 element. In all plates, the *Tana*1 transposon appeared interspersed across all the chromosome regions from centromers to telomers. Hybridization signals were found in acrocentric, metacentric, sub-metacentric and micro chromosomes.

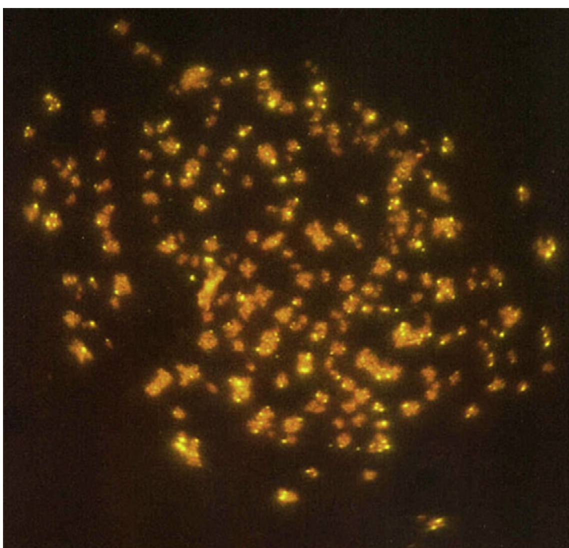


Fig. 5. Metaphase plate of *A. naccarii*. Chromosomes are evidenced in red, fluorescence in situ hybridization in yellow.

4. Discussion

4.1. Description of a new Tc1-like transposable element

We describe for the first time a new *Tc*1-like element isolated in sturgeons. The *Tana*1 element presents all the classic characteristics of *Tc*1-like elements (Plasterk et al., 1999), with a sequence of the full-length element of 1588 nucleotides with long terminal repeats of 210 nucleotides and a single transposase coding region corresponding to 341 amino acids that includes all the functional domains of *Tc*1-like transposons (the two HTH motifs, the GPPR sequence, the nuclear localization signal NLS and the DDE motif). The total length of the element is similar to most *Tc*1-like elements (close to 1.6 kb; Radice et al., 1994) and so is the length of the transposase (around 340 amino acids in all species with the exception of the *Tes1* element in hagfish, which is 407 amino acids long; Heierhorst et al., 1992). Long inverted repeats have also been described in other fish elements including *Tdr*1 in zebrafish *Danio rerio* (208 bp; Izsvak et al., 1995) and *Tsn*1 in lake trout *Salvelinus namaycush* (225 bp; Reed, 1999), while short inverted repeats are known in other fish elements (*Tbr*1 in zebrafish, *Tss*1–2 and 1–3 in Atlantic salmon; Radice et al., 1994).

On the other hand, the *Tana*1 element presents some unique characteristics not reported in other *Tc*1-like elements. Firstly, *Tana*1 shows a partial overlap of 70 bp between the transposase and the second TIR (at the 3′ extremity of the transposase), which means that the stop codon of the transposase is found within the second inverted repeat. In all other *Tc*1-like elements, the TIRs flank the transposase without overlapping as there is always a non-coding region separating the transposase and the TIRs. Secondly, the spacing of the DDE motif within the catalytic domain in *Tana*1 is unique, with 38 amino acids (DD38E) separating the second aspartic acid and the glutamic residues. The atypical spacing was observed in all sturgeon *Tana*1 sequences and differed from the rest of *Tc*1/*mariner* elements by a four amino acid (EQGV) insertion (Fig. 2).

Excluding mutations in defective copies, all *Tc*1-like and most *mariner* elements contain a characteristic 34 amino acid spacing (DD34E and DD34D; Capy et al., 1996), although *IS* elements with a distance of 37 and 39 amino acids have been described in insects and plants (Shao and Tu, 2001). The DDE signature and its spacing seems to be highly conserved even across class I and class II elements, with the number of amino acids between the last two residues ranging between 35 and 36 amino acids in LTR-retrotransposons (Capy et al., 1996). While the presence of a common DDE(D) catalytic motif and a similar spacing of residues points to a common ancestor for all class I and II transposable elements, the variability observed in spacing across subfamilies suggest that the presence of the DDE motif is more essential than its spacing. The catalytic activity of DDE motifs with longer spacing such as *Tana*1 should be intact as shown in *IS* elements with a distance of 37 and 39 amino acids (Shao and Tu, 2001), which were found in complete non-defective and highly similar copies suggestive of recent transposition.

Phylogenetic analysis showed that the *Tana*1 element belongs to the *Tc*1 family (Fig. 3). However, the high genetic distance with the rest of *Tc*1 elements and the presence of a unique DDE spacing (DD38E) suggests that *Tana*1 constitutes a new subfamily within the *Tc*1 family. The copies isolated from 9 out of the 13 sturgeon species investigated in our study presented the characteristic DDE motif, while 3 copies (including one of the two *A. naccarii* sequences) presented a change in the third residue of the DDE motif, a lysine in place of glutamic, turning it into a DDK motif. Using site-directed mutations in *mariner* elements, Lohe et al. (1997) showed that these three amino acids are essential for all catalytic activities

and that a change in the third residue completely obliterates transposase function. Transposase inactivation through changes in the DDE motif seems to be associated to spatial constraints (the presence of the correct amino acid residue is required for the precise spatial position within the transposase fold) rather than the loss of catalytic activity (Plasterk et al., 1999). While the change from DDE to DDK might render that particular transposase defective, transposition could still occur mediated by a tranposase protein encoded by a different source. When a given TE becomes non-autonomous (i.e. loses the ability to synthesize functional transposase), the element can continue to jump to different locations within the genome as long as other autonomous TEs have an intact gene that encodes and produces an active transposase.

Interestingly, the *Tana*1 element was also identified in the genome of Atlantic salmon *Salmo salar* following a search in GenBank. While the element has not yet been described in Atlantic salmon, the high similarity found in terms of sequence identity and structure, including the unique DD38E spacing, suggests that *Tana*1 is not exclusive of sturgeons but also present in Atlantic salmon.

4.2. Tana1 is complete and potentially active

To date, only one *Tc*1-like element in fish has been demonstrated to be functionally active, the *Passport* transposon isolated from the European plaice *Pleuronectes platessa* (Clark et al., 2009). Most *Tc*1-like transposons identified in fish are defective in that they contain either insertions/deletions that result in sequence frameshifts or point mutations that result in a premature stop codon or a nonsense codon in the transcribed mRNA. For instance, Radice et al. (1994) characterized five TEs in fish, one from zebrafish *Danio rerio*, one from rainbow trout *Oncorhynchus mykiss* and three from Atlantic salmon *Salmo salar*, all of which contained corrupted reading frames. Similarly, one full length and two partial-length copies of the *Tsn*1 transposon in lake trout *Salvelinus namaycush* contained stop codon mutations and were no longer functional (Reed, 1999).

