GENETIC CHARACTERIZATION OF THE IBERIAN POPULATIONS OF TWO INVASIVE MOLLUSKS: ZEBRA MUSSEL AND ASIATIC CLAM

Luis Peñarrubia Lozano

Universitat de Girona

Doctoral Thesis

Genetic characterization of the Iberian populations of two invasive mollusks: zebra mussel and asiatic clam

Luis Peñarrubia Lozano

2016

Universitat de Girona

Universitat de Girona

Doctoral thesis

Genetic characterization of the Iberian populations of two invasive mollusks:

zebra mussel and asiatic clam

Luis Peñarrubia Lozano

2016

Doctorate program in Experimental Sciences and Sustainability

Thesis supervisors:

Dr. Oriol Vidal Fàbrega

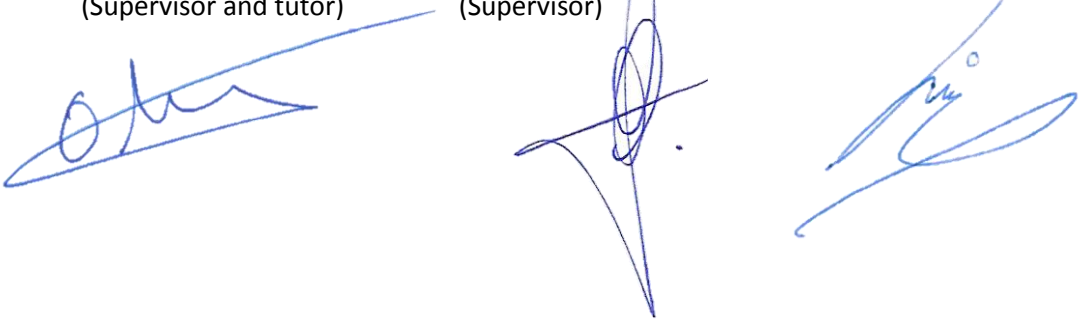
(Supervisor and tutor)

Dr. Jordi Viñas de Puig

(Supervisor)

PhD candidate:

Luis Peñarrubia Lozano



The present thesis contains two sections of additional supplementary materials, including 28 files.

This thesis is submitted in fulfilment of the requirements to obtain the doctoral degree from the Universitat de Girona.

Universitat de Girona

Hereby, Dr. Oriol Vidal Fàbrega and Dr. Jordi Viñas de Puig, of the University of Girona,

CERTIFY:

That this doctoral thesis, entitled **"Genetic characterization of the Iberian populations of two mollusks: zebra mussel and asiatic clam."**, that Luis Peñarrubia Lozano has submitted to obtain the doctoral degree from the University de Girona, has been completed under their supervision and meets the requirements to opt for the *International Doctor* mention.

Dr. Oriol Vidal Fàbrega

Professor in Genetics

Dr. Jordi Viñas de Puig

Professor in Genetics

Girona, 2016

DEDICATORIA

Después de seis años haciendo una tesis doctoral, podría escribir una segunda explicando la innumerable gente que ha colaborado, tanto en lo profesional como en lo personal, en este largo proceso para, finalmente, tenerla escrita. Es por ello que se merecen todo mi agradecimiento que aquí dejo plasmado.

En primer lloc, m'agradaria agrair de tot cor a l'Oriol i a en Jordi, ja que la seva dedicació cap a aquest treball de manera incondicional ha estat imprescindible en tots els sentits. També vull agrair a tota la gent del Laboratori d'Ictiologia Genètica tot el seu suport, tan fonamental en determinats moments. Una abraçada especial per a Rosa, Carles, David, i per suposat per a Núria, que la seva gran ajuda ha estat fonamental. Sense ells, aquesta tesi no hauria estat possible.

I would like to emphasize all the support and advice received by Jaime during my stay in USA. Thanks to all the TAMUG team and all the good people I met there.

Quiero agradecer también al resto de investigadores en formación que me han acompañado durante estos años, gracias UdGpeople!!! Los que os consideréis compañeros míos, os lo agradezco. Esto no se olvida.

Gracias también a todos los integrantes de grupos tan especiales para mí como los Costro-Calvos, los Discapacity Surfers, los Biolocos, y por supuesto a *NyctiBios Association*. Crecer a vuestro lado estos últimos años ha sido todo un honor.

Por supuesto no puedo olvidarme de Jesús Scrofa. Simplemente, gracias por existir.

He dejado para el final aquesllas personas que siempre se han preocupado, mi familia, especialmente a mis padres Javier y Mª Jesús, todo su apoyo y comprensión. Esto va por vostros.

"Magic is just science that we don't understand yet."

Arthur C. Clarke

ACKNOWLEDGMENTS

Research project

This thesis was funded by the research project "Genetic identification and larvae quantification in invasive mollusks: the zebra mussel and the Asian clam", financied in the period 2009 –2013 by the Spanish Ministry of Science and Innovation (MICINN) with code number CGL2009-09407. This project was carried in the *Laboratori d'Ictiologia Genètica* of the University of Girona, under the supervision of Dr. C. Pla.

Fellowships

L. Peñarrubia received a Researcher Training Fellowship (FPI) with reference number: BES-2010-037446) from the Spanish Ministry of Sciences and Innovation (MICINN) for developing his PhD during the period 2010 – 2014. This FPI fellowship is associated to the research project "Genetic identification and larvae quantification in invasive mollusks: the zebra mussel and the Asian clam" (CGL2009-09407).

L. Peñarrubia also received a Scholar Visitor Fellowship (reference EEBB-I-12-05670) from the Spanish Ministry of Sciences and Innovation (MICINN) in the period September-December 2012 to visit the Molecular Ecology Fisheries Genetics Laboratory of the A&M University at Galveston (TX), USA, under the supervision of Dr. J. Alvarado Bremer. This FPI fellowship is also associated to the research project "Genetic identification and larvae quantification in invasive mollusks: the zebra mussel and the Asian clam" (CGL2009-09407).

Sampling availability

We thank many colleagues for the zebra mussel and Asian clam adult collections: J.C Pérez-Quintero (Universidad de Huelva), R. Araujo (CSIC), J. Checa (Equipos y Suministros, S.L.), I. Planas and R. Puig (Nyctibios Association), R. Miranda (Universidad de Navarra, Spain), C. Alcaraz and M. Pla (IRTA, Government of Catalonia, Spain), R. Mansfield (University of Manchester, UK), B. Gallardo and C. McLaughlan (University of Cambridge, UK), E. Goretti (Università degli Studi di Perugia, Italy), A. Binelli (University of Milan, Italy), J. F. Fruget (Université de Lyon, France), J. N. Beiseland (Université de Lorraine, France), J. Alvarado-Bremer (A&M University at Galveston, TX, USA), C. Robertson (River Studies Program – Inland Fisheries Division, Texas Parks and Wildlife Department, San Marcos, TX, USA), and O. Popa and A. M. Krapaland (National Museum of Natural History, Romania).

We are also indebted to C. Durán (Ebro Hydrografic Confederation, Spain) for the logistic advice, and A. Terrats and C. Solà (Catalan Water Agency from the Government of Catalonia, Spain) for the filtered-water environmental samples and their disponibility.

Finally, we would specially thank to A. Bij de Vaate (Waterfauna Hydrobiologisch Adviesbureau, The Netherlands) for the *D. rostriformis* individuals and for his collaboration in some analyses.

LIST OF PUBLICATIONS

The results of This PhD Thesis have been submitted or published in scientific journals included in the Journal Citation Report (JCR) of the Institute of Scientific Information.

Chapter 1: **Peñarrubia L.**, Sanz N., Pla C., Vidal O. and Viñas J., 2015. **Using Massive Parallel Sequencing for the development, validation, and application of population genetic markers in the invasive bivalve zebra mussel (*Dreissena polymorpha*).** *PLoS ONE* **10**: e0120732.

Chapter 2: **Peñarrubia L.**, Araguas R. M., Pla C., Sanz N., Viñas J. and Vidal O., 2015. **Identification of 246 microsatellites in the Asiatic clam (*Corbicula fluminea*).** *Conservation Genetics Resources* **7**: 393–395.

Chapter 3: **Peñarrubia L.**, Viñas J., Sanz N., Smith B. L., Pla C. and Vidal O., 2016. **SNP identification in two invasive species: zebra mussel (*Dreissena polymorpha*) and Asian clam (*Corbicula fluminea*).** Unpublished.

Chapter 4: **Peñarrubia L.**, Pla C., Viñas J., Vidal O. and Sanz N., 2016. **Genetic characterization of the invasive zebra mussel (*Dreissena polymorpha*) in the Iberian Peninsula.** *Hydrobiologia* **779**: 227–242.

Chapter 5: **Peñarrubia L.**, Pla C., Viñas J., Vidal O., Araguas R. M. and Sanz N., 2016. **Genetic characterization of the Asian clam species complex (*Corbicula*) invasion in the Iberian Peninsula.** *Hydrobiologia* DOI: 10.1007/s10750-016-2888-2.

Chapter 6: **Peñarrubia L.**, Alcaraz C., Bij de Vaate A., Sanz N., Pla C., Viñas J. and Vidal O., 2016. **Validated methodology for quantifying infestation levels of dreissenid mussels in environmental DNA (eDNA) samples.** Unpublished.

TABLE OF ABBREVIATIONS

Abbreviation	Description
µg	Micrograms
µL	microliter
µM	Micromolar
π	Nucleotide Diversity
12S rDNA	12S nuclear ribosomal DNA gene
16S rDNA	16S nuclear ribosomal DNA gene
28S rDNA	28S nuclear ribosomal DNA gene
A	Number of Alleles per locus
A sampling	Autumn period
ACA	Catalan Water Agency
AFLP	Amplified Fragment Length Polymorphism
AICc	Akaike Information Criterion corrected value
AL	Alagon, Ebro River, Spain
AMOVA	Analyses of Molecular Variance
Ar	Allelic Richness
As	Amplicon Size
BIC	Bayesian Information Criterion
bp	Base pairs
BR	Brazos river, USA
CA	Cambre, Mero River, Spain
Cf	*Corbicula fluminea*
CHE	Ebro Hydrographic Confederation
CHG	Guadalquivir Hydrographic Confederation
CHJ	Jucar Hydrographic Confederation
CHS	Segura Hydrographic Confederation
COI mtDNA	Cytochrome Oxidase I, mitochondrial gene
COIR	COI Reference Haplotype
CT	Cycle Threshold
CV	Coefficient of Variation
cyt b	Cytochrome b
D_A	genetic distance, Nei et al., 1983
DAISE	Delivering Alien Invasive Species In Europe
DE	Ebro Delta, Ebro River, Spain
df	Degrees of Freedom
DL	Detection Level
D–loop mtDNA	Mitochondrial Control Region DNA
DNA	Deoxyribonucleic acid
dNTP	Deoxi-Nucleoside Triphosphate
Dp	*Dreissena polymorpha*
E	Efficiency
eDNA	Environmental DNA

emPCR	Emulsion PCR
EST	Expressed Sequence Tag
EU	Europe
F	Forward PCR primer
F	ANOVA statistics
FDR	False Discovery Rate
F_{ST}	Pairwise population differentiation
FW	Freshwater Haplotype
G'_{ST}	Overall Pairwise population differentiation
GA	Rivera Grande, Guadiana River, Spain
Gb	Gigabases
GBS	Genotyping by Sequencing
GO	Gene Ontology
GR	Guadalupe River, USA
GTR	General Time Reversible Model
h	hours
h	Number of haplotypes
H	Haplotypes
H1	Histone 1 gene
H2B	Histone 2B gene
H3	Histone 3 gene
hd	Haplotype Diversity
HKY	Hasegawa–Kishino–Yano Model
Ho	Observed Heterozygosity
HRMA	High Resolution Melting Analysis
Hs	Expected heterozygosity
HWE	Hardy-Weinberg Equilibrium
InDel	Insertion-deletion
IP	Iberian Peninsula
ISSG	Invasive Species Specialist Group
IUCN	International Union for Conservation of Nature
k	Number of genetically Homogeneous Groups
k	Number of Nucleotide Differences
Kb	kilobases
L	Liter
LD	Linkage Disequilibrium
M	molar
m^2	Square meter
MARS	Methionyl-tRNA Synthetase gene
MCMC	BayesianMarkov ChainMonte Carlo
MDS	Multidimensional Scaling Analysis
$MgCl_2$	Magnesium Chloride
MICINN	Ministry of Sciences and innovation
min	minutes
ML	Maximum Likelihood

mL	mililiter
mM	milimolar
MMARM	Ministry of Environment, rural and Marine
MNV	Multiple Nucleotide Variants
MPS	Massive Parallel Sequencing
mtDNA	Mitochondrial DNA
n	Sample size
N50	Median of Massive Sequencing sequences
NA	Number of Null Alleles
N_A	Number of Alleles per locus
ND	Non-Detected
ng	Nanograms
NGS	Next Generation Sequencing
NJ	Neighbor-Joining
NV	Nucleotide Variant
P	Significance Value
PAL	Potential Amplifiable Loci
PCR	Polymerase Chain Reaction
PIC	Polymorphic Information Content
ϕ_{ST}	Pairwise population differentiation
PSV	Paralogous Sequence Variant
q	Allele frequence
QL	Quantification Level
qPCR	Quantitative PCR
R	Reverse PCR primer
R^2	Coefficient of Regression
RA	RA Corbicula Lineage
RAD-seq	Restriction-site-Associated DNA Sequencing
RAPD	Randomly Amplified Polymorphic DNA
RB	Ribarroja Reservoir, Ebro River, Spain
RFLP	Restriction Fragment Length Polymorphism
RH	Rhone River, France
RlcB	RlcB Corbicula Lineage
RO	Danube River, Romania
rRNA	Ribosomal RNA
S	Size Range
s	seconds
S	Total Number of Variable Sites
S sampling	Spring Period
SB	Sabine River, USA
SC	SC Corbicula Lineage
SD	Standard Deviation
SEM	Standard Error of the Mean
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variants

SR	28S Reference Haplotype
SSR	Single Sequence Repeat
STR	Short Tandem Repeats
t	T-student statistic
T92	Tamura 3-Parameter Model
Ta	Annealing temperatures
TE	Ter River, Spain
Tm	Melting Temperature
TMS	Theoretical Mutational Steps
TPM	Two-Phase Model of Mutation
TRF	Tandem Repeats Finder Software
tRNA	Transfer RNA
TY	Trinity River, USA
UK	United Kingdom
URA	Vasc Water Agency
USA	United States of America
VC	Variants Coverage
w/d	Without Data
XE	Xerta, Ebro River, Spain
ZG	Canal Imerial de Aragón, Zaragoza, Ebro River, Spain

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

Summary

Biological invasions are considered one of the major threats to conservation because of the great ecological effects such as habitat degradation and loss of native biodiversity. It has been estimated that this deterioration is much more intense in freshwater habitats, where a global decline of biodiversity is observed.

The zebra mussel (*Dreissena polymorpha* Pallas, 1771) and the Asian clam (*Corbicula fluminea* Müller, 1774) are considered two of the worst invasive aquatic species worldwide. The zebra mussel is native to the Ponto-Caspian region, while the Asian clam is considered native to Asia, Middle East, Australia and Africa. In the last decades the two species have successfully expanded worldwide, and they have caused major ecological and economic impacts on aquatic ecosystems. Both species are present in the Iberian Peninsula since several years. The zebra mussel was first detected in the middle reaches of the Ebro River in 2001. Since then this species has expanded across the Ebro River basin as well as other adjacent basins of the Northeast of the Iberian Peninsula. The Asian clam was first recorded in 1980 in the mouth of the Tajo River. While the range distribution of the zebra mussel is found mainly in the Northeast of the Iberian Peninsula, the Asian clam is much more widespread and can be found in all major Iberian basins. However, introduction history and colonization routes of these two invasive species in the Iberian Peninsula remain mainly unknown.

In this thesis, we have focused on understanding the possible sources of invasion of the two species, and how they have colonized and have expanded across the Iberian basins, using molecular and population genetics techniques. In this respect, we have been developed first new molecular markers to increase the number of genetic markers in the two species. We have focused on optimizing new sets of microsatellites using bioinformatic analysis of the results obtained by massive genome sequencing. Additionally, we use the information generated by the massive sequencing to identify and validate new SNPs markers.

Second, after selecting the best molecular markers described in each of the two species, we characterized the genetic structure of the Iberian populations to infer their possible invasion routes. Our results show that the invasion of the zebra mussel in the Iberian Peninsula is explained by a unique and recent invasion episode. Since its introduction in the middle reaches of the Ebro River, this species has spread throughout the Ebro basin, and other rivers, using a continuous expansion model. Our results also show how the zebra mussel invasion in the Iberian Peninsula may be a different episode from the rest of the European invasion, and therefore contradict previous studies that placed the origin populations in France.

Regarding the Asian clam, our results suggest that there have been at least two colonization episodes in the Iberian Peninsula. The first covers almost the entire range distribution in the Peninsula, while the origin of the second episode is more limited to the lower Ebro River in the

east of the Peninsula, posteriorly expanding upstream throughout the basin. The two colonization episodes converge on the middle reaches of the Ebro River, from where Asian clam populations present in adjacent basins seem to have come.

Finally, we have developed and optimized a genetic method based on environmental DNA and Real Time PCR, to detect larvae of dreissenids in water samples.