By contrast, the *Tana*1 element was found intact and complete in the genome of *A. naccarii*, *A. sturio*, *A. stellatus* and *A. nudiventris*. The transposase presented no internal stop codons or frameshift mutations and the long inverted repeats were completely identical. This is suggestive of a current or recent activity of *Tana*1, since non-functional transposons are expected to rapidly accumulate random mutations not only in the transposase but also in the inverted repeats, disrupting their complementarity. For the rest of species, sequences showed either stop codons or sequence frameshifts but it should be noted that only one single colony per species was sequenced. Coupled with the integrity of the element, a recent activity of this transposon is sustained by the very high similarity observed among the different copies of *Tana*1 isolated within and across species and by the presence of the entire element in the sturgeon transcriptome. Although further assays should be conducted to test if the catalytic activity of the DDE motif is still intact despite its unique spacing, we believe that there is sufficient evidence to consider the element as putatively active.

Finally, using a relative quantification Real-Time PCR approach, we estimated the number of copies of *Tana*1 in *A. naccarii* to range between 764 and 1568 copies. Similar copy numbers were estimated by Leaver (2001) using dot blotting in three *Tc*1-like transposons, *Passport* in the European plaice *Pleuronectes platessa* (300 copies), *SSTN* in Atlantic salmon *Salmo salar* (1200 copies) and *RTTN* in the frog *Rana temporaria* (500 copies). The localization of *Tana*1 on Adriatic sturgeon chromosomes showed a dispersed distribution. A similar pattern has been observed for other *Tc*1-like transposons mapped in fish genomes (Capriglione et al., 2002; Teixeira et al., 2009). On the other hand, compartmentalization of transposable elements has been observed in species with reduced genome size such as the pufferfish *Tetraodon nigroviridis* (DaSilva et al., 2002), which might be the result of a process of containment of transposable elements into the heterochromatic regions removing them from euchromatin. This is in contrast with most vertebrate genomes, where transposable elements are scattered in the euchromatin as well as in the heterochromatin (DaSilva et al., 2002). According to this trend, the widespread distribution of *Tana*1 observed in sturgeons might be expected since sturgeons have large genome sizes (Ludwig et al., 2001), with an estimated genome size of 3.06×10^9 bp for *A. naccarii* (Wuertz et al., 2006).

4.3. Ancient vertical or recent horizontal transmission?

Evidence that TE families include representatives in more than one phylum has been interpreted by assuming ancestral relatively conserved TEs with particular evolutionary characteristics or postulating recent horizontal transmission between species (Capy et al., 1994a,b). After comparison of 17 *Tc*1-like transposons from species representing three phyla (nematodes, arthropods and chordates) in a highly conserved transposon family, Radice et al. (1994) concluded that the highly diverged sequences found were consistent with an ancient origin of the family possibly occurring before the divergence of the major phyla, with no need to invoke horizontal transmission. This is in contrast with the existence of virtually identical *Tc*1/*mariner* elements in the genomes of insect species that are thought to have diverged over 250 million years ago (Robertson and Lampe, 1995). A similar pattern has been found in *P* transposable elements between insect species (Daniels et al., 1990), suggestive of recent spread from one species to the other by horizontal transmission.

The presence of *Tana1* in all the 13 sturgeon species investigated in our study (including the three genera of the Acipenseridae: *Acipenser*, *Huso*, *Scaphirhynchus*, which are presently distributed in different areas of Nearctic and Palearctic zoogeographical regions; Bemis and Kynard, 1997) strongly suggests vertical transmission as the more likely process of dissemination of *Tana*1 in this group of living fossils. The ubiquity of *Tana*1 points to an ancient origin of the element that existed in the common ancestor to all present sturgeons. At that point, *Tana*1 might have contributed to the evolution of the particular genome of sturgeons (e.g. fragmentation into a large number of chromosomes including many microchromosomes). However, under the hypothesis of ancient origin a high divergence at the DNA level is expected between present-day sequences (Radice et al., 1994), which is not observed in our data. For example, high values of sequence identity are found between *A. naccarii* and the most distant species within the *Acipenser* group (*A. sturio*), unexpected for two species that diverged 170 million years ago (Krieger et al., 2008). One plausible explanation could be the assumption of a slowed evolutionary rate in sturgeons, which has been predicted on the basis of life history traits (long generation time, large body size, ectothermy, late maturity and low metabolic rate) and molecular data (limited karyotypic and genetic differences) (Krieger and Fuerst, 2002).

Our data show a general discordance between the phylogenetic trees of *Tana*1 and their sturgeon hosts. However, sequence similarity deserves careful analysis since the inclusion of paralogous copies (divergent copies of the original element) in the analysis could produce a mis-match between the TE tree and the host tree (Loreto et al., 2008). Moreover, the occurrence of introgression events has been largely documented in sturgeons (Arefjev, 1997; Ludwig et al., 2003) and this might have redistributed the genetic diversity of *Tana*1 across species as hypothesized for other repeated sequences (Robles et al., 2004), with an effect similar to that caused by horizontal transmission, in which, however, the transfer occurs without reproduction.

True horizontal transmission might have also played a role in the dissemination of *Tana*1 as evidenced by the presence of a complete copy in the genome of Atlantic salmon (GenBank AGKD01009409) with a 91% sequence similarity. The presence, like in this case, of high sequence similarity between TEs of very distantly related species is considered as one of the main evidences of horizontal transmission (Loreto et al., 2008; Schaack et al., 2010). The possibility that the two *Tana*1 forms descent by a common ancestor of sturgeons and salmonids can reasonably be excluded by the fact that the evolutionary rate of salmonids can not explain this degree of conservation.

Vertical and horizontal transmission are not mutually exclusive and may have concurred in shaping *Tana*1 evolution. A more extensive investigation on the sequence variability of *Tana*1 expanding the number of species studied should be conducted in order to explore the alternative evolutionary scenarios proposed.

Acknowledgments

We are indebted to Dr. Anna Forlani for the valuable help in the preliminary phase of the transposon characterization. We thank Prof. Francesco Fontana and Dr. Massimo Lanfredi for providing materials, facilities and technical support for the FISH experiments. This work was supported by a University of Padova Grant CPDA087543/08 to Leonardo Congiu.

References

Arefjev, V.A., 1997. Cytogenetics of interploid hybridization of sturgeons. J. Appl. Ichthyol. 15, 27.

Bemis, W.E., Kynard, B., 1997. Sturgeon rivers: an introduction to acipenseriform biogeography and life history. Environ. Biol. Fish. 48, 167–183.

Birstein, V.J., 2002. Sturgeon species and hybrids: can hybrids produce caviar? Environ. Pol. Law 32, 210–214.

Birstein, V.J., DeSalle, R., 1998. Molecular phylogeny of the Acipenserinae. Mol. Phylogenet. Evol. 9, 141–155.

Boscari, E., Barbisan, F., Congiu, L., 2011. Inheritance pattern of microsatellite loci in the polyploid Adriatic sturgeon (*Acipenser naccarii*). Aquaculture 321, 223–229.

Bowen, N.J., Jordan, K., 2002. Transposable elements and the evolution of eukaryotic complexity. Curr. Issues Mol. Biol. 4, 65–76.

Capriglione, T., Odierna, G., Caputo, V., Canapa, A., Olmo, E., 2002. Characterization of a *Tc*1-like transposon in the Antarctic ice-fish, *Chionodraco hamatus*. Gene 295, 193–198.

Capy, P., Anxolabehere, D., Langin, T., 1994a. The strange phylogenies of transposable elements: are horizontal transfers the only explanation? Trends Genet. 10, 7–12.

Capy, P., Langin, T., Bigot, Y., Brunet, F., Daboussi, M.J., Periquet, G., David, J.R., Hartl, D.L., 1994b. Horizontal transmission *versus* ancient origin: *mariner* in the witness box. Genetica 93, 161–170.

Capy, P., Vitalis, R., Langin, T., Higuet, D., Bazin, C., 1996. Relationships between transposable elements based upon the integrase–transposase domains: is there a common ancestor? J. Mol. Evol. 42, 359–368.

Clark, K.J., Carlson, D.F., Leaver, M.J., Foster, L.K., Fahrenkrug, S.C., 2009. *Passport*, a native *Tc*1 transposon from flatfish, is functionally active in vertebrate cells. Nucl. Acids Res. 37, 1239–1247.

Claudianos, C., Brownlie, J., Russell, R., Oakeshott, J., Whyard, S., 2002. *MaT* – a clade of transposons intermediate between *mariner* and *Tc*1. Mol. Biol. Evol. 19, 2101–2109.

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, e18249.

Daniels, S.B., Peterson, K.R., Strausbaugh, L.D., Kidwell, M.G., Chovnik, A., 1990. Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. Genetics 124, 339–355.

DaSilva, C., Hadji, H., Ozouf-Costaz, C., Nicaud, S., Jaillon, O., Weissenbach, J., Crollius, H.R., 2002. Remarkable compartmentalization of transposable elements and pseudogenes in the heterochromatin of the *Tetraodon nigroviridis* genome. Proc. Natl. Acad. Sci. USA 99, 13636–13641.

Doolittle, W.F., Sapienza, C., 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284, 601–603.

Feschotte, C., Pritham, E.J., 2007. DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. 41, 331–368.

Feschotte, C., Jiang, N., Wessler, S.R., 2002. Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet. 3, 329–341.

Fontana, F., Rossi, R., Lanfredi, M., Arlati, G., Bronzi, P., 1997. Cytogenetic characterization of cell lines from three sturgeon species. Caryologia 50, 91–95.

Fontana, F., Rossi, R., Lanfredi, M., Arlati, G., Bronzi, P., 1999. Chromosome banding in sturgeons. J. Appl. Ichthyol. 15, 9–11.

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.

Heierhorst, J., Lederis, K., Richter, D., 1992. Presence of a member of the *Tc*1-like transposon family from nematodes and *Drosophila* within the vasoticin gene of a primitive vertebrate, the Pacific hagfish *Eptatretus stouti*. Proc. Natl. Acad. Sci. USA 89, 6798–6802.

Ikeda, R., Kokubu, C., Yusa, K., Keng, V.W., Horie, K., Takeda, J., 2007. *Sleeping Beauty* transposase has an affinity for heterochromatin conformation. Mol. Cell. Biol. 27, 1665–1676.

Ivics, Z., Izsvak, Z., Minter, A., Hackett, P.B., 1996. Identification of functional domains and evolution of *Tc*1-like transposable elements. Proc. Natl. Acad. Sci. USA 93, 5008–5013.

Izsvak, Z., Ivics, Z., Hackett, P.B., 1995. Characterization of a *Tc*1-like transposable element in zebra fish (*Danio rerio*). Mol. Gen. Genet. 247, 312–322.

Kidwell, M.G., 2002. Transposable elements and the evolution of the genome size in eukaryotes. Genetica 115, 49–63.

Kidwell, M.G., Lisch, D.R., 2000. Transposable elements and host genome evolution. Trends Ecol. Evol. 15, 95–99.

Kidwell, M.G., Lisch, D.R., 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 55, 1–24.

Krieger, J., Fuerst, P.A., 2002. Evidence for a slowed rate of molecular evolution in the order Acipenseriformes. Mol. Biol. Evol. 19, 891–897.

Krieger, J., Hett, A.K., Fuerst, P.A., Artyukhin, E., Ludwig, A., 2008. The molecular phylogeny of the order Acipenseriformes revisited. J. Appl. Ichthyol. 24, 36–45.

Lanfredi, M., Congiu, L., Garrido-Ramos, M.A., de la Herran, R., Leis, M., Chicca, M., Rossi, R., Tagliavini, J., Ruiz-Rejon, C., Ruiz-Rejon, M., Fontana, F., 2001. Chromosomal location and evolution of satellite DNA family in seven sturgeon species. Chromosome Res. 9, 47–52.

Leaver, M.J., 2001. A family of *Tc*1-like transposons from the genomes of fish and frogs: evidence from horizontal transmission. Gene 271, 203–214.

Li, H. et al., 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25, 2078–2079.

Lohe, A.R., De Aguilar, D., Hartl, D.L., 1997. Mutations in the mariner transposase: the D,D(35)E consensus sequence is nonfunctional. Proc. Natl. Acad. Sci. USA 94, 1293–1297.

Loreto, E.L.S., Carareto, C.M.A., Capy, P., 2008. Revisiting horizontal transfer of transposable elements in *Drosophila*. Heredity 100, 545–554.

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.

Ludwig, A., Congiu, L., Pitra, C., Fickel, J., Gessner, J., Fontana, F., Patarnello, T., Zane, L., 2003. Nonconcordant evolutionary history of maternal and paternal lineages in Adriatic sturgeon. Mol. Ecol. 12, 3253–3464.

Miskey, C., Izsvak, Z., Kawakami, K., Ivics, Z., 2005. DNA transposons in vertebrate functional genomics. Cell. Mol. Life Sci. 62, 629–641.

Ning, Z., Cox, A.J., Mullikin, J.C., 2001. SSAHA: a fast search method for large DNA databases. Genome Res. 11, 1725–1729.

Orgel, L.E., Crick, F.H., 1980. Selfish DNA: the ultimate parasite. Nature 284, 604–607.

Plasterk, R.H.A., Izsvak, Z., Ivics, Z., 1999. Resident aliens: the *Tc*1/mariner superfamily of transposable elements. Trends Genet. 15, 326–332.