Resumen

Las invasiones biológicas son consideradas una de las principales amenazas para la conservación debido a los grandes efectos ecológicos que provocan tales como la degradación del hábitat y la pérdida de la biodiversidad autóctona. Se ha estimado que este deterioro es mucho más intenso en aguas dulces, en las que se observa un decline global de su biodiversidad.

El mejillón cebra (*Dreissena polymorpha* Pallas, 1771) y la almeja asiática (*Corbicula fluminea* Müller, 1774) están consideradas dos de las peores especies acuáticas invasoras en todo el mundo. El mejillón cebra es originario de la región de los mares Negro, Azov y Caspio, mientras que la almeja asiática se considera nativa de Asia, Oriente Medio, Australia y África. En las últimas décadas las dos especies se han expandido con éxito en todo el mundo, y han causado grandes impactos a nivel ecológico y económico en ecosistemas acuáticos. Ambas están presentes en la Península Ibérica desde hace años. El mejillón cebra fue detectado por primera vez en la parte baja del río Ebro en el embalse de Ribarroja en 2001. Desde entonces esta especie se ha expandido por toda la cuenca del río Ebro así como también en otras cuencas adyacentes del noreste de la Península Ibérica. La almeja asiática fue detectada por primera vez en 1980 en la desembocadura del río Tajo. Mientras el rango de distribución del mejillón cebra se encuentra principalmente en el noreste de la Península Ibérica, la almeja asiática está mucho más extendida y puede encontrarse en todas las principales cuencas ibéricas. Sin embargo, no se conocen de manera cierta ni la historia de su invasión ni las rutas de colonización de las dos especies en la Península Ibérica.

En esta tesis, nos hemos enfocado en entender las posibles fuentes de invasión de las dos especies, y como han colonizado y se han expandido a través de las cuencas ibéricas, utilizando técnicas de genética molecular y de poblaciones. En este sentido, se han desarrollado en primer lugar nuevos marcadores moleculares para incrementar el número de marcadores genéticos disponibles en las dos especies. Nos hemos centrado en optimizar nuevos juegos de microsatélites a partir del análisis bioinformático de los resultados obtenidos de la secuenciación masiva del genoma. Adicionalmente, usamos la información generada por la secuenciación masiva para detectar y validar nuevos marcadores SNPs.

En segundo lugar, después de seleccionar los mejores marcadores moleculares descritos en cada una de las dos especies, se ha caracterizado la estructura genética de las poblaciones ibéricas para inferir sus posibles rutas de invasión. Nuestros resultados muestran que la invasión del mejillón cebra en la Península Ibérica se explica con un episodio de invasión único y reciente. Desde su introducción en la parte media del río Ebro, esta especie se ha expandido por toda la cuenca, además de otros ríos, utilizando un modelo continuo de expansión. Nuestros resultados también muestran como la invasión del mejillón cebra en la Península Ibérica podría constituir un episodio de invasión diferente al resto de Europa, y por lo tanto contradicen estudios previos que situaban el origen a partir de poblaciones de Francia.

Respecto a la almeja asiática, nuestros resultados sugieren que ha habido al menos dos episodios de colonización en la Península Ibérica. El primero engloba casi todo el rango de distribución de la almeja en la Península, mientras que el origen del segundo episodio se encuentra más limitado al este de la Península en la parte baja del río Ebro, desde donde se ha expandido río arriba por toda la cuenca. Los dos episodios de colonización convergen en el tramo medio del río Ebro, desde donde parece que han surgido las poblaciones de almeja asiática presentes en cuencas adyacentes.

Finalmente, se ha desarrollado y optimizado un método genético, basado en DNA ambiental y Real Time PCR que permite la detección de larvas de especies de dreissenidos en masas de agua.

Resum

Les invasions biològiques són considerades una de les principals amenaces per a la conservació a causa dels grans efectes ecològics que provoquen, com ara la degradació de l'hàbitat i la pèrdua de la biodiversitat autòctona. S'ha estat estimat que aquest deteriorament és molt més intens en aigües dolces, en les quals s'observa un declini global de la seva biodiversitat.

El musclo zebrat (*Dreissena polymorpha* Pallas, 1771) i la cloïssa asiàtica (*Corbicula fluminea* Müller, 1774) estan considerades dues de les pitjors espècies aquàtiques invasores a tot el món. El musclo zebrat és originari de la regió dels mars Negre, Azov i Caspi, mentre que la cloïssa asiàtica es considera nativa d'Àsia, l'Orient Mitjà, Austràlia i África. En les últimes dècades les dues espècies s'han expandit amb èxit arreu del món, i han causat grans impactes a nivell ecològic i econòmic en ecosistemes aquàtics. Totes dues són presents a la Península Ibèrica des de fa anys. El musclo zebrat va ser detectat per primera vegada a la part baixa del riu Ebre a l'embassament de Riba-roja l'any 2001. Des de llavors aquesta espècie s'ha expandit per tota la conca del riu Ebre així com també en altres conques adjacents del nord-est de la Península Ibèrica. La cloïssa asiàtica va ser detectada per primera vegada el 1980 a la desembocadura del riu Tajo. Mentre el rang de distribució del musclo zebrat es troba principalment al nord-est de la Península Ibèrica, la cloïssa asiàtica està molt més escampada i pot trobar-se en totes les principals conques ibèriques. No obstant això, no es coneixen de manera certa ni la història de la seva invasió ni les rutes de colonització de cap de les dues a la Península Ibèrica.

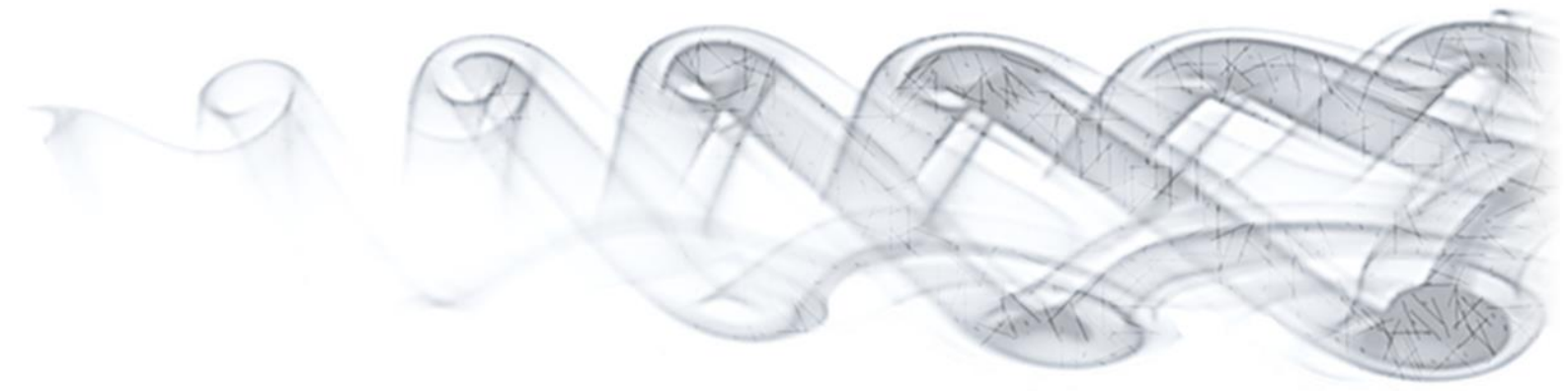
En aquesta tesi, ens hem enfocat a entendre les possibles fonts d'invasió de totes dues espècies, i com han colonitzat i s'han expandit a través de les conques ibèriques, fent servir tècniques de genètica molecular i de poblacions. En aquest sentit, s'han desenvolupat en primer lloc nous marcadors moleculars per incrementar el nombre de marcadors genètics disponibles. Ens hem centrat en optimitzar un nou joc de microsatèl·lits a partir de l'anàlisi bioinformàtica dels resultats obtinguts de la seqüenciació massiva del genoma. Addicionalment, vam fer servir la informació generada per la seqüenciació massiva per detectar i validar nous marcadors SNPs.

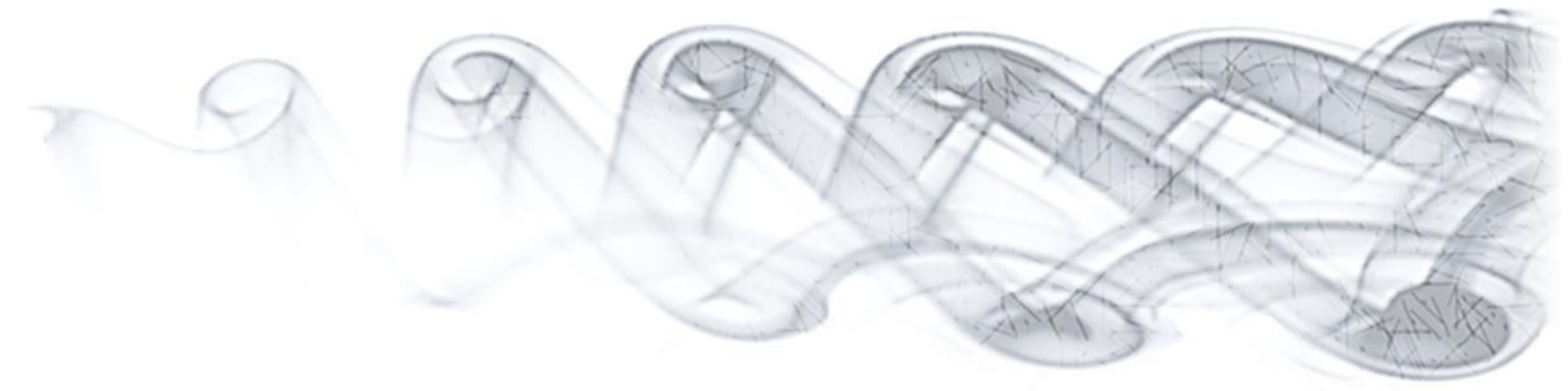
En segon lloc, després de seleccionar els millors marcadors moleculars descrits a cadascuna de les dues espècies, s'ha caracteritzat l'estructura genètica de les poblacions ibèriques per inferir les possibles rutes de invasió. Els nostres resultats mostren que la invasió del musclo zebrat a la Península Ibèrica s'explica amb un episodi d'invasió únic i recent. Des de la seva introducció a la part mitjana del riu Ebre, aquesta espècie s'ha expandit per tota la conca, a més d'altres rius, fent servir un model continu d'expansió. Els nostres resultats també mostren com la invasió del musclo zebrat a la Península Ibèrica podria constituir un episodi d'invasió diferent a la resta d'Europa, i per tant contradiuen estudis previs que situaven l'origen a partir de poblacions de França.

Respecte a la cloïssa asiàtica, els nostres resultats suggereixen que hi ha hagut almenys dos episodis de colonització a la Península Ibèrica. El primer engloba gairebé tot el rang de distribució de la cloïssa a la Península, mentre que l'origen del segon episodi es troba més limitat a l'est de la Península a la part baixa del riu Ebre, des d'on s'ha expandit riu amunt per tota la conca . Els dos episodis de colonització convergeixen en el tram mitjà del riu Ebre, des d'on sembla que han sorgit les poblacions de cloïssa asiàtica presents en conques adjacents.

Finalment, s'ha desenvolupat i optimitzat un mètode genètic, basat en DNA ambiental i en Real Time PCR que permet la detecció de larves d'espècies de dreissenids en masses d'aigua.

1. SCIENTIFIC BACKGROUND





1. SCIENTIFIC BACKGROUND

1.1. Invasive species

An invasive species is defined as a non-native species whose introduction into new areas causes environmental and/or economic harms (adapted from https://www.invasivespeciesinfo.gov/whatis.shtml). Invasions are considered a main threat for conservation because their large ecological effects such as habitat degradation, decreased diversity and high economic impacts on the ecologic communities they invade (Sakai *et al*. 2001). Introduction of a species in a non-native area usually occurs because of anthropogenic factors rather than by natural events (Cristescu 2015). Furthermore, these impacts have been estimated to be greater in freshwater habitats, which are already suffering a global decline in biodiversity (reviewed in Vaughn and Hakenkamp 2001). Introduction of aquatic organisms away of their native ranges has occurred intentionally or unintentionally over the last centuries for connecting river basins and associated passive transport (Bij de Vaate *et al*. 2002).

1.1.1. Stages of successful biological invasions

Interestingly, not all of the nonindigenous species arriving to a new ecosystem become successful invaders. In fact, approximately only 10 % (in a range between 5 to 20 %) of the migrant species are able to develop populations dense enough to persist in the new habitat (Bij de Vaate *et al*. 2002). In general, biological invasions occur as a combination of different steps following transport of individuals from occupied areas (either native or already invaded) into a new location. These invasion processes include three main stages: (1) introduction, (2) establishment, and (3) secondary spread of the invasive species in the new habitat (Sakai *et al*. 2001; Allendorf and Lundquist 2003). Each stage of the invasive process is associated with different "competition" traits that may act as the key point for survival, and thus they are extremely relevant in sustaining the final invasion (Blackburn *et al*. 2015).

Specifically, the first stage is an initial introduction of the invasive organism. This introduction is usually carried out by intentionally or unintentionally transport of individuals. It is clear that the larger number of transported individuals, the higher probability of invasion success (Blackburn *et al*. 2015). The second stage requires a successful establishment of a self-sustaining population (Sakai *et al*. 2001). This is more likely to become successful if the new habitat has conditions similar to those of the native one (Suarez and Tsutsui 2008). Populations originated from a small number of individuals may suffer from reduced genetic diversity, genetic drift, and bottlenecks; all of which could cause declines in their adaptation to the new habitat (Sakai *et al*. 2001; Blackburn *et al*. 2015). Then, the number of transported individuals also influences this step. Finally, the last stage of successful invasions includes distribution within the new range, usually by gradual short-distance dispersal by passive diffusion, but by long dispersal jumps, often derived from human transport, as well (Blackburn *et al*. 2015). This stage will exhibit exponential population growth, sometimes to densities higher than those in its native range (Bij de Vaate *et al*. 2002).

In conclusion, several factors are needed to be accomplished for a successful colonization in a new environment (Sakai *et al*. 2001; Allendorf and Lundquist 2003). In general, the most important are: (1) the species capacity to be transported over large distances (Suarez and Tsutsui 2008), and (2) climate similarity between native and non-native habitats (Karatayev *et al*. 2007). In this sense, the increase of worldwide transportation together with the climate change is contributing to the easy transport and settlement of putative invasive species (Vaughn and Hakenkamp 2001; Sousa *et al*. 2014).

1.1.2. Zebra mussel and Asian clam species description

Two clear examples of successful invasive species are the zebra mussel (*Dreissena polymorpha* Pallas, 1771) and the Asian clam (*Corbicula fluminea* Müller, 1774). The zebra mussel and the Asian clam belong to the Subclass Heterodonta, Order Veneroidea. Specifically, the zebra mussel is within the Superfamily Dreissenoidea and Family Dreissenidae, whereas Asian clam belongs to the Superfamily Cyrenoidea and Family Cyrenidae.

The zebra mussel is native of the Ponto-Caspian region, including the Black, Caspian and Azov Seas, and the lower section of Danube River (Albrecht *et al*. 2007). On the other hand, the Asian clam is considered native of Asia, the Middle East, Australia and Africa (McMahon 1982).

1.1.2.1. *Life cycle characteristics*

The zebra mussel and the Asian clam possess the ability to colonize a vast range of habitats because of their high genetic variability and phenotypic plasticity (Figure 1.1), high tolerance to different conditions, short generation times, rapid growth, sexual maturity at early ages, and high fecundity (Sousa *et al*. 2014). In addition to these traits, their nonspecific food preference and the benefits of gregarious behavior (Karatayev *et al*. 2005) -with colonies of 150,000 individuals/m^2 for the zebra mussel (Astanei *et al*. 2005), and up to 2,000 individuals/m^2 for the Asian clam (Sousa *et al*. 2008)- may contribute to settlement of new large colonies.

Both species have the capacity of sexual reproduction with early sexual maturity and high fecundity rate, releasing offspring numbers of up to 1 million eggs per female/season in the case of the zebra mussel (Astanei *et al*. 2005). However, in the invaded areas the Asian clam reproduces exclusively by androgenesis (Pigneur *et al*. 2012). Androgenesis is a type of asexual reproduction consisting in the fecundation of one oocyte by an immature sperm with the subsequent elimination of the nuclear maternal DNA, and thus producing an offspring with a nuclear genome identical to the genome of the male progenitor (Pigneur *et al*. 2012). As a consequence, a single individual can release up to 90,000 offspring clones genetically identical to the parental progenitor (Pigneur *et al*. 2014a).

1.1.2.2. *Invasiveness potential*

These two species have several life history-traits useful during the different stages of a successful invasion (described in the section 1.1.1). Large numbers of individuals of zebra mussels and Asian clams can be easily carried along by human transport activities, allowing the

opportunistic transfer of these invasive bivalves (Sousa *et al*. 2014). Larvae of both species resist transport in ballast waters, which may explain their expansion throughout North American and European continents (Karatayev *et al*. 2007; Brown and Stepien 2010, Pigneur *et al*. 2014a). In addition, adult individuals can be passively transported by commercial and recreational ships, attached to the hulls in case of zebra mussels (Johnson and Carlton 1996) or by commercial traffic as food resource in case of Asian clams (Karatayev *et al*. 2007). This clearly accomplishes the requirements of the initial introduction step.