Radice, A.D., Bugaj, B., Fitch, D.H.A., Emmons, S.W., 1994. Widespread occurrence of the *Tc*1 transposon family: *Tc*1-like transposons from teleost fish. Mol. Genet. Genom. 6, 606–612.

Reed, K.M., 1999. *Tc*1-like transposable elements in the genome of lake trout (*Salvelinus namaycush*). Mar. Biotechnol. 1, 60–67.

Robertson, H.M., Lampe, D.J., 1995. Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. Mol. Biol. Evol. 12, 850–862.

Robles, F., de la Herran, R., Ludwig, A., Ruiz Rejon, C., Ruiz Rejon, M., Garrido-Ramos, M.A., 2004. Evolution of ancient satellite DNA in sturgeon genomes. Gene 338, 133–142.

Rochard, E., Williot, P., Castelnaud, G., LePage, M., 1991. Elements de systematique et de biologie des populations sauvages d'esturgeons. In: Williot, P. (Ed.), Acipenser. Actes du Premier Colloque International sur l'Esturgeon. CEMAGREF Publications, Paris, pp. 475–507.

Schaack, S., Gilbert, G., Feschotte, C., 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol. Evol. 35, 537–546.

Schnable, P.S. et al., 2009. The B73 maize genome: complexity, diversity and dynamics. Science 32, 1112–1115.

Shan, X., Liu, Z., Dong, Z., Wang, Y., Chen, Y., Lin, X., Long, L., Han, F., Dong, Y., Liu, B., 2005. Mobilization of the active MITE transposons *mPing* and *Pong* in rice by introgression from wild rice (*Zizania latifolia*). Mol. Biol. Ecol. 23, 976–990.

Shao, H., Tu, Z., 2001. Expanding the diversity of the *IS630-Tc1-mariner* superfamily. Discovery of a unique DD37E transposon and reclassification of DD37 and DD39 transposons. Genetics 159, 1103–1115.

Slotkin, R.K., Martienssen, R., 2007. Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8, 272–285.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731-273.

Teixeira, W.G., Ferreira, I.A., Cabral-de-Mello, D.C., Mazzuchelli, J., Valente, G.T., Pinhal, D., Poletto, A.B., Venere, P.C., Martins, C., 2009. Organization of repeated DNA elements in the genome of the cichlid fish *Cichla kelberi* and its contributions to the knowledge of fish genomes. Cytogenet. Genome Res. 125, 224–234.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 25, 4876–4882.

Welsh, A.B., Blumberg, M., May, B., 2003. Identification of microsatellite loci in lake sturgeon *Acipenser fulvescens*, and their variability in green sturgeon, *A. medirostris*. Mol. Ecol. Notes 3, 47–55.

Wessler, S.R., 2006. Transposable elements and the evolution of eukaryotic genomes. Proc. Natl. Acad. Sci. USA 103, 17600–17601.

Wuertz, S., Gaillard, S., Barbisan, F., Carle, S., Congiu, L., Forlani, A., Aubert, J., Kirschbaum, F., Tosi, E., Zane, L., Grillasca, J.P., 2006. Extensive screening of sturgeon genomes by random screening techniques revealed no sex-specific marker. Aquaculture 258, 685–688.

Yuan, Y.W., Wessler, S.R., 2011. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. Proc. Natl. Acad. Sci. USA 108, 7884–7889.

6. RIASSUNTO

Genetica della conservazione dello storione Cobice (*Acipenser naccarii*)

Introduzione

La conservazione *ex situ* rappresenta una delle ultime chance di sopravvivenza per tutte le specie considerate a rischio di estinzione (Russello e Amato, 2004). Il presente lavoro nasce proprio dalla necessità di un piano di incroci *ex situ* a lungo termine specifico per la salvaguardia di una delle specie endemiche del bacino settentrionale dell'Adriatico e di tutti i suoi affluenti.
Il progetto prende in esame la specie tetraploide di storione Cobice (*Acipenser naccarii*), internazionalmente conosciuta come *Adriatic sturgeon* (classe Attinopterigi, superordine Condrostei), attualmente considerata sull'orlo dell'estinzione. L'abbondanza di questa specie è infatti progressivamente diminuita negli ultimi decenni a causa di un eccessivo sfruttamento della pesca, dell'introduzione di specie aliene e del continuo degrado degli ambienti naturali (Ludwig *et al.*, 2003). Dal 1996, lo storione Cobice e tutte le altre specie di storione sono sotto la tutela della IUCN (*International Union for Conservation of Nature*), inserite a diversi livelli di criticità nella Lista Rossa per le specie in via di estinzione. Nonostante dal 1998 tutto ciò che riguarda il commercio di prodotti o individui stessi appartenenti alla specie *A. naccarii* sia regolato anche dalle normative della CITES (*Conservation in International Trade in Endangered Species of Wild Fauna and Flora*), nel Marzo 2010 la classificazione dello storione Adriatico nella Lista Rossa è passata da "*Vulnerable*" a "*Critically Endangered*", livello di criticità antecedente l'estinzione (Congiu *et al.*, 2011), considerando l'assenza di avvistamenti di avannotti imputabili a riproduzioni avvenute in natura.
Con il progetto Life Cobice (2004) ha avuto inizio uno studio volto alla salvaguardia di questa specie in linea con le direttive nazionali e internazionali indicate nella *Ramsar Declaration on Global Conservation*, secondo cui ogni pratica futura finalizzata alla conservazione della specie dovrà essere supportata da solide basi genetiche. Durante il progetto Life Cobice è stato possibile caratterizzare geneticamente l'unico stock (N=50) di sicura origine selvatica (chiamato F0) attualmente esistente, stabulato dal 1977 presso l'Azienza Agricola V.I.P. (Orzinuovi). Dal 1988, presso questa azienda, vengono effettuate periodiche riproduzioni controllate e tutti gli esemplari allevati in Europa discendono direttamente da questo stock (Congiu *et al.*, 2011). Considerando l'esiguo numero di individui F0 (13) ad oggi sopravvissuti la pianificazione di un programma di conservazione *ex situ* a lungo termine, obiettivo del presente progetto, risulta prioritaria ed è strettamente legata alla gestione della variabilità esistente negli stock di F1 ritenuti nei diversi allevamenti.
Lo scopo è quello di garantire la disponibilità di un parco riproduttori (costituito da individui della generazione F1) che rappresenti il più possibile la variabilità esistente nello stock originario e di permettere la pianificazione di campagne di rilascio controllate.
Al fine di proporre un programma di allevamento specifico per lo storione Cobice sono stati considerati anche diversi aspetti legati al livello ploidia della specie, alle caratteristiche biologiche e ai vincoli economici e logistici legati alle pratiche di acquacultura.

Risultati e discussione

Inizialmente l'attenzione è stata rivolta alla determinazione del pattern di segregazione allelica nella specie *A. naccarii.*
Le analisi genetiche nello storione Adriatico, e più in generale negli Acipenseridi, sono complicate da un genoma complesso il cui grado di ploidia è tuttora oggetto di dibattito (Ludwig *et al.*, 2001). La

specie in esame viene considerata tetraploide funzionale e presenta un genoma composto da circa 240 cromosomi (Fontana *et al*., 2008).

La condizione di tetraploidia può essere raggiunta mediante meccanismi di allo- o auto-tetraploidizzazione ai quali vengono associate differenti modalità di segregazione allelica nei gameti, rispettivamente disomica e tetrasomica (Stift *et al*., 2008). Le modalità di segregazione allelica non sono mai state investigate nella specie in esame, inoltre, nell'ottica di un progetto volto alle analisi di parentela, dove i genotipi attesi nella progenie variano al variare del pattern di segregazione, sono un prerequisito fondamentale per l'ottimizzazione di un metodo di allocazione parentale, per la messa a punto di un metodo di generazione di prole virtuale e per la gestione di un piano di incroci a lungo termine.

Le analisi condotte a questo scopo hanno permesso di rigettare l'ipotesi di disomia e ogni altra segregazione di tipo intermedio e di confermare la tetrasomia come pattern più probabile.

Questi risultati, riportati in dettaglio nella corrispondente pubblicazione allegata al lavoro di tesi **(Boscari *et al.,* 2011 - Annex A)**, hanno fornito un importante contributo alla messa a punto del metodo di allocazione applicato durante la fase di caratterizzazione degli stock e ricostruzione dei pedigree. Inoltre, i dati suggeriscono che il raggiungimento dell'attuale livello di ploidia in questa specie sia avvenuto mediante auto-poliploidizzazione a partire da un'unica specie.

Successivamente, ci siamo concentrati su un altro aspetto di interesse generale per gli storioni che riguarda in dettaglio i metodi per l'identificazione di specie e soprattutto di ibridi. Eventi di ibridazione sono comuni in natura e, negli ultimi decenni, l'allevamento di ibridi interspecifici ha visto un notevole incremento grazie soprattutto alla loro precoce maturità sessuale che abbatte i costi di mantenimento e produzione.

In acquacoltura è consuetudine l'allevamento simultaneo di più specie e di ibridi aumentando la possibilità di incorrere in errori nell'identificazione degli individui anche considerando che, ad oggi, non sono disponibili metodi genetici affidabili. La maggior parte degli strumenti disponibili in letteratura si basa sul DNA mitocondriale il quale, essendo ereditato per via matrilineare non può essere impiegato per l'identificazione di ibridi (Ludwig, 2008). Pertanto, la disponibilità di un buon metodo di identificazione è di primaria importanza come strumento di controllo di attività commerciali (Ludwig *et al.*, 2008) ma ha anche una diretta ricaduta conservazionistica volta alla conferma della purezza degli individui destinati all'utilizzo come riproduttori.

In questo lavoro, è stato messo a punto un nuovo metodo per l'identificazione di specie e di ibridi basato sulla caratterizzazione di marcatori nucleari e sulla costruzione di un pannello di primer specie-specifici costruiti su SNP diagnostici identificati sul primo introne del gene RPS7 (*Ribosomal Protein S7*). Inoltre, per l'identificazione di ibridi sono stati incorporati anche dati relativi al d-Loop mitocondriale e al gene codificante per la Vimentina (in collaborazione con dott. Nikolai Mugue , Russian Federal Institute).

Tra i risultati ottenuti, riportati nella pubblicazione **Boscari *et al.* (2014) (Annex B)**, particolarmente rilevante nell'ambito di questo progetto di tesi è l'identificazione del primo marcatore specifico per la specie *A. naccarii*.

Il metodo proposto si è rivelato molto utile nell'identificazione di individui ibridi all'interno degli stock analizzati per la progettazione del piano di incroci.

Ai fini della costruzione del piano di incroci, la variabilità genetica di tutti gli individui F1 disponibili in allevamento è stata stimata caratterizzando la regione di controllo mitocondriale e genotipizzando 7 loci microsatellite. I profili degli individui F1 sono stata quindi confrontati con quelli degli individui F0 (Band-Sharing), tenendo conto del modello di eredità tetrasomica precedentemente osservato

(Boscari *et al.*, 2011). Al fine di migliorare le prestazioni del metodo Band Sharing, utilizzato per l'allocazione parentale in Congiu *et al.* (2011), è stato sviluppato un nuovo strumento per l'analisi del pedigree che consente la stima per ogni individui di un indice di compatibilità per tutte le coppie di putativi parentali. I dettagli del metodo sono riportati nel manoscritto in appendice **(Boscari *et al.*, *Manuscript* - Annex D)**.

Le conseguenze della conduzione di programmi di recupero senza supporti genetici sono stati valutati nel Parco del Ticino (TRP), in Italia, dove in passato erano state eseguite delle campagne di recupero per questa specie. La bassissima eterogeneità osservata in questi stock F1 (Stock_1 con Ni = 137 e Stock_2 con Ni = 116), in parte rilasciati e in parte destinati ad essere utilizzati per le future riproduzioni, suggerisce fortemente la necessità di una completa riorganizzazione degli stock e di un coordinamento degli sforzi condotti dai diversi impianti, prestando particolare attenzione alla conservazione a lungo termine della diversità genetica disponibile. I dettagli di questo lavoro sono riportati nel manoscritto in appendice **(Boscari e Congiu, 2014 - Annex C)**.

Al contrario, i risultati ottenuti dalla caratterizzazione dello stock V.I.P. di 445 individui, ritenuti presso l'azienda V.I.P. dalle riproduzioni condotte negli ultimi 20 anni, hanno mostrato una considerevole variabilità in confronto allo stock parentale. Su 445 individui, 382 sono stati assegnati correttamente a 30 famiglie e, inferendo le relazioni di parentela sulla base della distanza, è stato possibile identificare altre 4 famiglie tra i non allocati.

Per la costruzione del piano di incroci (oggetto del manoscritto in appendice - **Annex D**) viene proposta una strategia in due fasi in cui un primo programma di selezione a breve termine, basato sui 13 individui F0 ad oggi sopravvissuti, è seguito da una seconda serie di incroci tra le famiglie F1. In quest'ultimo caso, le famiglie F1 dello stock V.I.P. sono state opportunamente selezionate per la formazione della "*breeders unit*" dalla quale verranno successivamente selezionate per gli incroci. Anche alcune famiglie dello Stock_1, i cui parentali non erano rappresentati nello stock V.I.P., sono state incluse nella "*breeders unit*".

Nel nostro caso la strategia proposta non è basata sulla scelta di singoli individui da incrociare bensì sulla scelta di combinazioni familiari. Ad ogni famiglia, denominata "*priority family*", è stato dato un ordine di priorità da seguire per la riproduzione. Ogni "*priority family*" è poi incrociata con la "*mating family*" selezionata sulla base della distanza genetica a 24 loci (considerando il profilo cumulativo dei parentali componenti ogni famiglia).