The second successful stablishment step is conditioned by intrinsic biological factors such as lag time (period between this initial colonization and an exponential growth in the following stage) and morphologic plasticity (capacity of generating adaptive phenotypes in a new area) (Suarez and Tsutsui 2008; Stapley *et al*. 2015). In this sense, both zebra mussel and Asian clam have shorter lag times than other competing bivalve species (Karatayev *et al*. 2015) and high plasticity levels (Pigneur *et al*. 2014a; Pigneur *et al*. 2014b).

In addition, both species have the capacity of release large numbers of offspring (Astanei *et al*. 2005; Sousa *et al*. 2008), this being a crucial feature for the last invasion step with secondary expansions in the new area (Karatayev *et al*. 2005). Then, once first introduction is already stablished, secondary introductions do not necessarily experience the same first "competition" traits for survival, and individuals just interact with the large number of individuals already established for increasing secondary spread in the new range (Blackburn *et al*. 2015). Moreover, the large number of relocated individuals plus the numerous offspring may help retaining high levels of genetic diversity, thus avoiding bottlenecks (Sakai *et al*. 2001). Populations with higher genetic diversity may present more capacity of adaptation to new conditions and higher adabtiblity to new area (Sakai *et al*. 2001), although other traits may be relevant as well. However, adaptative success depends on several traits, and the association between genetic diversity levels and adaptation capacity to new conditions is not completely correlated (Roman and Darling 2007).

1.1.2.3. *Invasive range*

Although both species can be sympatric in the invaded areas, their distribution ranges usually differ. One of the major factors affecting bivalves' distribution and abundance is the availability of a suitable substrate, and the two species differ in habitat tolerance and preference (Karatayev *et al*. 2005). Thus, while zebra mussels grow in high densities in lakes and reservoirs because they need to attach to hard surfaces, Asian clams usually dominate rivers and small streams because they bury themselves in the sand or other soft substrates. Minimum temperature for growth and development is similar (10—11^{o}C), but zebra mussel presents smaller tolerance limits (2-33 ^{o}C) than Asian clam (0-37 ^{o}C). In this same sense, salinity tolerance and pH tolerances are higher in the Asian clam (salinity concentration = 10-17 ‰; low pH tolerance = 5.6) than in zebra mussel (salinity concentration = 4-6.2 ‰; low pH tolerance = 7.4). Accordingly, zebra mussel range distribution remains more limited than Asian clam (Karatayev *et al*. 2007). For instance, based on the US Geological Survey official North American website for invasive species monitoring (http://nas.er.usgs.gov/), zebra mussel distribution is mainly situated in the East North America, and zebra mussels have not been

detected in open waters of the West. In contrast, Asian clam individuals are well distributed across the entire North American continent. In Europe, the Delivering Alien Invasive Species In Europe (DAISIE) project developed by the official European Commission for invasive species monitoring (http://www.europe-aliens.org/) situates the Asian clam distribution reaching a more southern distribution range than the zebra mussel.

1.1.3. Ecological Impacts

Bivalves produce the majority of the successful invasions found in freshwater ecosystems (Howard and Cuffey 2006; Lopes-Lima *et al*. 2014). They can remove large amounts of phytoplankton, bacteria and particulate organic matter from the water column, which they transfer to the substrate. This would mean that subsequently inorganic and organic component of the sediment increases, and the structure of the benthic community becomes affected (Vaughn and Hakenkamp 2001; Karatayev *et al*. 2007). Because of this, some authors consider bivalves as ecosystem engineers (Karatayev *et al*. 2005; Sousa *et al*. 2014), and when invading, they are responsible for a wide range of environmental changes, affecting from individuals to ecosystems (Sousa *et al*. 2014). These impacts change from hydrology and biogeochemical cycling to biotic interactions through several mechanisms. Therefore, the main ecological impacts produced by the zebra mussel and the Asian clam are consequence of the biofilter activity of their massive colonies producing a significant disequilibrium (Hakenkamp *et al*. 2001; Sousa *et al*. 2008; Lucy *et al*. 2012). Specifically, the reduction in plankton and thus photosynthesis induce changes in water clarity and oxygen availability in the benthos (Malmqvist 2002; Howard and Cuffey 2006). Moreover, benthic algal abundance becomes enhanced and the submerged vegetation increases (Spooner and Vaughn 2006).

At the level of animal communities, competition for habitat and food can lead to replacement of autochthonous bivalve species during invasion (Sakai *et al*. 2001; Allendorf and Lundquist 2003). Native unionids, chironomid and sphaeriid population densities are negatively affected by zebra mussel invasion (Descy *et al*. 2003). This decrease may depend on invasive species colony density, time since invasion, and type of bottom sediments (Karatayev *et al*. 2007). Previous studies have observed that severe unionid mortality (>90%) occurs when zebra mussel density and mean infestation intensity reach 6,000 individuals/m^2 and the proportion outcomes 100 dreissenids/unionid (Ricciardi *et al*. 1995; Strayer and Malcom 2007).

Finally, dense populations of invasive bivalve species can also alter predatory behavior of native fish and birds (Strayer *et al*. 2004), which may alter their dietary intake to include these new species. However, no evidence of any long-term decline in invasive populations has been observed, even in punctual situations of 90% depletion by predation (Karatayev *et al*. 2005; Sousa *et al*. 2014), with a quick reestablishment of invasive populations. Paradoxically, consumptions of larvae of Asian clam and zebra mussel by fishes and birds may facilitate their spread as they can survive gut passage (Galtin *et al*. 2013).

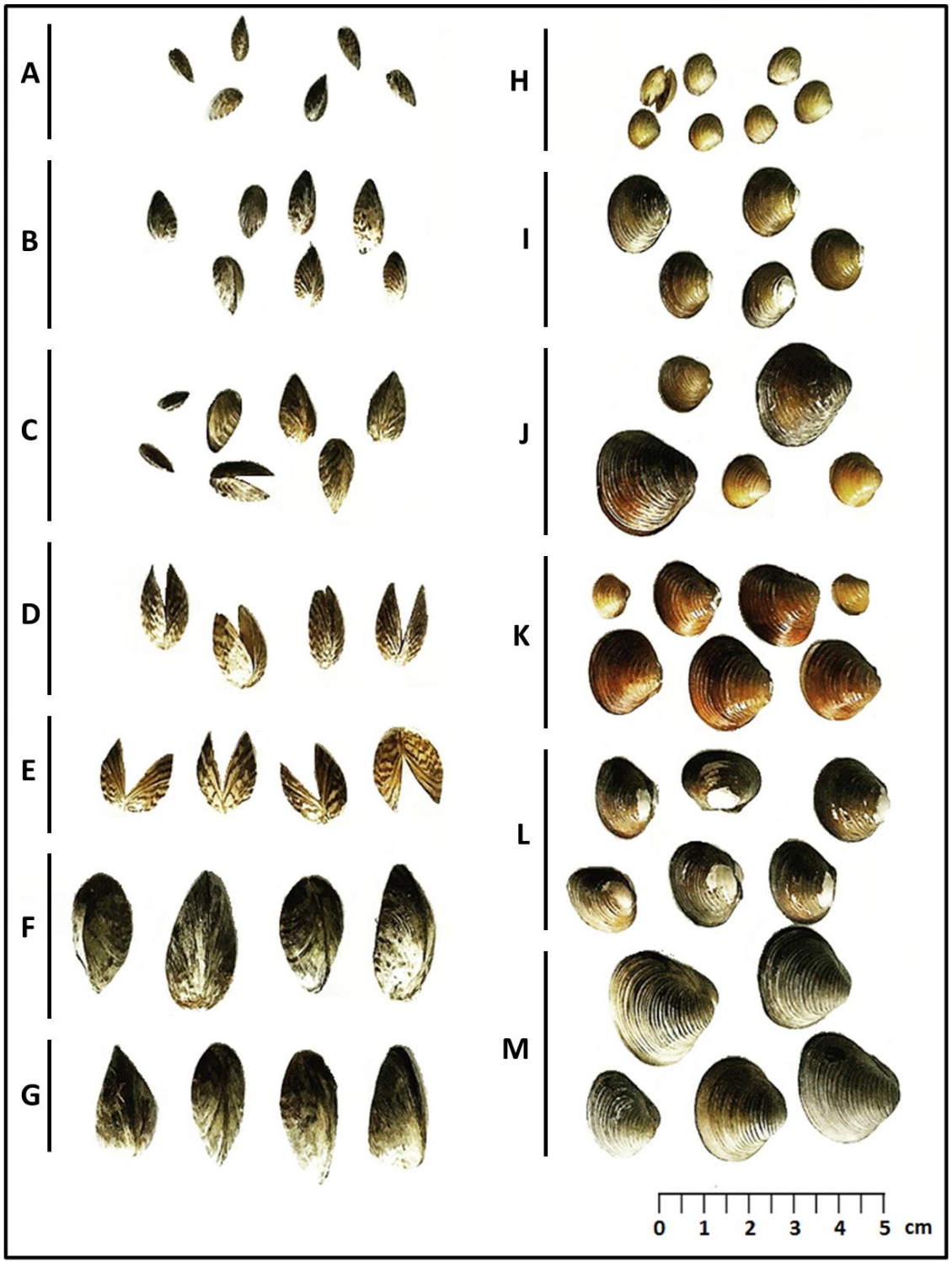


Figure 1.1. Pictures of zebra mussel (A-G) and Asian clam (H-M) adult individuals. **A**: Trasimeno Lake, Umbria, Italy; **B**: Rhône River, Lyon, France; **C**: Ebro River, Burgos, Spain; **D**: Danube River, Moldavia, Romania; **E**: Lugano Lake, Lombardy, Italy; **F**: Llobregat River, Barcelona, Spain; **G**: Ebro River, Tarragona, Spain; **H**: Ter River, Girona, Spain; **I**: Guadiana River, Huelva, Spain; **J**: Ebro River, Zaragoza, Spain; **K**: Miño River, Pontevedra, Spain; **L**: Mero River, A Coruña, Spain; **M**: Rhône River, Lyon, France.

1.1.4. Economic impacts

The ability of zebra mussels to attach to solid substrates, creating layers of shelves, may end up collapsing artificial waterways. Nuclear and hydroelectric power stations are the most harmed facilities by fouling of pipes, but they can also harm ship hulls, aquaculture cages, or other navigational structures (Oscoz *et al*. 2010). Conversely, the economic effects of Asian clams mostly come from macrofouling of water conduction systems in fossil-fueled or nuclear power stations, and enhancement of sedimentation rates in irrigation channels (Lucy *et al*. 2012). Overall, up to $2 billion are spent every year in the USA for control, replacement, and repair water structures (Pimentel *et al*. 2005). In the Iberian Peninsula during the 2005-2009 time period, the cost derived from zebra mussel invasion has been estimated in €11.5 million, including operation problems in affected facilities, cleaning, and control treatments (Durán *et al*. 2012).

1.2. Invasion of the Iberian Peninsula

1.2.1. Distribution

The zebra mussel was first reported in the Iberian Peninsula in the low reaches of the Ebro River, in the Ribarroja Reservoir in 2001 (Ruíz-Altaba *et al*. 2001). The relatively late Iberian invasion contrasts with the detection timing in other Western European habitats. For instance, first detection in UK and France were in 1824 and 1826, respectively (revised in Strayer and Smith 1993). This late invasion of zebra mussel in the Iberian Peninsula can probably be attributed to the Pyrenees acting as a barrier (Rajagopal *et al*. 2009; Bij de Vaate *et al*. 2013). The origin of the Iberian invasion is supposed to be occurred through adults or larvae carried along by sport fishing boats or by fish transport trailers proceeding from France (Rajagopal *et al*. 2009). Since then, this species has expanded across all Ebro River basin (Durán and Anadón 2008; Oscoz *et al*. 2010), to also adjacent basins in the Northeast of the Iberian Peninsula, such as Mijares River, 105 km south of Ebro River (Navarro *et al*. 2013); and Llobregat River, 215 km north of Ebro River (ACA 2015). Recently (2009-2011 period), zebra mussels have also been detected in the Guadalquivir River in the South of the Iberian Peninsula (Figure 1.2A).

The first detection of the Asian clam in the Iberian Peninsula was documented in 1980 in the Tajo River estuary (Mouthon 1981). Since then, this species has been continuously spreading. While the zebra mussel is restricted to the Northeast of the Iberian Peninsula, the Asian clam distribution has a wider extension being found in all main river ecosystems (Figure 1.2B). These range distribution differences between both species could be explained by their invasion history. Since the Asian clam was firstly recorded 30 years before of first citation of zebra mussel had more time to spread along the Iberian basins. In addition, because of Asian clam larvae have the capacity of crawling away (rather than float in plankton) in its first life stages, it can easily avoid the obligated unidirectionally downstream spread following the river flow (Karatayev *et al*. 2005) and thus have the capacity of spreading also upstream.

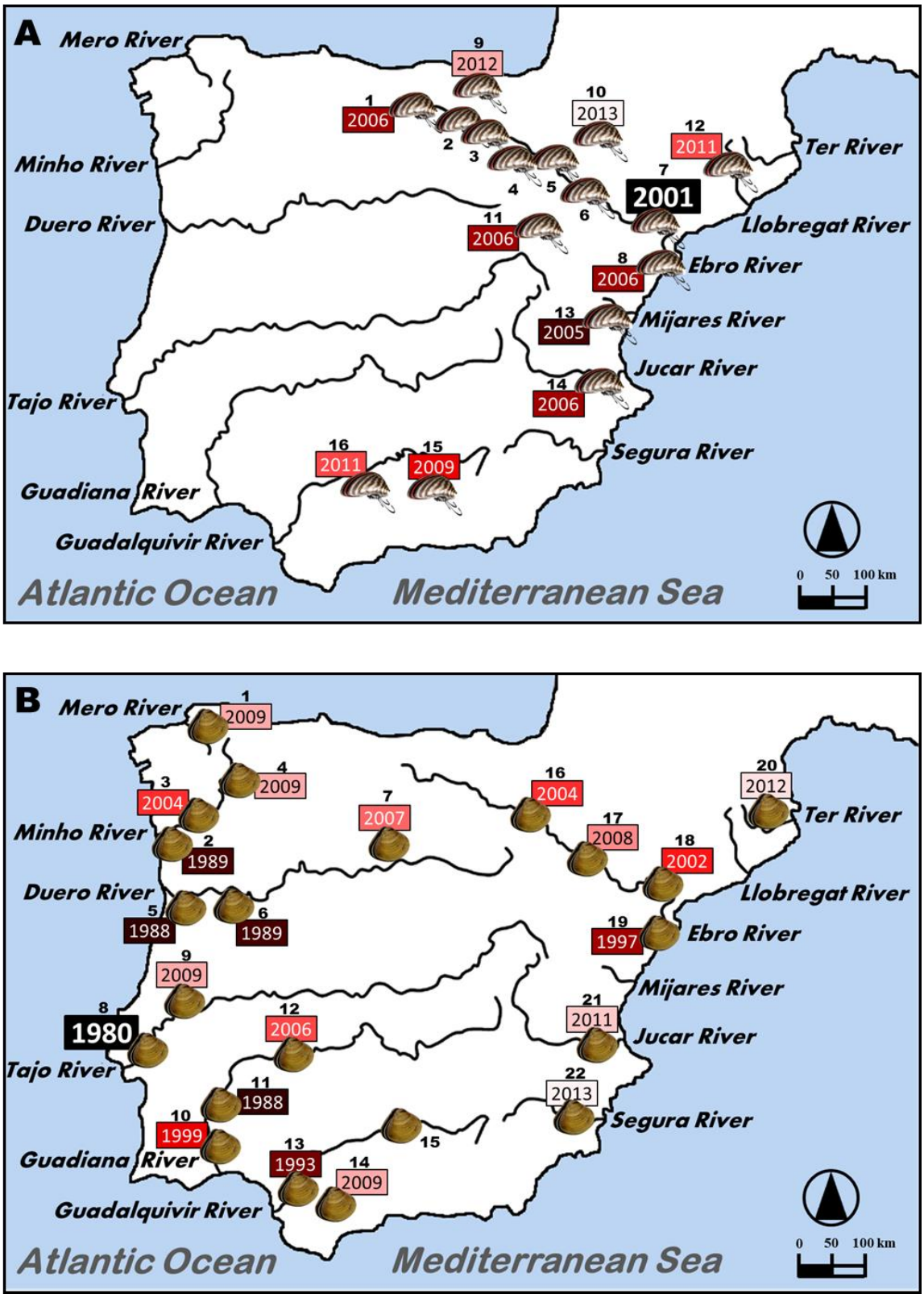


Figure 1.2. Invasion history in the Iberian Peninsula. A: Zebra mussel. **B**: Asian clam. See Additional Table 7.2.1. and Additional Table 7.2.2. for location numbers. Temporality of the invasions is represented from dark (older) to light (newer) red. First detection in the Iberian Peninsula for each species is indicated in black.

1.2.2. Control and management

Understanding the invasion history of the zebra mussel and the Asian clam in the Iberian Peninsula could be essential for environmental management and conservation. Up to seven species of native freshwater bivalve, belonging three major families (Sphaeriidae, Unionidae, and Margaritiferidae), are present in the Iberian Peninsula, and all of them are being negatively affected by zebra mussel and Asian clam presence (Araujo 2004; Oscoz *et al*. 2006). Due to the fact that eradication of invasive bivalve is extremely difficult once populations are established (Sousa *et al*. 2014), the implementation of control and management plans focused on detecting initial infestation stages is crucial to avoid a full establishment of the invasive species (Lucy 2006; Durán *et al*. 2010) and contain further expansions (Ricciardi 2003).

In this sense, the relatively late invasion of the zebra mussel in the Iberian Peninsula has allowed developing governmental plans of control and monitoring strategies to prevent large-scale expansion. The National Strategy for the Control of Zebra Mussel in Spain (MMARM 2007) started in 2007 and focused on avoiding initial establishment of new populations based on a quick detection. The main objectives of the government are to prevent the spread of zebra mussels into uncontaminated bodies of water, as well as finding out more about the biology and behavior of the species in order to design control measures as effective as possible (Durán *et al*. 2010). To that end, the National Strategy included integrated ecological and cost-effective management plan involving information and awareness campaigns, strategies of the construction of disinfection stations, closing uncontrolled accesses, and research on the species and its interaction with the environment in affected and in unaffected areas (Durán *et al*. 2010).

Because of zebra mussel distribution mainly remains in the Northeast of the Iberian Peninsula, governmental organisms such as the Ebro Hydrographic Confederation (CHE) of the Agriculture, Food and Environment Ministry of Spain, and the Catalan Water Agency (ACA) of the Agriculture, Fisheries, Food and Environment Department of the Autonomous Community of Catalonia, monitorize the Ebro River and the rest of the Catalan basins (ACA 2015; CHE 2015), by conducting annual campaigns of larval and adult zebra mussels detection (ACA 2015; CHE 2015). Current monitoring plans involve adult individual detection by direct visual diagnostics and veliger larvae detection in filtered water volumes through microscopic methods (Lucy 2006).

By contrast, Asian clam remains as a secondary priority in Spain's plans against invasive species, probably because the higher visual impacts produced by the zebra mussel mask the economic impacts produced by Asian clams. Although some governmental organisms have monitoring plans for adult detection and they gather data about the expansion into the Iberian Peninsula, no particular actions are taken to diminish densities and/or protect newly infected waterbodies. Fortunately, strategies carried out to control the zebra mussel spread concurrently prevent extensive spreading of the Asian clam, because as described before the dissemination of larvae by anthropogenic factor is the main spreading mechanism and is shared by both species (Karatayev *et al*. 2007).

As we have seen, spread mechanisms of zebra mussel and Asian clam include the transport of juvenile states (Karatayev *et al*. 2007; Brown and Stepien 2010, Pigneur *et al*. 2014a), and thus early monitoring and detection of larvae may be crucial to prevent new establishments. For this reason, governmental ACA and CHE organisms also include in their prevention efforts water bodies where zebra mussel is not already present (ACA 2015; CHE 2015), and they specifically focus on larvae detection (Figure 1.3). However, as the current methodologies are based on microscopically visual detection (Lucy 2006), false positives are common due to the morphological similarity of zebra mussel larvae with other organisms (Baldwin *et al*. 1996). In addition, these protocols present limitations to detect larvae at low concentrations. Improved molecular protocols could generate more accurate larvae quantification results and, subsequently, detect water bodies with real potential risk of invasion.

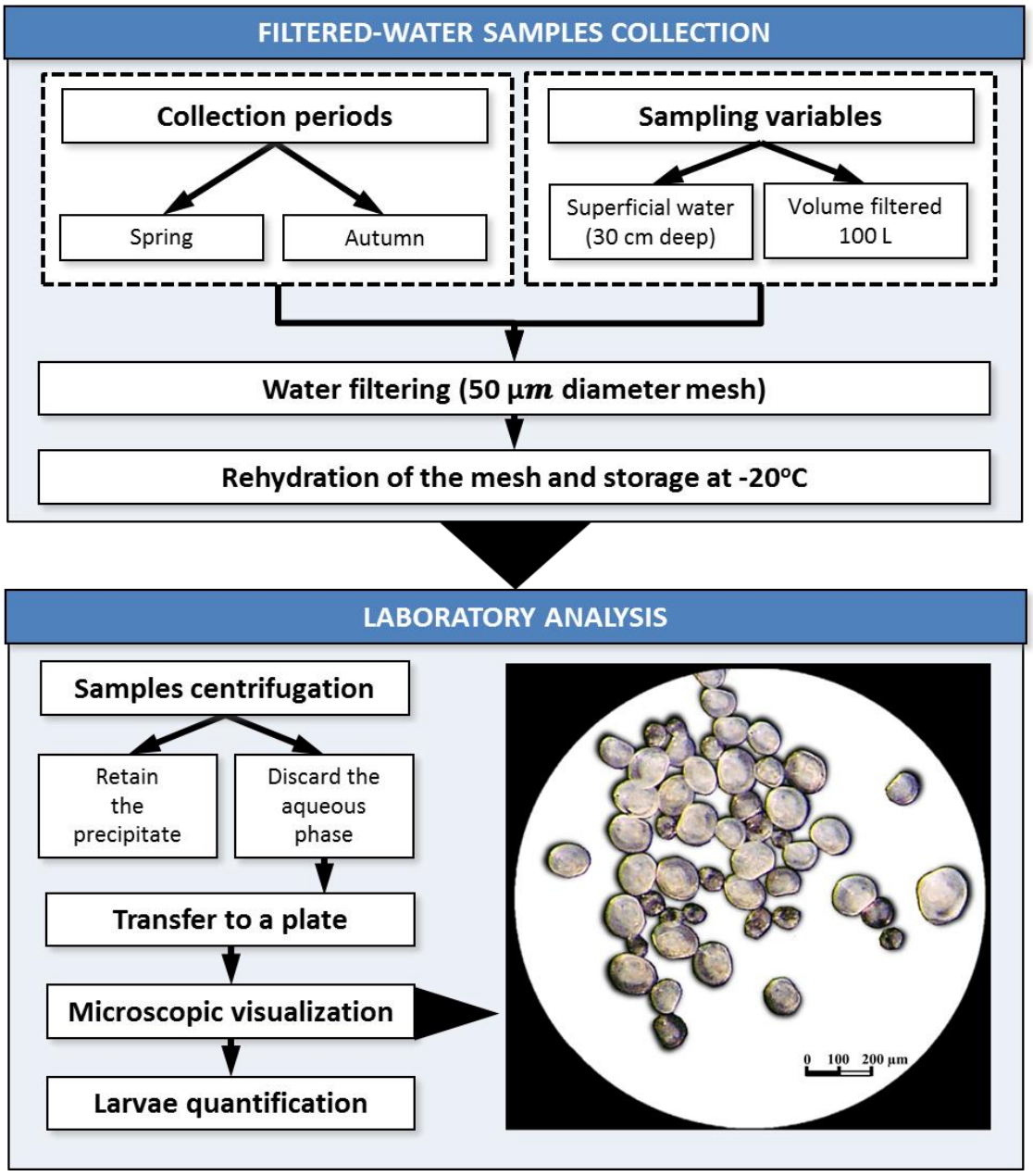


Figure 1.3. Workflow of the protocol used by the official Spanish governmental agencies for the detection and quantification of zebra mussel larvae in water samples.

1.3. Genetics of invasive species

Conservation genetics is an interdisciplinary science that aims to apply genetic methods to the conservation and restoration of biodiversity (adapted from Avise 1994). Researchers involved in conservation genetics come from a variety of fields including population genetics, molecular ecology, biology, evolutionary biology, and systematics of endangered or invasive species. Specifically, genetic studies in invasive species may aim to (1) describe the origin and the temporal and spatial patterns of invasion trough the description of the distribution of the genetic variation (Sakai *et al*. 2001; Allendorf and Lundquist 2003), (2) improve identification, management, prevention and restoration strategies through the understanding of the population dynamics during the process of invasion (Roux and Wieczorek 2008; Handley *et al*. 2011), and (3) study key evolutionary processes at unusual timescales if enough markers or functionaly mutations are assessed (Colautti and Lau 2015; Lau and Terhorst 2015). In this sense, assessment of genetic diversity may predict the population's capacity to adapt to new or changing environmental conditions (Romiguier *et al*. 2014). For instance, invasive species usually present a loss of genetic diversity because of bottlenecks and founder effects, and this loss becomes higher when introductions derived from single episodes and the number of introduced individuals is low (Sakai *et al*. 2001). However, multiple introductions can often be critical to be successful establishment and spread of introduced species, as they may be important sources of genetic variation necessary for adaptation in new environments (Dlugosch and Parker 2008; Suarez and Tsutsui 2008).

1.3.1. Population genetics analyses

Population genetics includes several methodologies to compute genetic diversity statistics and to reconstruct the past demography of a set of populations using allele frequencies distribution among populations (Excoffier and Heckel 2006). We could summarize these approaches in two main types: (1) descriptive genetics for the assessment of genetic variability among individuals, and (2) population structure characterization for analyzing the genetic relationships among populations.

1.3.1.1. Descriptive genetics

Assessment of population genetic diversity parameters is estimated from allele frequencies analyses. It includes concordance with Hardy-Weinberg Equilibrium (HWE) expectations, Linkage Disequilibrium (LD) between pairs of loci, heterozygosity degree, number of alleles or number of polymorphic loci, allelic richness (number of alleles standardized to the smallest sample size in the study) and private alleles detection for any location (Excoffier and Heckel 2006). In addition, recent genetic bottlenecks can be detected from allele frequency data sets by testing for reduced allele number and excess of heterozygosity compared to that expected for an equilibrium population (Cristescu 2015). The knowledge of these parameters allows to infer the genetic diversity level in each location, and also to compare different locations.

1.3.1.2. Population structure inference

Inter-population genetic diversity analyses involve comparisons among populations for inferring genetic structure using estimates of F-statistics and genetic distances. Phylogenetic trees of populations can be constructed based on genetic-distance measures and parsimony networks. These phylogenies are used to infer genetic relationships among native and introduced populations and to investigate patterns of range expansions (Cristescu 2015).

In addition, specialized genetic approaches such as assignment tests allow studying patterns of colonization events (Excoffier and Heckel 2006). These assignment tests are based on clustering genotypes into groups, and involve (1) allocating individuals to previously defined populations and comparing the observed allele frequencies, or (2) allocating them to virtual populations with estimated allele frequencies based on bayesian frameworks (Handley *et al*. 2011). This information allows inferring the genetic structure among a set of populations and, in consequence, estimating colonization routes.

1.3.2. Molecular markers

Molecular markers are DNA fragments that can be identified with a specific localization in the genome and that present genetic variations (polymorphisms) among individuals, species, or higher order taxonomic groups (Liu and Cordes 2004). Choosing the correct molecular marker depends on the desired sensitivity for specific biological questions (Liu and Cordes 2004; Roux and Wieczorek 2008), a feature that is related to polymorphism and mutation rate. Common genetic markers include mitochondrial DNA (mtDNA), Single Sequence Repeat (SSR), and Single Nucleotide Polymorphism (SNP) markers. Other less used markers are Restriction Fragment Length Polymorphism (RFLP), Randomly Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), and allozymes (Sunnucks 2000; Allendorf and Luikart 2009).

1.3.2.1. mtDNA

Mitochondrial DNA is a circular, double-stranded closed DNA molecule located in the mitochondria. In general, the structure (gene order) of mtDNA is conserved among eukaryotes, (Hoelzel 1998), and it encodes 37 genes. Of these, 13 are protein-coding genes, 22 are for transfer RNA (tRNA) and two are for the small and large subunits of ribosomal RNA (rRNA) (Hoelzel 1998).

As a result of the lack of a repair mechanism during replication, mutations accumulate more rapidly in mitochondrial than in nuclear DNA (Avise 1994; Liu and Cordes 2004) with a relatively high mutation fixation rate about 5–10 times higher than nuclear DNA (Arif *et al*. 2011). Genetic drift and no recombination combined with MtDNA is maternal inheritance and haploid, generate an effective population size four times lower than nuclear DNA (Avise 1994) making it more prone to random changes in frequency due to drift. Sequences from known mtDNA can be easily amplified by conserved or specific PCR primers, and polymorphisms in the inner amplified sequence may reveal differences among individuals (Tjensvoll *et al*. 2005).

Thus, mtDNA has been furtherly used for intra- and inter- population variation assessment, species identification, and DNA barcoding in a diverse range of taxa (Avise 1994; Tjensvoll 2005). The fact that mitochondrial DNA is maternally inherited enables genealogical analyses to trace maternal lineage far back in time and it is useful for studying phylogenetic relationships (Avise 1994). Because of these properties, mtDNA has been the most popular marker of molecular diversity in animals over the last three decades (Galtier *et al*. 2009).

However, there are different mtDNA fragments used as molecular markers depending on their characteristics. First, mitochondrial ribosomal RNA markers (12S rDNA and 16S rDNA) are highly conserved and have been used to understand the genetic diversity of higher categorical levels such as phyla or subphyla (12S rDNA) or middle categorical levels such as families or genera (16S rDNA, Arif and Khan 2009). Second, the mitochondrial protein–coding genes are regarded as powerful markers for genetic diversity analysis at lower categorical levels (including families, genera and species) due to their faster evolutionary rates compared to ribosomal RNA genes (Arif *et al*. 2011). Animal mitochondria DNA contains 13 protein-coding genes, but only two, Cytochrome b (cytb) and the cytochrome oxidase I (COI) genes, are frequently used for species identification and biodiversity analysis (Arif *et al*. 2011). Finally, mtDNA contains a non–coding region, the control region (D–loop), having a relevant role in replication and transcription of mtDNA. The D–loop region segments contain larger level of variation than protein–coding sequences because of reduced functional constraints and relaxed selection pressure (Arif and Khan 2009). Sequence analysis of the mtDNA D–loop fragment has been generally used to measure the genetic diversity to establishing phylogenic relationships among individuals in the same species.

1.3.2.2. SSRs

SSR markers, also known as microsatellites, are one of the most used nuclear markers. They are short sequence motifs (ranging from 2 to 6 nucleotides), repeated in tandem usually between 5 and 100 times (Tautz and Schlötterer 1994). The most commonly proposed model for microsatellite mutation is the slipped-strand mispairing of DNA chains during replication (Kelkar *et al*. 2010). According to this model, nascent DNA strand misaligns with the template strand because of the presence of tandem repeats structures. Displacing some tandem repeats in template or new strand may result in addition or deletion of repeat units, respectively (Kelkar *et al*. 2010). Because polymorphisms are generated by the different number of repeats, a large number of alleles may be found in one locus. Theoretically, we could obtain one different allele from each number of tandem repeats in a single locus.

DNA sequences flanking variable fragment among individuals are usually conserved and thus can be used for primer design (Roux and Wieczorek 2008). SSRs are highly variable, consequence of their high mutation rates (10^{-2}-10^{-6} per generation), and multiple alleles are found in a single locus. Therefore, they can resolve population structure, even among closely related populations (Allendorf and Luikart 2009). In addition, they are usually very abundant in the genome and have a co-dominant inheritance (DeWoody and Avise 2000). Utilities of SSRs include (1) genetic mapping, (2) individual DNA identification and parentage assignment, (3)

phylogeny, population, and conservation genetics, (4) molecular epidemiology and pathology, (5) quantitative trait loci mapping, and (6) marker-assisted selection (Chistiakov *et al*. 2006).

1.3.2.3. SNPs

Nuclear SNP markers are point mutations derived from single base substitutions (Liu and Cordes 2004). Theoretically, a single SNP could produce up to four alleles, but they typically are biallelic in nature (Liu and Cordes 2004). They are usually neutral base variants with no changes in the protein sequence. These markers present co-dominant inheritance and they are the most abundant polymorphism in the genome, approximately one SNP is found every 500 bp in many animal species (Allendorf and Luikart 2009). However, their mutation rate at a single base is low (10^{-8} changes per nucleotide per generation, Allendorf and Luikart 2009). Their high quantities and the potential genotyping by high throughput genotyping automated allele detection systems (with the possibility of combining and adding new data to existing sets of genotypes) increase their usability in genetic variability detection, and they are likely to replace SSRs as the marker of choice for many applications.

1.3.3. Genetic analyses of invasive zebra mussel and Asian clam species

There are several genetic studies inferring sources and pathways of colonizations of the zebra mussel. Initial studies using allozyme markers focused on detecting general patterns of genetic structure among North American and European locations identifyed multiple colonization events in North America. In fact, these invasions were originated from previously invaded European locations (Mardsen *et al*. 1995; Lewis *et al*. 2000). In addition, nuclear and mitochondrial markers have been used to also describe that colonizations in North America involved multiple and accumulative founding sources (Stepien *et al*. 2002; Stepien *et al*. 2005). High SSR genetic diversity values found in North America and shared allele frequencies with European populations confirmed multiple North American invasion events from specific European basins (Brown and Stepien 2010) .

MtDNA has also been used to examine the evolutionary history in both native and invasive European areas (Gelembuik *et al*. 2006; May *et al*. 2006). These studies show that all invasive populations in Central-North Europe derive directly from native Ponto-Caspian sea basin. In addition, authors revealed a clear genetic structure in Europe, mainly explained by the existence of specific expansion corridors from the native area to Western Europe in the two last centuries (Bij de Vaate *et al*. 2002). In the Iberian Peninsula, a suggested source for the invasion is the Rhone River in France (Rajagopal *et al*. 2009). This hypothesis would follow the same patern of invasion found in French rivers, a continuous linear spread from northern to southern France (Tarnowska *et al*. 2013).

In contrast, population structure among Asian clams remains unresolved, an uncertainity probably conditioned by their particular asexual reproduction system (Pigneur *et al*. 2012). Whereas within the native range Asian clam lineages can present both sexual and asexual reproduction, in the invasive range only asexual androgenesis has been described (Pigneur *et*

al. 2014a). This asexual reproduction system allows crossed fecundation among lineages, which difficults the genetic comparisons (Pfenninger *et al*. 2002; Hedtke *et al*. 2008).