Il piano di incroci è stato simulato in silico. Differenti scenari sono stati testati stimando l'efficienza della trasmissione allelica nella generazione F2 al fine di valutare diverse strategie di scelta delle famiglie da incrociare e il numero ottimale di individui per famiglia da riprodurre.

I risultati indicano che in generale, nel breve termine, la strategia migliore per massimizzare il mantenimento della variabilità genetica presente nello stock F0 è la selezione della "*mating family*" sulla base della distanza genetica e che il numero consigliato di individui da riprodurre per famiglia è tre.

Inoltre, la sostenibilità del piano di incroci in acquacoltura è stata simulata sulla base di vincoli economici e logistici permettendo così la valutazione del supporto economico richiesto per anno.

Sulla base dello spazio disponibile in allevamento e considerando i tassi di mortalità per le diverse fasi di sviluppo il lavoro propone anche il numero di individui da tenere in cattività come riproduttori (considerando di avere circa venti individui per ogni famiglia di 10 anni di età) e da rilasciare.

In conclusione, la strategia qui presentata propone il primo piano di incroci mirato alla gestione coordinata delle azioni intraprese dalle diverse aziende coinvolte nella conservazione di questa

specie, con l'obiettivo finale di raggiungere la costituzione di popolazioni auto-sufficienti idonee al rilascio.
Il metodo proposto può inoltre essere considerato come guida per la salvaguardia anche di altre specie tetraploidi di storione a rischio di estinzione.

7. REFERENCE LIST

Anders PJ, Drauch-Schreier A, Rodzen J, Powell MS, Narum S, Crossman JA (2011). A review of genetic evaluation tools for conservation and management of North American sturgeons: roles, benefits, and limitations. *Journal of Applied Ichthyology*, 27(Suppl.2): 3-11.

Araki H, Cooper B and Blouin MS (2007). Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. Science, 318: 100-103.

Ballou JD and Lacy RC (1995). Identifying genetically important individuals for management of genetic variation in pedigreed populations. In: Ballou JD, Gilpin M, Foose TJ (Eds.), Population management for survival and recovery: analytical methods and strategies in small population conservation. Columbia University Press, New York, pp. 76-111.

Beck BB, Rapaport LG, Stanley Price MR, Wilson AC (1994). Reintroduction of captive-born animals. In: Olney PJS, Mace GM, Feistner ATC (Eds.), Creative conservation: interactive management of wild and captive animals. Chapman and Hall, London, pp. 265-286.

Bemis W and Kynard B (1997). Sturgeon rivers: an introduction to acipenseriforme biogeography and life history. *Environmental Biology and Fishes*, 48: 167-183.

Billard R and Lecointre G (2001). Biology and conservation of sturgeon and paddlefish. *Reviews in Fish Biology and Fisheries*, 10: 355-392.

Birstein VJ and Bemis WE (1997). How many species are there within the genus *Acipenser?* In: Birstein VJ, Waldman JR, Bemis WE (Eds.), Sturgeon Biodiversity and Conservation. Kluwer Academic Publishers, Dordrecht, pp. 157-163.

Birstein VJ and DeSalle R (1998). Molecular phylogeny of Acipenserinae. Molecular Phylogenetic and Evolution, 9: 141-155.

Birstein VJ and Vasil'ev VP (1987). Tetraploid-octoploid relationships and karyological evolution in the order Acipenseriformes (Pisces): karyotypes, nucleoli, and nucleolus-organizer regions in four acipenserid species. *Genetica,* 73: 3-12.

Boscari E, Barbisan F, Congiu L (2011). Inheritance pattern of microsatellite loci in the polyploidy Adriatic sturgeon (*Acipenser naccarii*). *Aquaculture,* 321: 223-229.

Boscari E and Congiu L (2014). The need for genetic support in restocking activities and *ex situ* conservation programs: the case of the Adriatic sturgeon in the Ticino River Park. *Proceedings of the Journal of Applied Ichthyology, Accepted.*

Boscari E, Barmintseva A, Pujolar JM, 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,* Doi:10.1111/1755-0998.12203. *In Press*

Bronzi P, Arlati G, Cataudella S, Rossi R (1994). Sturgeon distribution in Italy. Presentation at the Interantional Conference on Sturgeon Biodiversity and Conservation. The American Museum of Natural History, New York, USA, July 28-30 1994.

Bronzi P, Rosenthal H, Arlati G, Williot P (1999). A brief overview on the status and prospects of sturgeon farming in Western and Central Europe. *Journal of Applied Ichthyology,* 15: 224-227.

Bronzi P, Rosenthal H, Gessner J (2011). Global sturgeon aquaculture production: an overview. *Journal of Applied Ichthyology,*27: 169-175.

Congiu L, Pujolar JM, 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: e18249.

Crow JF and Kimura M (1970). Introduction to population genetics theory. Harper and Row, New York.

DeSalle R and Amato G (2004). The expansion of conservation genetics. *Nature Reviews Genetics,* 5(9): 702-712.

Doukakis P, Babcock EA, Pikitch EK, Sharov AR, Baimukhanov M, Erbulekov S, Bokova Y, Nimato A (2009). Management and recovery options for Ural River beluga sturgeon. *Conservation biology*, 23(3): 769-777.

Doukakis P, Pikitch EK, Rothschild A, DeSalle R, Amato G, Kolokotronis SO (2012). Testing the effectiveness of an international conservation agreement: marketplace forensic and CITES caviar trade regulation. *PLoS One*, 7: e40907.

Dudu A, Suciu R, Paraschiv M, Georgescu SE, Costache M, Berrebi P (2011). Nuclear markers of Danube sturgeons hybridization. *International Journal of Molecular Science,* 12: 6796-6809.

Estoup A, Gharbi K, SanCristobal M, Chevalet C, Haffray P, Guomard R (1998). Parentage assignment using microsatellite in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorynchus mykiss*) hatchery populations. *Canadian Journal of Fisheries and Aquatic Sciences,* 55: 715-725.

Fischer J and Lindenmayer DB (2000). An assessment of the published results of animal relocations. *Biological Conservation,* 96: 1-11.

Fontana F (2002). A cytogenetic approach to the study of taxonomy and evolution in sturgeons. Journal of Applied Ichthyology, 18: 226-233.

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: 1008-1012.

Fontana F, Tagliavini J, Congiu L (2001). Sturgeons genetics and cytogenetics: recent advancements and perspectives. *Genetica*, 111(1-3): 359-373.

Fontana F, Zane L, Pepe A, Congiu L (2007). Polyploidy in Acipenseriformes: cytogenetic and molecular approaches. In Pisano E, Ozouf-Costaz C, Foresti F, Kapoor BG (Eds.), Fish Cytogenetics. Science Publisher, Inc, Enfield, New Hampshire, USA, pp. 285-403.