In this context, genetic studies for the Asian clam have been focusing on phylogenetic analyses rather than population dynamics analyses. In general, three main lineages across the invasive area have been described based on shell morphology and molecular analysis (Pfenninger *et al*. 2002; Hedtke *et al*. 2008). A combination of mitochondrial the COI gene and a 28S nuclear ribosomal DNA fragment is commonly used to identify these lineages, and it has been used to infer some colonization routes as well (Park and Kim 2003; Lee *et al*. 2005; Pigneur *et al*. 2011). On average, mitochondrial analyses suggest that all three invasive lineages present a similar spread history among European locations, whereas in America their range distribution remains restricted for some lineages (Pigneur *et al*. 2014a). However, the comparison of mitochondrial and nuclear markers has allowed the detection of androgenetic genome mismatches likely due to hybridization events among all lineages (Pigneur *et al*. 2014a).

1.3.4. Choosing the best molecular marker

Historically, the most common molecular markers in invasive species studies have been nuclear SSRs and mitochondrial DNA COI gene sequences (Handley *et al*. 2011). Both kinds of markers can be used to describe invasion episodes and for taxonomic identification. The utility of molecular markers can be measured by their Polymorphic Information Content (PIC) index (Botstein *et al*. 1980). PIC value refers to the capacity of detecting polymorphism in a population, based on the global number of detectable alleles among all loci used and their frequencies distribution. All molecular markers can be compared based on their PIC values. MtDNA can present a large number of alleles per locus but the total amount of markers is very limited, thus yielding lower PIC values than those found in nuclear markers such as SSRs or SNPs. The highest PIC values are found in SSR markers due to the larger number of alleles per locus, followed by SNPs because of their abundance (Liu and Cordes 2004). In addition, mtDNA also is less informative than nuclear markers, because it generates incomplete inheritance patterns based only on maternal inheritance. Furthermore, mutation model of mtDNA generates useful information for ancient phylogeographic episodes, but has less power of resolution in inferring recent changes in populations (Avise 2000), thus being a less appropriate marker to be implemented in the study of population dynamics in invasive species.

In contrast, the use of SSRs in population genetics may provide unique genotypes for every individual within a population. For this reason SSRs are also the most common genetic tool for parental analysis and related protocols (Liu and Cordes 2004; Roux and Wieczorek 2008). However, the total number of SSRs that can be used simultaneously in a single study is limited due to this molecular marker type requires their previous time-consuming identification and characterization (Zane *et al*. 2002; Zhan *et al*. 2008), and only a limited number of loci can be multiplexed simultaneously (Guichoux *et al*. 2011). In this sense, SNPs are becoming the marker of choice for several applications, mainly due to their adaptability to automation (Morin *et al*. 2004). However, several SNPs are needed to generate sufficient genetic information. On average, genetic information derived from a similar number of random SSRs

and SNPs, SSRs are four to twelve times more informative (Liu *et al*. 2005; Defaveri *et al*. 2013). Based on the total number of alleles, 100 SNPs could present the same discriminatory power than a set of 10-20 SSRs (Kalinowski 2002). Specifically, in extreme scenarios where moderate levels of population differentiation are present, it can easily be detected by a combination of 30 SSRs whereas a minimum of 80 SNPs are required (Morin *et al*. 2009). By contrast, using only between 6 and 10 SSRs with moderate allelic diversity are required for optimizing population assignment (Bernatchez and Duchesne 2000). Since high-throughput screening of SNPs is not always possible, SSRs might provide the most cost-effective and time efficient means of screening for structure population (Defaveri *et al*. 2013).

1.3.5. SNP and SSR discovery

Traditional methodologies of SSRs isolation require the construction of microsatellite-enriched libraries with tandem repeat probes, and they usually display low efficiencies, ranging from 12 % to less than 0.04 % SSRs out of initial screened sequences depending on the species (Zane *et al*. 2002; Zhan *et al*. 2008). In parallel, traditional SNP detection protocols also involve expensive, labor intensive and time-consuming steps. They are based on resequencing methods producing sequences to be aligned and subsequently compared for polymorphism detection among individuals (Kwok and Chen 2003; Twyman 2005). However, since the availability of large data bases of Expressed Sequence Tag (EST), molecular marker characterization protocols had been facilitated. Screening EST data bases for bioinformaticaly markers detection became less time and cost consuming, and all efforts can be easily focused on validating them for polymorphism detection among individuals.

1.3.6. Massive Parallel Sequencing implementation

Massive Parallel Sequencing (MPS) technologies (also known as Next-Generation Sequencing or NGS) appeared in 2005 with the sequencing technology developed by Margulies *et al*. (2005) and commercialized by 454 Life Sciences. Over the past ten years, other MPS technologies have been developed and all of them have become widely available, reducing costs of DNA sequencing for obtaining large data sets (Mardis 2008a; Ansorge 2009). With MPS technologies, molecular markers isolation and characterization procedures have become easier, faster and cheaper. Moreover, the high power of the *de novo* assembly strategies of MPS approaches allow screening these large databases in search of sets of SSR or SNP markers, even in non-model species when no previous genetic information is available (Everett *et al*. 2011; Gardner *et al*. 2011).

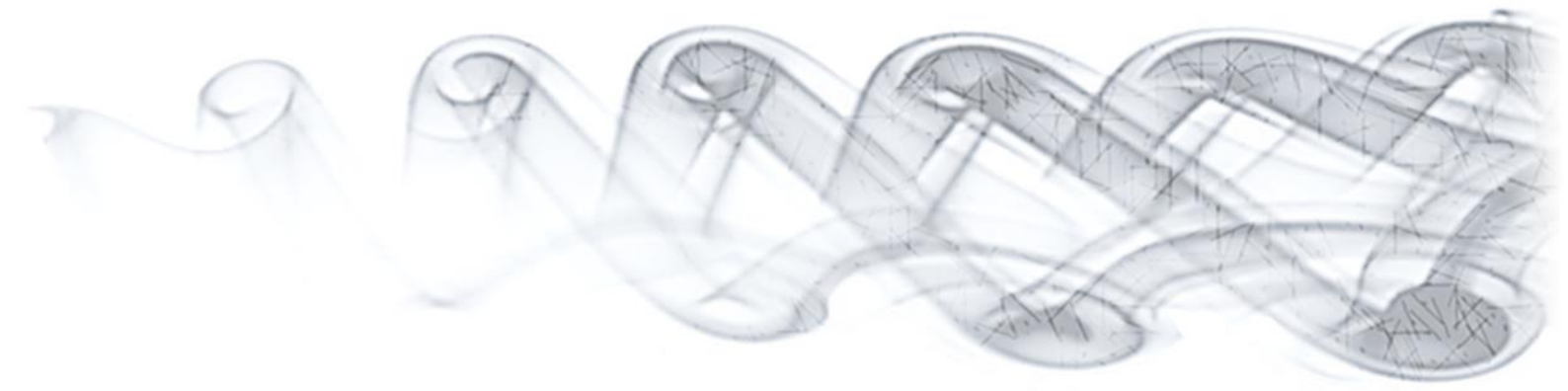
Two MPS methods are the most widely used to identify molecular markers: Roche 454 Pyrosequencing system (http://www.454.com), and Illumina Sequencing by synthesis (http://www.illumina.com). The molecular basis of 454 Pyrosequencing is emulsion PCR (emPCR) and a subsequent Pyrosequencing reaction (Droege and Hill 2008). After initial DNA fragmentation (into 300 – 800 bp size fragments), thousands of DNA strands are massively amplified by emPCR. In this emulsion step, oil bubbles containing all PCR reagents and a single DNA segment attached to an agarose microbead allows the DNA amplification in pararell (See Additional Figure 7.1.1). Next, positive amplified microbeads are isolated, and are

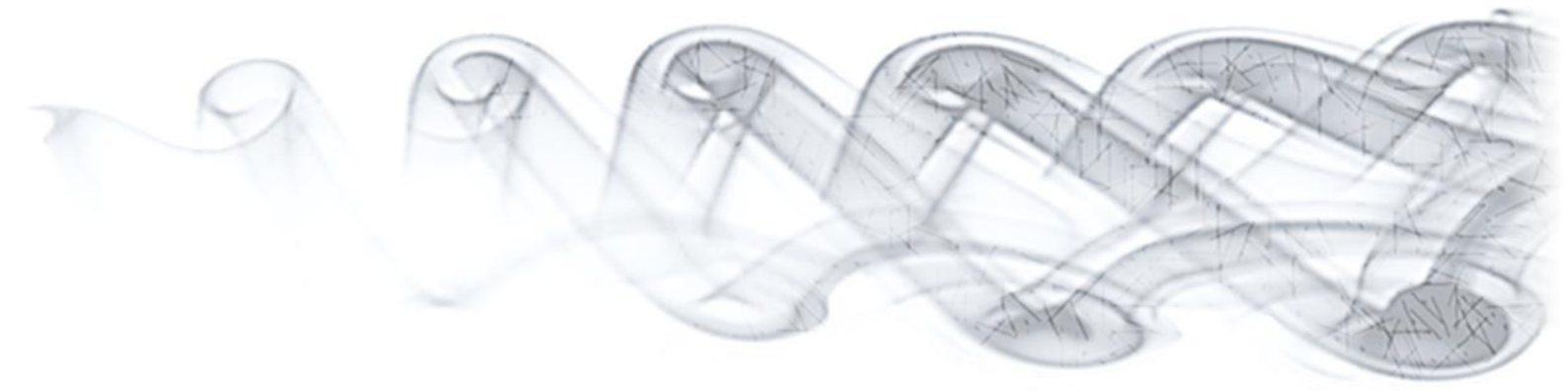
subsequently used for pyrosequencing (Mardis 2008b). At this point, the specific nucleotide incorporation reaction in pyrosequencing in each microbead generates a luminous peak immediately detected by 454 Sequencer Systems (Additional Figure 7.1.1). The major achievement of 454 Pyrosequencing is the longer read sizes obtained compared to other MPS platforms (700-800 bp size). In contrast, major drawbacks involve errors in the sequencing reaction producing false insertion-deletions (InDels) in homopolymeric regions, and the higher per-base costs compared to other systems (Nowrousian 2010).

In contrast, Illumina technology is based on the concept of bridge PCR and subsequent sequencing by synthesis. Here, single DNA fragments are initially joined to a solid surface plate, where they occupy single spots. Sequence linkers attached to the plate join both DNA fragment ends, thus creating a loop structure (Shendure and Ji 2008). Then, this structure is amplified (the bridge PCR reaction) and each DNA fragment ends up forming a specific cluster. These clusters will be the template for the subsequent sequencing reaction, which involves fluorescently labeled nucleotides carrying a terminating/inhibiting group. At the end of each round of the sequencing reaction, there is a cleavage step to remove the terminating/inhibiting group, so a new nucleotide can be added to the sequence during the next cycle (Shendure and Ji 2008). In this sense, each cycle consists in a single-base extension, and cycles are repeated up to 200 times making sequence reads up to 200 bp size from each DNA fragment. When paired-end was lately incorporated in Illumina sequencers, both sides of the bridge PCR fragment were sequenced, doubling the genetic information acquired (Ansorge 2009).The major achievements of Illumina MPS sequencers are the lower per-base costs and the high quantity of sequences generated (more than 40 million) compared to other systems. In contrast, this method presents nucleotide substitution errors during the reaction, needing more sequencing coverage for base confirmation (Mardis 2008b).

Additional MPS platforms have been developed, including SOLiD sequencers (Applied Biosystems) based on ligase enzymes reactions, and Iont Torrent sequencers (Life Technologies) based on the detection of pH changes after nucleotide incorporations (Huang and Marth 2011; Rothberg *et al*. 2011). However, their use for the development of molecular markers has been masked by 454 and Illumina platforms. The larger read sequences of the 454 have allowed this MPS to be the most usually used for new SSRs characterization (Guichoux *et al*. 2011; Zalapa *et al*. 2012). In contrast, lower costs of Illumina have made this MPS the most common for population genetics, ecological genetics and evolution using SNP (Davey *et al*. 2011; Nielsen *et al*. 2011; Shanin *et al*. 2012). With both MPS technologies and subsequent bioinformatic analysis, markers across almost any genome of interest are available to be discovered rapidly and at a low cost, even without a reference genome (Ratan *et al*. 2010; Helyar *et al*. 2011).

2. OBJECTIVES





2. OBJECTIVES

The zebra mussel and the Asian clam have been present in the Iberian Peninsula for several decades, generating high ecological and economic impacts in ecosystems. However, introduction history and colonization routes of these invasive species in the Iberian Peninsula remain mainly unknown. In the present thesis, we focus on understanding possible source(s) of both invasions, and how both species have colonized and spread across Iberian basins based on subsequent genetic procedures.

Specifically, our objectives are:

1. Development of new genetic markers using MPS technologies for the zebra mussel and the Asian clam.

1.1. Identification and characterization of a new set of SSR markers in zebra mussel for population genetics analyses (**chapter 1**).

1.2. Identification and characterization of new SSR markers in Asian clam (**chapter 2**).

1.3. Identification and characterization of new SNPs in zebra mussel and Asian clam species to increase the number of available markers in both invasive species (**chapter 3**).

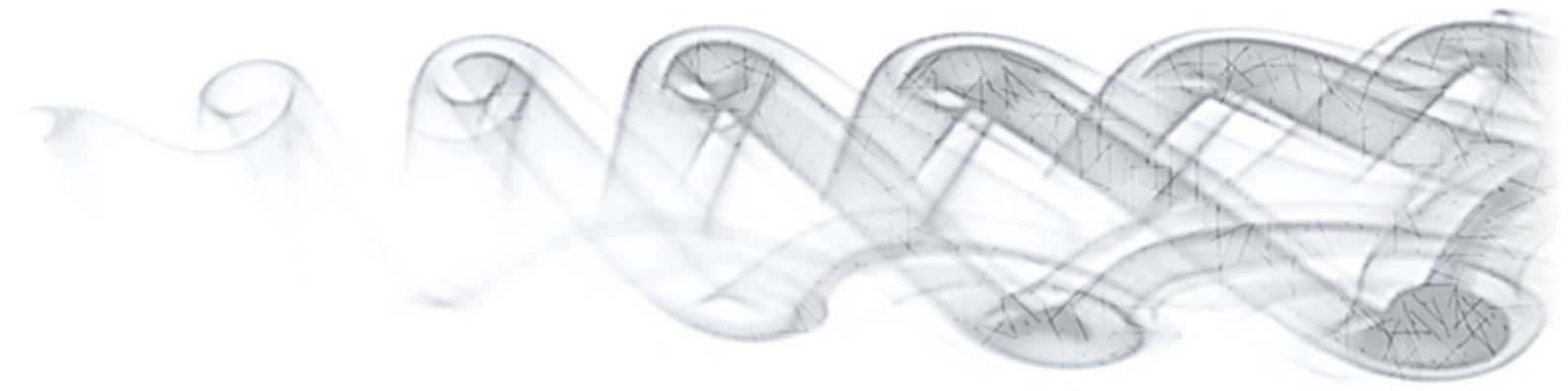
2. Genetic characterization of Iberian populations to understand their invasion history in the Iberian Peninsula.

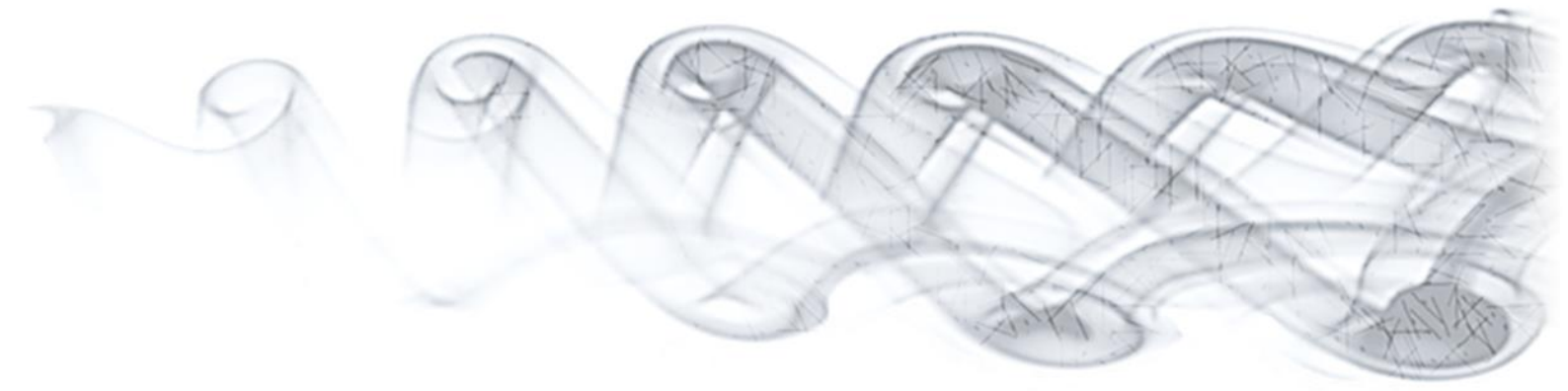
2.1. Genetic characterization of the zebra mussel distribution in the Iberian Peninsula (**chapter 4**).

2.2. Genetic characterization of the Asian clam distribution in the Iberian Peninsula (**chapter 5**).

3. Development a genetic method for monitoring the zebra mussel detection in environmental samples (**chapter 6**).

3. RESULTS





3.1

Using Massive Parallel Sequencing for the development, validation, and application of population genetics markers in the invasive bivalve zebra mussel (*Dreissena polymorpha*)

Peñarrubia L., Sanz N., Pla C., Vidal O. and Viñas J., 2015. **Using Massive Parallel Sequencing for the development, validation, and application of population genetic markers in the invasive bivalve zebra mussel (*Dreissena polymorpha*).** *PLoS ONE* **10**: e0120732.

PLOS ONE

RESEARCH ARTICLE

Using Massive Parallel Sequencing for the Development, Validation, and Application of Population Genetics Markers in the Invasive Bivalve Zebra Mussel (*Dreissena polymorpha*)

Luis Peñarrubia, Nuria Sanz, Carles Pla, Oriol Vidal, Jordi Viñas*

Laboratori d'Ictiologia Genètica, Departament de Biologia, Universitat de Girona, Campus Montilivi, Girona, Spain

* jordi.vinas@udg.edu

OPEN ACCESS

Citation: Peñarrubia L, Sanz N, Pla C, Vidal O, Viñas J (2015) Using Massive Parallel Sequencing for the Development, Validation, and Application of Population Genetics Markers in the Invasive Bivalve Zebra Mussel (*Dreissena polymorpha*). PLoS ONE 10(3): e0120732. doi:10.1371/journal.pone.0120732

Academic Editor: Wolfgang Arthofer, University of Innsbruck, AUSTRIA

Received: October 10, 2014

Accepted: January 26, 2015

Published: March 17, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was carried out within the objectives of the research project CGL200909407 of the Spanish MICINN (Ministerio de Ciencia e Innovación), http://www.idi.mineco.gob.es/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

The zebra mussel (*Dreissena polymorpha*, Pallas, 1771) is one of the most invasive species of freshwater bivalves, due to a combination of biological and anthropogenic factors. Once this species has been introduced to a new area, individuals form dense aggregations that are very difficult to remove, leading to many adverse socioeconomic and ecological consequences. In this study, we identified, tested, and validated a new set of polymorphic microsatellite loci (also known as SSRs, Single Sequence Repeats) using a Massive Parallel Sequencing (MPS) platform. After several pruning steps, 93 SSRs could potentially be amplified. Out of these SSRs, 14 were polymorphic, producing a polymorphic yield of 15.05%. These 14 polymorphic microsatellites were fully validated in a first approximation of the genetic population structure of *D. polymorpha* in the Iberian Peninsula. Based on this polymorphic yield, we propose a criterion for establishing the number of SSRs that require validation in similar species, depending on the final use of the markers. These results could be used to optimize MPS approaches in the development of microsatellites as genetic markers, which would reduce the cost of this process.

Introduction

The zebra mussel (*Dreissena polymorpha*, Pallas, 1771) is a successful invasive bivalve that is native to the brackish and fresh waters that drain into seas of the Ponto-Caspian region (Black, Caspian, and Azov Seas) [1]. This species is now considered one of the World's 100 most invasive species according to the International Union for Conservation of Nature—Invasive Species Specialist Group (IUCN—ISSG) [2]. *Dreissena polymorpha* has several biological attributes that facilitate the success of its invasions, including rapid growth with early sexual maturity, dispersal by larvae, unspecific food preference, and gregarious behavior [3]. The invasive behavior of this species is further enhanced by several transport-related anthropogenic factors, primarily the ballast water of boats that move through inland waterways [4].

The freshwater invasion of *D. polymorpha* is of public interest because of the associated major economic and ecological damage caused by this species. Once individuals are introduced to a new area, high-density aggregations [3] have many adverse socioeconomic and ecological consequences including habitat destruction, the loss of species diversity, and the extinction of native species [5]. The socio-economic effects of dense colonies of *D. polymorpha* include the inability usage of freshwater superficies for recreational purposes and the blockage of artificial water-conducts, among others. The latter creates serious working problems to industries and water supply systems, as well as to the watering and refrigeration systems of hydroelectric, thermic, and nuclear structures [6]. Understanding the population genetic structures in invasive species have revealed their usefulness for inferring source regions, routes of invasion and possible threatened locations by phylogeographic analysis. In recent years, molecular markers have been increasingly used on Dreissenid species in studies of taxonomy, phylogeny, genetic diversity, and phylogeography [7]. The mitochondrial gene of the Cytochrome Oxidase subunit I (COI) and nuclear Amplified Fragment Length Polymorphism (AFLP) markers have been used to examine the evolutionary history of *D. polymorpha* species in Europe [8–9], revealing highly structured populations. Nuclear allozyme markers have been used to analyze the North American invasion indicating that its presence was consequence of accumulative invasion events from Europe [10–11].

Microsatellites [12], also known as Simple Sequence Repeats (SSRs) or Short Tandem Repeats (STRs), are one of the most informative types of molecular markers due to their high variability, consequence of their high mutation rates, and power to resolve population structure, even among closely related populations [13]. Other advantages of microsatellite markers are that they are abundant in the genome, co-dominant and easy to detect by PCR [14–15]. Consequently, microsatellite loci have been the most used genetic markers for population genetic analysis, and they are used in a diverse range of applications, such as parentage analyses, genetic mapping, and conservation genetics [16]. These markers have also proven to be useful for the genetic characterization of invasive species [17–18], where genetic diversity is reduced [19], as well as to infer the sources and pathways of the introduced populations in aquatic ecosystems [20–21]. However, despite the critical situation caused by the expansion of *D. polymorpha* across Europe and North America, only a few microsatellites have been described for this species [22–24] and their analyses have corroborated the results obtained by mitochondrial and allozyme markers [20, 25].

For most non-model species, few or none microsatellite are described, and they must be isolated *de novo*. One of the causes of having limited number of microsatellite loci is consequence of the high cost and labor-intensive approaches required in traditional methodologies of microsatellites isolation [26–27], but are being overcome by new sequencing technologies. For instance, large databases of genomic sequences have been obtained by Massive Parallel Sequencing (MPS) methods (also known as Next-Generation Sequencing or NGS), which could be screened using bioinformatics tools to detect large sets of microsatellite markers [28], even in invasive mollusk species [29].

All of the microsatellites identified in *D. polymorpha* employed traditional hybridization probes. The implementation of new microsatellites by MPS methods may provide new markers for assessing the dispersion capacity and invasion routes of *D. polymorpha*, which would help with the establishment of management strategies to control the invasive potential of this species [30]. Thus, the main objective of this study was to identify and validate new polymorphic microsatellite loci using MPS platforms to increase the number of molecular markers for *D. polymorpha*. This process involved establishing procedures to set up and implement a protocol for the identification and validation of new markers for application in successive non-model species and organisms. The second objective was to determine the number of

microsatellite markers needed for several genetic applications. The newly developed markers are expected to be used in subsequent studies focusing on the population genetic structure of *D. polymorpha*, to determine its invasion history and dispersal routes through Europe. This information could potentially help predict possible new areas of invasion, which would allow the implementation of preventative management to prevent accidental introduction.

Methods

Phase A: MPS Data Processing

DNA Extraction and Sample Selection. Total DNA was isolated separately from three *D. polymorpha* individuals by two phenol-chloroform and ethanol precipitations, as described in [31]. RNase treatment was performed between the two DNA precipitations by adding 2 μL of RNase (10 μL/ml) in each DNA tube, and implementing an incubation step of 30 min at 37°C. DNA quality was analyzed by 0.6% Agarose Gel Electrophoresis, and DNA quantity was evaluated using the Nanodrop spectrophotometer ND-1000 (Thermo Fischer Scientific, Inc. Waltham, Massachusetts, U. S. A.) (Fig. 1). Finally, the best of the three DNA extractions was sent for the massive DNA sequencing at CRAG (Centre de Recerca en Agrigenòmica, Barcelona, Spain).

Sequencing by the 454 GS System. DNA was sequenced using a 454 GS FLX platform using 1/8 of a plate. Reads were assembled using GS DE NOVO ASSEMBLER software, version 2.5.3. (Fritz Hoffmann-La Roche Ltd. Basel, Switzerland) (Fig. 1). Two strategies were used to compare different yield when recovering reads into the contigs. First, we used the default align and assemble parameters: minimum read length = 20 base pairs (bp); seed step = 12 bp; seed length = 16 bp; seed count = 1; minimum match length = 40 bp; overlap minimum match identity = 90%; alignment identity score = 2; alignment difference score = 3; and minimum output sequence size = 100 bp. Second, we ran an assembly to optimize the number of assembled reads into the contigs. We used the same assembled parameters and included the options of "Large or Complex Genome" to specify that the species was eukaryotic and "Heterozygotic Mode," due to the diploid nature of *D. polymorpha*.

Phase B: SSR Isolation and Primer Design

Identification and Microsatellite Selection of Tandem Repeats. We used TANDEM REPEATS FINDER (TRF) software version 4.04 [32] to identify simple tandem repeats (Fig. 1). TRF target SSR sequences were aligned to a predefined library of consensus tandem repeats sequences. We selected alignments of 30 nucleotides, with values of match = 2, mismatch = 7, and indels = 7, and a maximum period size of five nucleotides for each microsatellite. This final parameter excludes hexanucleotide SSRs. Homopolymeric repeats were discarded, due to the extremely high number of sequencing errors made within homopolymeric runs using the 454 GS FLX platform [33].

Tandem repeat structures identified by TRF were pruned to obtain the Potential Amplifiable Loci (PAL), after removing structures composed of less than five tandem repetitions and regions with less than 30 base pairs in each of the flanking sequences and with positive primer annealing positions. In addition, we considered two neighboring SSRs separated by less than 100 nucleotides as the same locus (a compound microsatellite).

Primer Design. PCR primers were designed (Fig. 1) using PRIMER3 software, version 0.4.0. [34]. BLASTN [35] analysis was run to discount possible homologies that could produce inespecifities in the PCR. Predicted amplicons spanned three size ranges (100–150, 200–300, and 350–450 base pairs) to facilitate possible implementation in future multiplex PCR without overlapping allele sizes. All forward PCR primers contained a sequence of 19 additional

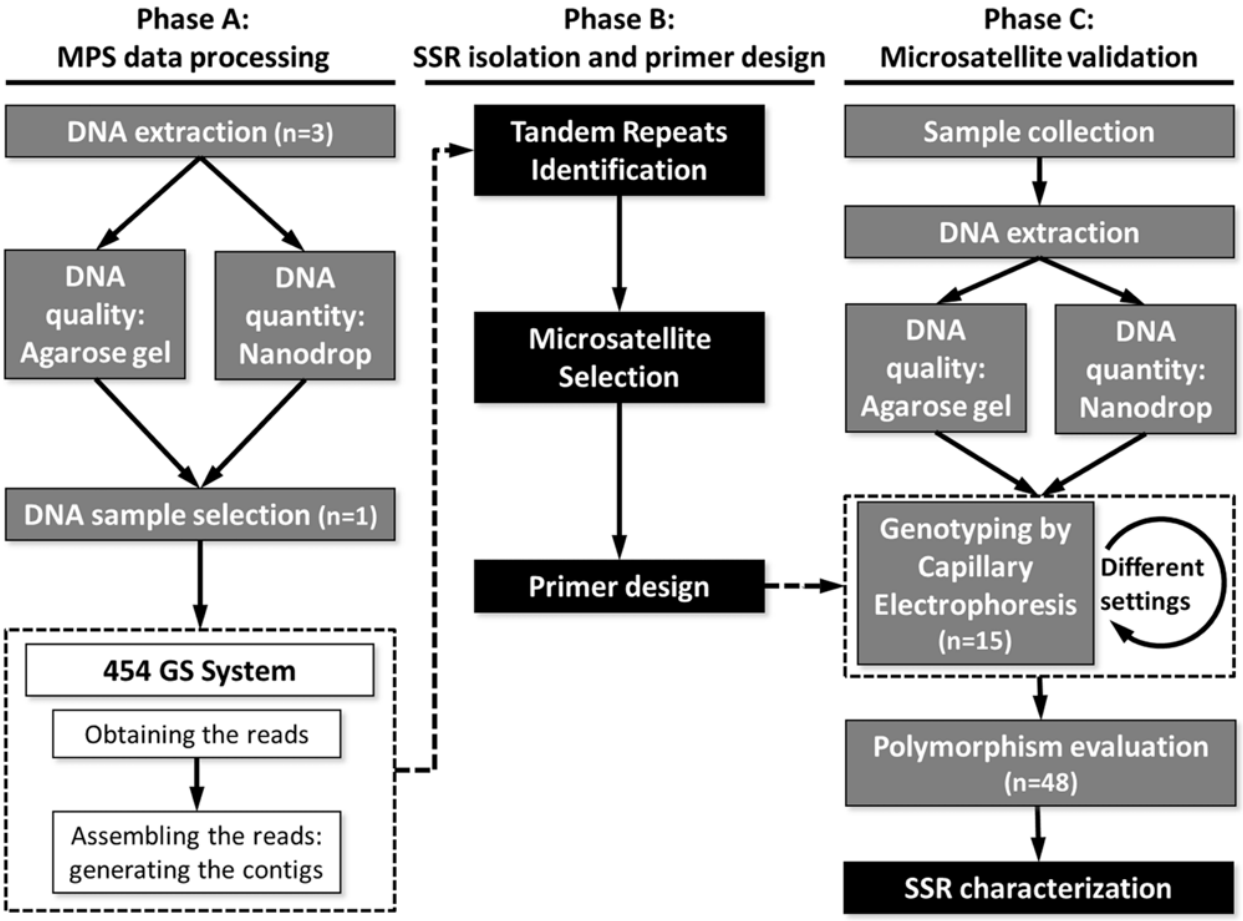


Fig 1. Experimental design and protocol pipeline. Gray steps: Laboratory protocols. White steps: 454 GS System. Black steps: Bioinformatic analysis.

doi:10.1371/journal.pone.0120732.g001

nucleotides (CACGACGTTGTAAAACGAC) at each 5′-end. A labelled primer consisting of the same sequence with a 6-FAM fluorochrome was included in each genotyping PCR, as described in Schuelke *et al.* [36].

Phase C: Microsatellite Validation

Sample Collection and DNA Extraction. Forty-eight *D. polymorpha* individuals were collected from six representative locations where they had been introduced in north-eastern Spain (S1 Table). Samples were collected and manipulated under permits provided by the Catalan Water Agency of the Agriculture, Fisheries, Food and Environment Department of the Autonomous Community of Catalonia and the Hydrographic Confederation of the Ebro River of the Agriculture, Food and Environment Ministry of Spain. All work was performed in compliance with and approved by the Ethics Committee of the University of Girona and met the requirements stated by the Spanish (RD53/2013) and Catalonian (D214/1997) laws of animal care and experimentation. The shell was removed, and the whole muscle bodies were preserved in 96% ethanol until processing. DNA extraction was performed using the Real Pure DNA Extraction Kit (Durvitz, Valencia, Spain). Quality was checked using 0.6% Agarose Gel Electrophoresis, and DNA quantity was measured with a Nanodrop spectrophotometer ND-1000 (Thermo Fischer Scientific, Inc. Waltham, Massachusetts, USA) (Fig. 1).

Genotyping and Polymorphism Evaluation. The PCR mix was prepared in a final volume of 20 μL, containing 1X Buffer (BIOLINE), 1.5 μM $MgCl_2$, 0.8 mM dNTPs, $5x10^{-3}$ μM primer Forward, 0.2 μM primer Reverse, 0.2 μM 6-FAM dye-probe, and $2.5x10^{-2}$ u/μL Taq polymerase (BIOLINE). DNA samples (25–100 ng) were added to each PCR mix. The thermal cycles consisted of an initial denaturing step of 3 min at 94°C, followed by 35 cycles of denaturing at 94°C for 30 s, annealing at 50°C or 60°C for 90 s, and extension at 72°C for 90 s, and a final extension period of 10 min at 72°C. Negative controls were included in all PCR runs to confirm that there was no cross-contamination. PCR products were read in a 3130 Genetic Analyzer (Applied Biosystems).

Validation analysis of PALs was conducted in two steps (Fig. 1). First, PCR screening of the putative SSRs was completed to check the specificity of the primers in 15 *D. polymorpha* individuals from four different locations separated in range of 167 and 562 km. If these initial PCR conditions failed or unspecific results were obtained, a second attempt was made using the same samples, but increasing the annealing temperature to 60°C. Successfully validated loci were genotyped in 48 *D. polymorpha* individuals to assess polymorphism and allele richness.

SSR Characterization. We evaluated the presence of null alleles in each locus using MICRO-CHECKER software, version 2.2.3 [37]. The observed (H_o) and expected (H_s) heterozygosities were assessed for each new polymorphic locus for conformance to Hardy-Weinberg and Linkage disequilibrium expectations by the exact test implemented in GENEPOP 4.0 software [38]. We calculated the Polymorphism Information Content (*PIC*) index to estimate the information values of each microsatellite marker [39]. Microsatellite sequences were added to the GenBank Database with accession numbers [GenBank: JQ812984-JQ812997].