Fontana F, Congiu L, Mufrak VA, Quattro JM, Smith TIJ, Ware K, Doroshov SI (2008). Evidence of hexaploid karyotype in shortnose sturgeon. *Genome,* 51: 113-119.

Fopp-Bayat D and Luczynski M (2006). Microsatellite DNA polymorphism in sturgeon species and their hybrids reared in Polish aquaculture farms. *Environmental Biotechnology,* 2: 11-19.

Frankham R (2008). Genetic adaptation to captivity in species conservation programs. Molecular ecology, 17: 325-333.

Frankham R, Ballou JD and Briscoe DA (2002). Introduction to conservation genetics. Cambridge university press, Cambridge, UK.

Fraser DJ (2008). How well can captive breeding programs conserve biodiversity? A review of salmonids. *Evolutionary Applications,* 1: 535-586.

Garcia de Leon FJ, Canonne M, Quillet E, Bonhomme F, Chatain B (1998). The application of microsatellite markers to breeding programs in the sea bass, *Dicentrarchus labrax. Aquaculture,* 159: 303-316.

Goldstein PZ, DeSalle R, Amato G, Vogler AP (2000). Conservation genetics at the species boundary. *Conservation Biology,* 14(1): 120-131.

Hayes B, He J, Moen T, Bennewitz J (2006). Use of molecular markers to maximize diversity of founder populations for aquaculture breeding programs. *Aquaculture*, 255:573-578.

Havelka M, Kaspar Vojtech, Hulak Martin, Flajshans M (2011). Sturgeon genetics and cytogenetics: a review related to ploidy levels and interspecific hybridization. *Folia Zoologica*, 60: 93-103.

Hayes B, He J, Moen T, Bennewitz J (2006). Use of molecular markers to maximise diversity of founder populations for aquaculture breeding programs. *Aquaculture,* 255: 573-578.

Hedrick PW (2001). Conservation genetics:where are we now? *Trends in ecology and evolution*, 16(11): 629-636.

Herran R, Robles F, Martinez-Espin E, Lorente JA, Rejon CR, Garrido-Ramos MA, Rejon MR (2004). Genetic identification of western Mediterranean sturgeons and its implication for conservation. *Conservation Genetics,* 5: 545-551.

Honda T, Nomura T, Mukai F (2004). Reduction of inbreeding in commercial females by rotational mating with several sire lines. *Genetics Selection Evolution,* 36: 509-526.

Huang CM and Liao IC (1990). Response to mass selection for growth rate in *Oreochromis niloticus. Aquaculture,* 85: 199-205.

Ireland SC, Beamesderfer RCP, Paragamian VL, Wakkinen VD, Siple JT (2002). Success of hatchery-reared juvenile white sturgeon (*Acipenser transmontanus*) following release in the Kootenai River, Idaho, USA. *Journal of Applied Ichthyology*, 18: 642-650.

Jenneckens I, Meyer JN, Horstgen-Schwark G, May B, Debus L (2001). A fixed allele at microsatellite LS-39 is characteristic for the black caviar producer *Acipenser stellatus. Journal of Applied Ichthyology*, 17: 39-42.

Knapp A, Kitschke C, Von Meibom S (2006). Proceedings of the international sturgeon enforcement workshop to combat illegal trade in caviar. Prepared by TRAFFIC Europe for the European Commission, Brussels, Belgium, Contract No. N_070201/2006/433267/MAR/E2.

Kozfkay CC, Campbell MR, Heindel JA, Baker DJ, Kline P, Powell MS, Flagg T (2008). A genetic evaluation of relatedness for broodstock management of captive, endangered Snake River sockeye salmon, *Oncorhynchus nerka. Conservation Genetics,* 9: 1421-1430.

Krieger J, Hett AK, Fuerst PA, Artyukhin E, Ludwig A (2008). The molecular phylogeny of the order Acipenseriformes revisited. *Journal of Applied Ichthyology,* 24: 36-45.

Lacy RC (2012). Extending pedigree analysis for uncertain parentage and diverse breeding system. *Journal of Heredity*, 103(2): 197-205.

Leus K (2011). Captive breeding and conservation. *Zoology in the Middle East*, 54: 151-158.

Ludwig A (2006). A sturgeon view on conservation genetics. *European Journal of Wildlife Research*, 52: 3-8.

Ludwig A (2008). Identification of Acipenseriformes species in trade. *Journal of Applied Ichthyology*, 24: 2-19.

Ludwig A, Belfiore NM, 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.

Ludwig A, Congiu L, Pitra C, Fickel J, Gessner J, Fontana F, Patarnello T and Zane L (2003). Nonconcordant evolutionary history of maternal and paternal lineages in Adriatic sturgeon. *Molecular Ecology,* 12: 3253-3264.

Lynch M and O'Hely M (2001). Captive breeding and the genetic fitness of natural populations. *Conservation Genetics,* 2: 363-378.

Lyon JP, Todd C, Nicol SJ, MacDonald A, Stoessel D, Ingram BA, Barker RJ, Bradshaw CJA (2012). Reintroduction success of threatened Australian trout cod (*Maccullochella macquariensis*) based on growth and reproduction.

Meirmans PG and Van Tienderen PH (2013). The effects of inheritance in tetraploids on genetic diversity and population divergence. *Heredity,* 110: 131-137.

Montgomery ME, Ballou JD, Nurthen RK, England PR, Briscoe DA, Frankham R (1997). Minimizing kinship in captive breeding programs. *Zoo Biology,* 16: 377-389.

Myers JM, Heggelund PO, Hudson G, Iwamoto RN (2001). Genetics and broodstock management of coho salmon. *Aquaculture*, 197: 43-62.

Nathanailides C, Tsoumani M, Papazogloy A, Paschos I (2002). Hatching time and post-hatch growth in Russian sturgeon *Acipenser gueldenstaedtii. Journal of Applied Ichthyology,* 18:651-654.

Nomura T and Yonezawa K (1996). A comparison of four systems of group mating for avoiding inbreeding. *Genetics Selection Evolution,* 28: 141-159.

Norris AT, Bradley DG, Cunningham EP (2000). Parentage and relatedness determination in farmed Atlantic salmon (*Salmo salar*) using microsatellite markers. *Aquaculture,* 182: 73-83.

O'Reilly PT, Herbinger C, Wright JM (1998). Analysis of parentage determination in Atlantic salmon (*Salmo salar*) using microsatellites. *Animal Genetics,* 29: 363-370.

Peng Z, Ludwig A, Wang D, Diogo W, Wei Q, He S (2007). Age and biogeography of major clades in sturgeons and paddlefishes (Pisces: Acipenseriformes). *Molecular Phylogenetics and Evolution,* 42: 854-862.