We also tested the 14 new polymorphic microsatellites in the preliminary characterization of the genetic structure of *D. polymorpha* in the Iberian Peninsula. We analyzed 48 *D. polymorpha* individuals from six representative locations where they are present in north-eastern Spain (S1 Table). The Hardy-Weinberg equilibrium (*HWE*) of genotypic distributions was measured at each site by the exact probability test [40] using GENEPOP software version 4 [38]. Benjamini & Yekutieli [41] False Discovery Rate (FDR) was applied to each test to adjust the significance levels for multiple simultaneous comparisons. Genetic diversity within each study site was estimated from direct counts as the number of alleles per locus (*A*), the estimated expected heterozygosity (H_s), and allelic richness (A_r) from allele frequencies using FSTAT 2.9.3 [42]. In addition, pairwise population differentiation (F_{ST}) [43] and significance values were calculated using FSTAT software. The number of genetically homogeneous groups (*K*) was estimated among all sampled collections using the Bayesian Markov Chain Monte Carlo approaching method of STRUCTURE software version 2.3.2 [44]. Runs for each possible *K* (1 to 6) were repeated five times. A burn-in period of 20,000 steps, followed by 100,000 Monte Carlo replicates, was simulated with the model of independent allele frequencies. The optimal *K* value was selected following the recommendations of Pritchard et al. [44], according to the posterior probability of the data. Finally, patterns of gene diversity distribution among *D. polymorpha* collections provided by the genetic results were estimated by Analyses of Molecular Variance (AMOVA) conducted in ARLEQUIN 3.5 [45]. In the AMOVA model, all six locations were grouped into two hydrographic basins (i. e., the Ebro River and Llobregat River). We tested whether genetic diversity was partitioned into three levels: within locations, among locations within river basins, and among river basins.

Results

Phase A: MPS Data Processing

454 GS System Sequencing Results. The 454 GS FLX run generated 110,593 reads (read range size = 24–691 bp, mean read size = 331 bp), with a total of 36,554,817 bases (Table 1). GS

DE NOVO ASSEMBLER software generated two assembled outputs based on the selected parameters (see Methods section). In the first output using the default parameters, 31,330 reads (28.33% of total) were assembled into 2,326 contigs (contig range size = 100–8,697 bp, mean contig size = 457 bp, N50 = 825 bp) and 70,208 singletons (Table 1). In comparison, the second output using modified parameters assembled 35,374 reads (31.99% of total) into 3,885 contigs (contig range size = 100–8,717 bp, mean contig size = 421 bp, N50 = 860 bp) and 68,463 singletons (Table 1). Since one of the main objectives of this study was to find as many *SSR* structures as possible, we selected the second assembly with the modified parameters "Large or Complex Genome" and "Heterozygotic Mode," because it generated a higher N50 value with a greater number of long contigs of the same sequence quality.

Phase B: SSR Isolation and Primer Design

Identification, Microsatellite Selection, and Primer Design of Tandem Repeats. *TRF* software identified a total of 299 single SSR structures. After removing homopolymers, 288 SSRs were identified with an overall estimated density of 0.18 SSRs/Kb. The SSRs were distributed in 6.60%, 37.85%, 39.58%, and 15.97% of di-, tri-, tetra-, and pentanucleotides respectively (S2 Table). Of these SSRs, 93 (32.29% of the 288 initial SSRs identified that had a density of 0.06 SSRs/Kb) were identified as Potential Amplifiable Loci (PAL), after removing sequences with less than 5 units of repetition or with less than 30 nucleotides at each end of the tandem repeat, assuming that repeat structures separated by less than 100 base pairs were the same locus, and with positive designed PCR primers (Table 1 and S3 Table).

Phase C: Microsatellite Validation

Genotyping and Polymorphism Evaluation. All 93 PALs were tested for PCR amplification in 15 representative *D. polymorpha* individuals out of the 48 individuals collected from six locations in the North-Eastern Iberian Peninsula (S1 Table). Twelve PALs failed to produce PCR products. The remaining 81 loci, which generated positive PCR amplification with a product within the expected size range, were genotyped in the set of 48 mussel samples. These loci represented 28.13% of the 288 initially identified SSRs (Table 1). Fourteen of these 81 loci showed polymorphism (Table 2). Blast analysis of these new 14 polymorphic microsatellites resulted in no matches with previously published microsatellite markers in *D. polymorpha*.

SSR Characterization. MICRO-CHECKER software detected null alleles or scoring errors in three (*Dp2*, *Dp43*, and *Dp72*) of the 14 microsatellite loci (Table 2). The *Dp2* locus presented null alleles at three of the locations, whereas *Dp43* and *Dp72* presented null alleles at just one location. The number of alleles detected at each locus ranged from two (*Dp7* and *Dp72*) to 13 (*Dp31*). Estimates of gene diversity and the *PIC* index varied among loci. The lowest value of expected heterozygosity was obtained for *Dp2* (H_s = 0.20), while the highest value was obtained for *Dp1* (H_s = 0.83). The lowest *PIC* index was in *Dp68* (*PIC* = 0.19), while the highest was in *Dp31* (*PIC* = 0.83) (Table 2). Only two markers (*Dp1* and *Dp2*) presented significant deviations from Hardy-Weinberg equilibrium at the 5% significance level, and all loci appeared to be unlinked, according to the lack of gametic disequilibria in each location (Table 2).

These validated microsatellites were used for the preliminary analysis of the genetic population structure of *D. polymorpha* in the Iberian Peninsula. All 48 individuals collected from the six locations were used, with a sample size of eight individuals from each sampling site (five locations from the Ebro River and one from the Llobregat River; S1 Table). All 14 microsatellite loci were polymorphic at all six locations. Analyses of allele frequencies (S4 Table) presented deviations from *HWE* in only one location at Ebro River. After adjusting for differences in sample size, permutation tests demonstrated lower average allelic richness (A_r) and genetic

Table 1. Summary of MPS results and of polymorphic microsatellites identification.

		Output description	Default assembly	Modified assembly
Phase A	Sequencing step	Reads number	110,593	
		Largest read size (bp)	691	
		Shorter read size (bp)	24	
		Reads mean size (bp)	331	
		Sequenced bases (Kb)	36,555	
	Assembling step	Aligned reads (of the total right reads)	31,330 (28.33%)	35,374 (31.99%)
		Aligned bases (Kb) (of the sequenced bases)	7,620 (20.84%)	9,030 (24.71%)
		Assembled reads	18,758	20,570
		Contigs number	2,326	3,885
		Bases number into contigs (kb)	1,063	1,635
		Largest contig size (bp)	8,697	8,717
		Shorter contig size (bp)	100	100
		Contig mean size (bp)	457	421
		Contig N50 value (bp)	825	860
		Singletons number	70,208	68,463
Phase B		Initial *SSR* identified	-	288
		Potential Amplifiable Loci (% initial SSR)	-	93 (32.29%)
Phase C		Positive PCR amplification (% initial SSR)	-	81 (28.13%)
		Polymorphic markers (% SSR)	-	14 (4.86%)

Phase A: MPS data processing; Phase B: SSR isolation and primer design; Phase C: Microsatellite validation

doi:10.1371/journal.pone.0120732.t001

diversity (H_s) in the Llobregat River (A_r = 2.06; H_s = 0.50) compared to the Ebro River basin (A_r = 2.14–2.33; H_s = 0.53–0.61). The lowest number of total alleles (A) across all 14 loci also corresponded to the Llobregat River (A = 43) in comparison to the Ebro River (A = 46–50) (S1 Table).

Estimated average genetic differentiation (F_{ST}) among locations was F_{ST} = 0.044. None of the pairwise F_{ST} values were significant after FDR correction (S5 Table). The Bayesian analyses of STRUCTURE showed an *ln P* value that was lower for K = 1 compared to for K = 2 and K = 3 (K = 1, mean *ln P* = -1429.80; K = 2, mean *ln P* = -1473.96; K = 3, mean *ln P* = -1582.28). Following the recommendations of Pritchard et al. [38], these results distinguished only one genetically homogeneous group of populations for all analyzed locations (K = 1). Finally, the AMOVA test revealed similar genetic variance between the tested river basins (3.20%) and among locations within river basins (3.89%). Furthermore, genetic variance was not significant (P = 0.062) in the comparison between the two river basins.

Discussion

New Microsatellites in *D. polymorpha*

This study presents a set of new polymorphic microsatellites in *D. polymorpha* (Table 2), which, together with previously described microsatellites [22–24], represent a useful tool for investigating genetic variation within and among populations of this species.

This study is the first to use MPS platforms to discover new SSRs in *D. polymorpha*. Interestingly, our approach yielded a higher number of initially identified SSR structures (288 SSRs in this case) compared to the average number of SSR structures identified using hybridization methodologies (mean = 47 SSRs, SD = 18.52) (Table 3). However, the polymorphic yield (ratio

Table 2. Polymorphic microsatellites description.

Locus	Repeat motif	*S*	Primer Sequences 5' → 3'	*Ta*	*A*	H_o/H_s	*PIC*	*NA*	*LD*	GenBank Accession
Dp1	$[ATA]_{21}$	279–299	*F*: GGATTTTTCTCCCGTGGAAT	60	5	0.83 / 0.81	0.76	No	No	JQ812984
			R: CGGTAGCGTTCTCTTCACAA							
Dp2	$[TGA]_{17}$	410–440	*F*: GCTACCGGAGCTCAACCTAA	60	9	0.20 / 0.66	0.62	Yes	No	JQ812985
			R: ACGTCGAACCCTGTCAAAAA							
Dp7	$[TAA]_{11}$ 4 $[ATT]_5$ 48 $[TAT]_{11}$ 71 $[ATA]_{14}$	411–416	*F*: GGAATACCGGGTGCTGTAGA	50	2	0.28 / 0.49	0.37	No	No	JQ812986
			R: GACGTGCGTCACAATAGGTG							
Dp30	$[TTG]_7$ 57 $[GTT]_6$	214–234	*F*: GCGTTGGTGTTGTGTACGTC	60	6	0.58 / 0.67	0.62	No	No	JQ812987
			R: CTGAGCATCTCACCGTCAAA							
Dp31	$[ATT]_{13}$	231–306	*F*: CGAGTTTCTTGCACGTTTCA	50	13	0.79 / 0.86	0.83	No	No	JQ812988
			R: TGTTATTTTAAGAAGGCCACATTG							
Dp39	$[GGCG]_{11}$	385–393	*F*: GACGTCATGGTTCTGAATGG	50	3	0.56 / 0.66	0.58	No	No	JQ812989
			R: CCGGACAAGCTCATTTATGG							
Dp42	$[GTTG]_9$	243–256	*F*: TCGCTTAACCTGACCAGTGA	50	3	0.60 / 0.59	0.52	No	No	JQ812990
			R: CCAAATATCAAGTTGCCTATCTTCA							
Dp43	$[TTA]_{10}$	224–259	*F*: TTGCTCATGATGAAATATGATGT	50	5	0.34 / 0.38	0.35	Yes	No	JQ812991
			R: ATGCGTTTCACTTTGGCATC							
Dp44	$[GACC]_8$	139–153	*F*: CCCCAAGCGTCTTGAGTATC	50	5	0.60 / 0.67	0.60	No	No	JQ812992
			R: TCCTGCCAAGCATGTATGAG							
Dp68	$[TGTTC]_5$	291–297	*F*: TGCTACACACCGTATTTGCTG	50	3	0.23 / 0.21	0.19	No	No	JQ812993
			R: ACACGTGGATGGTGTGAAGA							
Dp72	$[GGTA]_8$	378–382	*F*: TGCACACACATCTTGACCTG	50	2	0.31 / 0.41	0.32	Yes	No	JQ812994
			R: GCTGAAGGCACAACATTTGA							
Dp74	$[CGTC]_9$	332–360	*F*: ATCCCCTCAAGACGTTTCCT	60	3	0.57 / 0.58	0.49	No	No	JQ812995
			R: ACCATACCGGTGGCATAAAA							
Dp86	$[CGTC]_5$	306–315	*F*: GCAAAGGGAGAAAACTGCAC	50	4	0.73 / 0.71	0.65	No	No	JQ812996
			R: CACTGTCACCGTCGCACATA							
Dp89	$[CGTC]_8$	260–284	*F*: TTTTCACACAGCAGCCAAAG	50	4	0.56 / 0.53	0.42	No	No	JQ812997
			R: TGAGAAATAGCCCGGACAAA							

Repeat motif, size range in base pairs (*S*), forward (*F*) and reverse (*R*) (5′– 3′) sequences, annealing temperatures in°C (*Ta*), number of alleles (*A*), Observed (H_o) and Expected (H_s) heterozygosities, *PIC* index, presence of null alleles (*NA*), linkage disequilibrium (*LD*) and GenBank accession number for the 14 polymorphic microsatellites loci tested on 8 individuals for each of the 6 populations (n = 48) in *D. polymorpha*.

doi:10.1371/journal.pone.0120732.t002

Table 3. Microsatellites described in *Dreissena polymorpha*.

		Initial SSR identified	PAL loci (% of initial SSR)	Polymorphic microsatellites	Polymorphic yield	Reference
Traditional techniques		65	16 (24.62%)	5	31.25%	[18]
		48	8 (16.67%)	8	100.00%	[19]
		28	8 (28.57%)	5	62.50%	[20]
	Mean	47	10.67 (23.40%)	6	64.58%	
	SD	18.52	4.62	1.73	34.42	
This study		288	93 (32.29%)	14	15.05%	This study

Initial SSR identified, PAL loci, Polymorphic microsatellites, and Polymorphic yield in all studies describing microsatellite loci in *D. polymorpha*. *SD*: Standard deviation. PAL: Potential Amplifiable Loci. Polymorphic yield is the ratio between the numbers of polymorphic microsatellites divided by the number of PCR validated loci.

doi:10.1371/journal.pone.0120732.t003

between the number of polymorphic microsatellites divided by the number of PCR validated loci) of hybridization techniques has an average of 64.58% of final polymorphic markers, which is substantially higher than the 15.05% obtained in our analysis using MPS (Table 3). Yet, the final number of polymorphic microsatellites obtained by MPS is greater than the expected total number of microsatellites obtained by hybridization techniques (Table 3).

Microsatellite isolation and validation in bivalve species is challenging for several reasons. For instance, we found about180 microsatellites per Mb in *D. polymorpha*. This value is situated in the lower range of 50–1,500 microsatellites/Mb expected in eukaryotes [46]. This phenomenon is probably related to the large genome of bivalves, particularly for *D. polymorpha* (1.70 ± 0.03 pg, about 1.6 Gb) [47–48], combined with the suspected low frequency of microsatellites in the bivalve genome [26, 49]. In addition, PCR design for microsatellite amplification is not always efficient because of the high number of repetitive elements in the flanking regions of mollusks [50]. On average, invertebrates require over twice as many sequences to obtain the same number of useable loci compared to other taxa, like plants and vertebrates [28].

In the current study, we obtained a slightly lower values for the average number alleles per locus (A = 4.79) and mean gene diversity (H_s = 0.56) compared to the value obtained by previous studies analyzing the population genetic structure of *D. polymorpha* (A = 12.70, H_s = 0.84, 5 *SSRs*, [25]; A = 35.09, H_s = 0.61, 11 microsatellites; [20]). The present study focused on validating and obtaining a first approximation of the usefulness of the developed markers in a very restricted number of *D. polymorpha* individuals (n = 48) compared to the sample sizes used by previous studies (n = 309 [25] and n = 386 [20]). This difference may partly explain the lower genetic diversity detected here. However, the overall genetic diversity obtained here was adequate for use in further genetic studies.

Microsatellites in Population Genetics Analyses and Other Applications

Microsatellite markers are the most used genetic markers for population genetic analysis, due to their high variability and power to resolve population structure, even among closely related populations [13]. To optimize the laboratory work in population genetic analyses, it is crucial to determine the minimum number of microsatellite markers used. Then, two issues must be considered: 1) how many microsatellites are needed to achieve a specific objective; and 2) how much effort is needed to develop the desired number of polymorphic microsatellites using MPS platforms. Chistiakov *et al.* [16] classified six main applications of microsatellites (Table 4): 1) genetic mapping; 2) individual DNA identification and parentage assignment; 3)

Table 4. Microsatellite validation effort depending of the application.

Genetic application using microsatellite markers	*n*	Range	Mean (*SD*)	Median	Initial SSR markers to be validated
Genetic mapping	19	31–2000	337.89 (442.37)	209	1388.70
Individual DNA identification and parentage assignment	15	1–14	7.73 (4.03)	8	53.16
Phylogeny, population and conservation genetics	30	1–33	8.77 (6.32)	7	46.51
Molecular epidemiology and pathology	26	1–227	26.15 (49.56)	9	59.80
Quantitative trait loci mapping	13	15–428	96.54 (77.38)	77	511.63
Marker-assisted selection	6	1–7	4.17 (1.94)	4	26.58

Studies analyzed (*n*), SSR markers used and validation effort to get markers required in each genetic application using microsatellite markers. Validation effort was determined using the median values of the number of SSR used in each microsatellite-containing analysis according with 104 studies revised in Chistiakov et al. [24]. The number of locus required to be validated were calculated using the median of SSR analyzed and the polymorphism yield (15.05%) calculated in this study.

doi:10.1371/journal.pone.0120732.t004

phylogeny, population, and conservation genetics; 4) molecular epidemiology and pathology; 5) quantitative trait loci mapping; and 6) marker-assisted selection. However, the authors did not infer the average number of microsatellites required for each specific objective. To gather this information, we reviewed all 104 studies included in the analysis by Chistiakov *et al.* [16] (S6 Table). We used the median value to avoid bias caused by extreme values present in some of the applications. As expected, genetic mapping requires the highest number of microsatellites (median = 209), whereas a median of just four microsatellites is required for marker-assisted selection studies. In molecular epidemiology and pathology, only one microsatellite is necessary when the diagnostic marker associated to a metabolic disorder is known, while several (hundred) markers are used to search for new diagnostic markers. For population genetic studies, which are the target application for *D. polymorpha*, a median of seven microsatellites is required (Table 4).