Pertoldi C, Bijlsma R, Loeschcke V (2007). Conservation genetics in a globally changing environments: present problems, paradoxes and future challenges. *Biodiversity and conservation*, 16: 4147-4163.

Philippart JC (1995). Is captive breeding an effective solution for the preservation of endemic species? *Biological Conservation,* 72: 281-295.

Plasterk RHA, Izsvak Z, Ivics Z (1999). Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends in Genetics*, 15: 326–332.

Rodzen JA, Famula TR, May B (2004). Estimation of parentage and relatedness in the polyploid white sturgeon (*Acipenser trasnsmontanus*) using a dominant marker approach for duplicated microsatellite loci. *Aquaculture*, 232: 165-182.

Rudnick JA and Lacy RC (2008). The impact of assumptions about founder relationships on the effectiveness of captive breeding strategies. *Conservation Genetics,* 9: 1439-1450.

Russello MA and Amato G (2004). *Ex situ* population management in the absence of pedigree information. *Molecular ecology*, 13: 2829-2840.

Sekino M, Sugaya T, Hara M, Taniguchi N (2004). Relatedness inferred from microsatellite genotypes as a tool for broodstock management of Japanese flounder *Paralichthys olivaceus. Aquaculture,* 233: 163-172.

Shao H, Tu Z (2001). Expanding the diversity of the IS630-Tc1-mariner superfamily. Discovery of a unique DD37E transposon and reclassification of DD37 and DD39 transposons. *Genetics*, 159: 1103–1115.

Snyder NFR, Derricksin SR, Beissinger SR, Wiley JW, Smith TB, Toone WD, Miller B (1996). Limitations of captive breeding in endangered species recovery. *Conservation Biology*, 10: 338-348.

Stift M, Berenos C, Kuperus P, Van Tiederen PH (2008). Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to *Rorippa* (yellow cress) microsatellite data. *Genetics,* 179: 165-182.

Teichert-Coddington DR and Smitherman RO (1988). Lack of response by *Tilapia nilotica* to mass selection for rapid early growth. *Transaction of the American Fisheries Society,* 117: 297-300.

Theodorou K and Couvet D (2010). Genetic management of captive populations: the advantages of circular mating. *Conservation Genetics,* 11: 2289-2297.

Tzika AC, Remy C, Gibson R, Milinkovitch MC (2009). Molecular genetic analysis of a captive-breeding program: the vulnerable endemic Jamaica yellow boa. *Conservation Genetics,* 10: 69-77.

Van Dyke F (2008). Conservation Biology: foundations, concepts, applications. Wheaton College, Illinois, USA. Second Editions, Springer Science and Buisiness Media.

Van Eenennaam AL, Murray JD, Medrano JF (1998). Mitotic analysis of the North American white sturgeon, *Acipenser transmontanus* Richardson (Pisces, Acipenseridae), a fish with a very high chromosome number. *Genome,* 41: 266-271.

Waldam JR and Wirgin II (1997). Status and restoration options for Atlantic sturgeon in North America. *Conservation Biology*, 12: 631-638.

Waldbeiser GC, Wolters WR (1999). Applicatio of polymorphic microsatellite loci in a channel catfish *Ictalurus punctatus* breeding program. *Journal of the World Aquaculture Society*, 30: 256-262.

Wang JH, Wei QW, Zou YC (2011). Conservation strategies for the Chinese sturgeon, *Acipensr sinensis*: an overview on 30 years of practices and future needs. *Journal of Applied Ichthyology*, 27: 176-180.

Wei Q, He J, Yang D, Zheng W, Li L (2004). Status of sturgeon aquaculture and sturgeon trade in China: a review based on two recent nationwide surveys. *Journal of Applied Ichthyology*, 20: 321-332.

Wei QW, Zou Y, Li P, Li L, (2011). Sturgeon aquaculture in China: progress, strategies and prospects assessed on the basis of nation-wide surveys (2007-2009). *Journal of Applied Ichthyology,* 27: 162-168.

Welsh A and May B (2006). Development and standardization of disomic microsatellite markers for lake sturgeon genetic studies. *Journal of Applied Ichthyology,* 22: 337-344.

Wendel JF (2000). Genome evolution in polyploids. *Plant Molecular Biology,* 42: 225-249.

William SE and Hoffman EA (2009). Minimizing genetic adaptation in captive breeding programs: a review. *Biological Conservation,* 142: 2388-2400.

Williot P, Sabeau L, Gessner J, Arlati G, Bronzi P, Gulyas T, Berni P (2001). Sturgeon farming in Western Europe: recent developments and perspectives. *Aquatic Living Resources,* 14: 367-374.

Winding JJ and Kaal L (2008). An effective rotational mating scheme for inbreeding reduction in captive populations illustrated by the rare sheep breed Kempisch Heideschaap. *Animal,* 2(12): 1733-1741.

Witzenberger KA and Hochkirch A (2011). *Ex situ* conservation genetics: a review of molecular studies on the genetic consequencies of captive breeding programmes for endangered animal species. *Biodiversity and Conservation,* 20: 1843-1861.

Woodworth LM, Montgomery ME, Briscoe DA, Frankham R (2002). Rapid genetic deterioration in captive populations: causes and conservation implications. *Conservation Genetics,* 3: 277-288.

Yuan YW and Wessler SR (2011). The catalytic domain of all eukaryotic cutand-paste transposase superfamilies. PNAS USA, 108: 7884–7889.

Zhao N and Chang JB (2006). Microsatellite loci inheritance in the Chinese sturgeon Acipenser sinensis with an analysis of expected gametes ratios in polyploidy organisms. *Journal of Applied Ichthyology*, 22:89-96.

8. ACKNOWLEDGEMENTS

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

First of all, thanks to my supervisor, Leonardo Congiu, for his help, encouragement, supervision and support.

Thanks also to Lorenzo Zane for his help and support.

Many thanks to my external supervisor, Giorgio Bertorelle, for his important advices and suggestions.

I want to thank all the people already cited in the acknowledgement of the papers here reported, in particular Giacinto Giovannini and sons, owners of the V.I.P. plant, for their active collaboration and for providing samples.

Thanks to Nikolai Mugue, Isabelle Dupanloup and Phaedra Doukakis for their collaborations.

I'd like to thank all my laboratory members: Ilaria Marino, Chiara Papetti, Marti Pujolar and Cecilia Agostini.

Thanks to all colleagues and lab-neighbours: Andrea Pilastro, Alessandro Grapputo, Paolo Bronzi, Federica Barbisan, Andrea Benazzo, Riccardo Corradin, Michele Vidotto, Andrea Di Nisio, Alessandro De Vigili, Clelia Gasparini, Alessandro Coppe, Silvia Cattelan, Marta Paterno, Chiara Caruso.

Thanks to my parents.

And of course, a special thanks to Nicola Franchi for supporting me in everything.