This information could be used to determine the validation effort required for a specific task. Based on the median value of SSR markers used in each genetic application with the polymorphic yield (15.05%) obtained in this study, we could estimate the number of initial SSR markers that would require validation in mollusk species depending on the desired genetic study (Table 4). Thus, for a population genetics study of eight microsatellites about 54 SSR markers must be validated. Then, the set of 14 polymorphic microsatellite markers developed on *D. polymorpha* clearly exceeds the median number of microsatellites required for genetic population analyses.

A preliminary analysis on the genetic structure (using a sample size of eight individuals per location) of these microsatellites in the Iberian Peninsula failed to show any population structure, but demonstrated the usefulness of these new markers. Our results suggest that all *D. polymorpha* individuals in the Iberian Peninsula belong to the same genetic population supporting a previous preliminary genetics study using a single microsatellite marker [51]. Since its first citation in Iberian Peninsula in the lower Ebro in 2001 [52], the expansion of *D. polymorpha* could be caused by a combination of a free-swimming larval dispersion and human mediated, particularly in ballast waters in vessels. Nevertheless, further analysis with a larger sample size from more locations, and even including some of the previously described microsatellites [22–24], is crucial to confirm this result.

Conclusions

In summary, this study identified and validated 14 new polymorphic microsatellites in *D. polymorpha*, which, added to those developed by Naish and Gosling, Feldheim *et al*, and Thomas

et al [22–24], increase the number of working microsatellites in this invasive species. Our results indicate that methodology based on MPS platforms and bioinformatic analysis could be used to identify a large number of markers, supporting its suitability in microsatellite isolation. This study showed that MPS approaches may be used to optimize to the development of microsatellites as genetic markers, minimizing the cost of this process.

In addition, we conducted a preliminary study of the genetic structure of *D. polymorpha* at six locations in Spain using all 14 new microsatellite markers. Our results showed that all six locations represent the same genetic population; however, further analysis with a larger sample size from more locations, along with more microsatellites, is crucial to confirm this result. In conclusion, this study demonstrates the usefulness of these new markers for population genetic analyses in *D. polymorpha*.

Data deposition

The 454 GS FLX reads of *D. polymorpha* were submitted to the NCBI Sequence Read Archive under accession number SRX803533. De novo assembly was also submitted in DDBJ/EMBL/GenBank Whole Genome Shotgun project when contigs size was longer than 500 nucleotides under accession number JWHF00000000. Monomorphic microsatellite sequences were submitted in Genbank with accession numbers KP274952–KP275018, and polymorphic microsatellite sequences with accession numbers JQ812984–JQ812997.

Supporting Information

S1 Table. Samples analyzed in this study with their genetic variability statistics.
(PDF)

S2 Table. Number (and percentage) of microsatellite types based on the size of the repeat motif.
(PDF)

S3 Table. Description of the 93 validated microsatellites.
(PDF)

S4 Table. Allele frequencies for the 14 polymorphic microsatellite loci in the six locations.
(PDF)

S5 Table. Pairwise FST values among populations.
(PDF)

S6 Table. Number of microsatellite markers used in the different genetic analyses.
(XLSX)

Acknowledgments

We are indebted to Gisela Mir from CRAG Laboratory for her help in obtaining MPS data, and Jesús Scrofa for his collaboration in data analysis. The authors thank Cesca Rivas for assistance with laboratory procedures.

Author Contributions

Conceived and designed the experiments: LP NS CP OV JV. Performed the experiments: LP NS. Analyzed the data: LP OV JV. Contributed reagents/materials/analysis tools: CP JV. Wrote the paper: LP JV.

References

1. Kinzelbach R. The main features of the phylogeny and dispersal of the zebra mussel *Dreissena polymorpha*. In: Neumann D. and Jenner HA, editors. The zebra mussel *Dreissena polymorpha*, ecology, biological monitoring and first application in water quality management. Stuttgart, Germany: Gustav Fischer Verlag; 1992. pp. 5–17.
2. Lowe S, Browne M, Boudjelas S, De Poorter M. 100 of the World's worst invasive alien species: a selection from the Global Invasive Species Database. Aliens. 2000; 12: 1–12.
3. Bij de Vaate A, Jazdzewski K, Ketelaars HAM, Ollasch S, Van der Velde G. Geographical patterns in range extension of Ponto-Caspian macroinvertebrate species in Europe. Can J Fish Aquat Sci. 2002; 59: 1159–1174.
4. Johnson LE, Ricciardi A, Carlton JT. Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. Ecol Appl. 2001; 11: 1789–1799.
5. Karatayev AY, Burlakova LE, Padilla DK. Zebra versus quagga mussels: a review of their spread, population dynamics, and ecosystem impacts. Hydrobiologia. 2015; 746: 97–112.
6. Duran C, Lanao M, Perez L, Chica C, Anadon A, Touya V. Estimación de los costes de la invasión del mejillón cebra en la cuenca del Ebro (período 2005–2009). Limnetica. 2012; 31: 213–230.
7. Lopes-Lima M, Teixera A, Froufe E, Lopes A, Varandas S, Sousa R. Biology and conservation of freshwater bivalves: past, present and future perspectives. Hydrobiologia. 2014; 735: 1–13.
8. May GE, Gelembuik GW, Panov VE, Orlova MI, Lee CE. Molecular ecology of zebra mussel invasions. Mol Ecol. 2006; 15: 1021–1031. PMID: 16599964
9. Rajagopal S, Pollux BJA, Peters JL, Cremers G, Moon-van der Staay SY, van Alen T, et al. Origin of Spanish invasion by the zebra mussel, *Dreissena polymorpha* (Pallas, 1771) revealed by amplified fragment length polymorphism (AFLP) fingerprinting. Biol Invasions. 2009; 11: 2147–2159.
10. Marsden JE, Spidle A, May B. Genetic similarity among zebra mussel populations within North America and Europe. Can J Fisheries Aquat Sci. 1995; 52: 836–847.
11. Lewis KM, Feder JL, Lamberti GA. Population genetics of the zebra mussel, *Dreissena polymorpha* (Pallas): local allozyme differentiation within Midwestern lakes and streams. Can J Fisheries Aquat Sci. 2000; 57: 637–643.
12. Tautz D, Schlötterer C. Simple sequences. Curr Opin Genet Dev. 1994; 4: 832–837. PMID: 7888752
13. Hosino AA, Bravo JP, Nobile PM, Morelli KA. Microsatellites as tools for genetic diversity analysis. In: Caliskan Mahmut, editor. Genetic diversity in microorganismes. Rijeka, Croatia: InTech; 2012. pp 149–170.
14. DeWoody JA, Avise JC. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. J Fish Biol. 2000; 56: 461–473.
15. Guichoux E, Lagache L, Wagner S, Chaumeil P, Leger P, Lepais O, et al. Current trends in microsatellite genotyping. Mol Ecol Resour. 2011; 11: 591–611. doi: 10.1111/j.1755-0998.2011.03014.x PMID: 21565126
16. Chistiakov DA, Hellemans B, Volckaret Filip AM. Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. Aquaculture. 2006; 255: 1–29.
17. Diez-del-Molino D, Carmona-Catot G, Araguas RM, Vidal O, Sanz N, García-Berthou E, et al. Gene flow and maintenance of genetic diversity in invasive mosquitofish (*Gambusia holbrooki*). PLoS ONE. 2013; 8: e82501. doi: 10.1371/journal.pone.0082501 PMID: 24358194
18. Sanz N, Araguas RM, Vidal O, Diez-del-Molino D, Fernández-Cebrián R, García-Marín JL, et al. Genetic characterization of the invasive mosquitofish (*Gambusia* spp.) introduced to Europe: population structure and colonization routes. Biol Invasions. 2013; 15: 2333–2346.
19. Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA, et al. The population biology of invasive species. Annu Rev EcoSyst. 2001; 32: 305–332.
20. Brown JE, Stepien CA. Population genetic history of the dreissenid mussel invasions: expansion patterns across North America. Biol Invasions. 2010; 12: 3687–3710.
21. Therriault TW, Orlova MI, Docker MF, Macisaac HJ, Heath DD. Invasion genetics of a freshwater mussel (*Dreissena rostriformis bugensis*) in eastern Europe: high gene flow and multiple introductions. Heredity. 2005; 95: 16–23. PMID: 15931246
22. Naish KA, Gosling E. Trinucleotide microsatellite loci for the zebra mussel *Dreissena polymorpha*, an invasive species in Europe and North America. Mol Ecol Notes. 2001; 1: 286–288.
23. Feldheim KA, Brown JE, Murphy DJ, Stepien CA. Microsatellite loci for dreissenid mussels (Mollusca: Bivalvia: Dreissenidae) and relatives: markers for assessing exotic and native populations. Mol Ecol Resour. 2011; 4:725–732. doi: 10.1111/j.1755-0998.2011.03012.x PMID: 21457480

24. Thomas G, Hammouti N, Seitz A. New polymorphic microsatellite loci for the zebra mussel *Dreissena polymorpha* (Pallas, 1771), a common bioindicator. J Shellfish Res. 2011; 30: 123–126.
25. Astanei I, Gosling E, Wilson J, Powell E. Genetic variability and phylogeography of the invasive zebra mussel, *Dreissena polymorpha* (Pallas). Mol Ecol. 2005; 14: 1655–1666. PMID: 15836640
26. Zane L, Bargelloni L, Patarnello T. Strategies for microsatellite isolation: a review. Mol Ecol. 2002; 11: 1–16. PMID: 11903900
27. Zhan A, Bao Z, Hu X, Lu W, Hu J. Methods comparison for microsatellite marker development: different isolation methods, different yield efficiency. J Ocean Univ China. 2008; 2: 161–165.
28. Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ. Rise of the machines—recommendations for ecologists when using next generation sequencing for microsatellite development. Mol Ecol Resour. 2011; 11: 1093–1101. doi: 10.1111/j.1755-0998.2011.03037.x PMID: 21679314
29. Peñarrubia L, Araguas RM, Pla C, Sanz N, Viñas J, Vidal O. Identification of 246 microsatellites in the asiatic clam (*Corbicula fluminea*). Conserv Genet Resour. 2014; PMID: 25067980
30. Vidal O, Sanz N, Araguas RM, Fernández-Cebrian R, Diez-del-Molino D, García-Marín JL. SNP diversity in introduced populations of the invasive *Gambusia holbrooki*. Ecol Freshw Fish. 2012; 21: 100–108.
31. Viñas J, Alvarado-Bremer JR, Pla C. Inter-oceanic genetic differentiation among albacore (*Thunnus alalunga*) populations. Mar Biol. 2004; 145: 225–232.
32. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999; 27: 573–580. PMID: 9862982
33. Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS ONE. 2012; 7: e30087. doi: 10.1371/journal.pone.0030087 PMID: 22347999
34. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 2000; 132: 365–386. PMID: 10547847
35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215: 403–410. PMID: 2231712
36. Schuelke M. An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol. 2000; 18: 233–234. PMID: 10657137
37. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes. 2004; 4: 535–538.
38. Raymond M, Rousset F. GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. J Hered. 2013; 86: 248–249.
39. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. 1980; 32: 314–331. PMID: 6247908
40. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. Biometrics. 1992; 48: 361–372. PMID: 1637966
41. Benjamini BY, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat. 2001; 29: 1165–1188.
42. Goudet J. FSTAT: a computer program to estimate and test gene diversities and fixation indices. J Heredity. 2001; 86: 485–486.
43. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984; 38: 1358–1370.
44. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155: 945–959. PMID: 10835412
45. Excoffier L, Laval G, Schneider S. ARLEQUIN (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform. 2005; 1: 47–50.
46. Richard GF, Kerrest A, Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol Mol Biol Rev. 2008; 72: 686–727. doi: 10.1128/MMBR.00011-08 PMID: 19052325
47. Dolezel J, Bartos J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size of trout and human. Cytometry Part A. 2003; 51: 127–128. PMID: 12541287
48. Gregory TR. Genome size estimates for two important freshwater molluscs, the zebra mussel (*Dreissena polymorpha*) and the schistosomiasis vector snail (*Biomphalaria glabrata*). Genome. 2003; 46: 841–844. PMID: 14608401
49. Saavedra C, Bachère E. Bivalve genomics. Aquaculture. 2006; 256: 1–14.

50. McInerney CE, Allcock AL, Johnson MP, Bailie DA, Prodohl PA. Comparative genomic analysis reveals species-dependent complexities that explain difficulties with microsatellite marker development in mollusks. Heredity (Edinburg). 2011; 106: 78–87. doi: 10.1038/hdy.2010.36 PMID: 20424639

51. Navarro A, Sanchez-Fontenla J, Cordero D, Faria M, Pena JB, Saavedra C, et al. Genetic and phenoptypic differentiation of zebra mussel populations colonizing Spanish river basins. Ecotoxicology. 2013; 22: 915–928. doi: 10.1007/s10646-013-1084-7 PMID: 23681738

52. Ruiz-Altaba C, Jiménez PJ, López MA. El temido mejillón cebra empieza a invadir los ríos españoles desde el curso bajo del río Ebro. Quercus. 2001; 188: 50–51.

3.2

Identification of 246 microsatellites in the Asiatic clam (*Corbicula fluminea*)

Peñarrubia L., Araguas R. M., Pla C., Sanz N., Viñas J. and Vidal O., 2015. **Identification of 246 microsatellites in the Asiatic clam (*Corbicula fluminea*).** *Conservation Genetics Resources* **7**: 393–395.

Peñarrubia L., Araguas R. M., Pla C., Sanz N., Viñas J. And Vidal O. "Identification of 246 microsatellites in the Asiatic clam (Corbicula fluminea)". *Conservation Genetics Resources.* Vol. 7 (2015): 393–395

http://dx.doi.org/10.1007/s12686-014-0378-2

http://link.springer.com/article/10.1007/s12686-014-0378-2

Received: 29 September 2014

Accepted: 28 October 2014

Published online: 6 November 2014

Abstract

Asiatic clam (*Corbicula fluminea*) is one of the most invasive freshwater bivalves. Despite the ecologic and economic impacts of this species, there are only a few polymorphic microsatellites. In this study, we screened the genome from *C. fluminea* in search of microsatellite markers using massive parallel sequencing. We identified 246 new microsatellites involving di-, tri-, tetra- and pentanucleotide single repeats and different compound microsatellites, and we used a validation protocol to characterize a sample of those. Nine of them presented two or more alleles indicating a low level of variability.

Keywords

Asiatic clam

Corbicula fluminea

Microsatellites

Massively parallel sequencing

3.3

SNP identification in two invasive species: zebra mussel (*Dreissena polymorpha*) and Asian clam (*Corbicula fluminea*)

Peñarrubia L., Viñas J., Sanz N., Smith B. L., Pla C. and Vidal O., 2016. **SNP identification in two invasive species: zebra mussel (*Dreissena polymorpha*) and Asian clam (*Corbicula fluminea*).** *Submitted*.

Note: style format of this chapter is presented following the specifications of the JCR journal in which it has been submitted.

SNP identification in two invasive species: zebra mussel (*Dreissena polymorpha*) and Asian clam (*Corbicula fluminea*).

Luis Peñarrubia[a], Jordi Viñas[a], Nuria Sanz[a], Brad L. Smith[b], Carles Pla[a], Oriol Vidal[a*].

a Laboratori d'Ictiologia Genètica. Department of Biology. Universitat de Girona, LEAR Building, Girona, 17071, Spain.

b Department of Wildlife and Fisheries Sciences, Texas A&M University, TAMU 2258, College Station, TX 77840-2258, USA

* Corresponding author at: Laboratori d'Ictiologia Genètica. Department of Biology. Universitat de Girona, LEAR Building room L-115, Girona, 17071, Spain. Tel.: +34 629065158.
E-mail address: oriol.vidal@udg.edu

Abstract

Single-nucleotide polymorphisms (SNPs) have been increased their use in population and conservation genetic studies, due to their advantages over neutral markers. Massive Parallel Sequencing (MPS) platforms have been reducing time and costs in SNP developing. After the SNP calling step, a subsequent validation is usually required. We have used a High Resolution Melting Analysis (HRMA), a simple method for mutation scanning and genotyping by detecting sequence variants through differences in melting temperatures.

We present a set of 5 and 4 novel polymorphic SNPs identified by MPS in two invasive species: zebra mussel (*Dreissena polymorpha*) and Asian clam (*Corbicula fluminea*). Improving the number of molecular markers described in these species is crucial to population genetics in order to understand their invasion history to predict future invasions.

Keywords

Corbicula fluminea, *Dreissena polymorpha*, High Resolution Melting Analysis, Massive Parallel Sequencing, Single Nucleotide Polymorphisms.

Single-nucleotide polymorphisms (SNPs) are becoming more common in population genetic studies (Morin et al. 2009) since the development of Massive Parallel Sequencing (MPS) protocols for their discovery and genotyping. Advantages include low-scoring error rates, high abundance, functionally relevance, and easy high-throughput genotyping (Liu et al. 2005), and in large sets can provide with better F_{ST} estimates than other molecular markers (Morin et al. 2009).

SNP identification in non-model species can be performed even without genomic references, due to the high power of the *de novo* assembly strategies of the MPS approaches (Everett et al. 2011). After the SNP calling step, and before their use in population studies, a subsequent validation is usually required. High Resolution Melting Analysis (HRMA), a relatively new, reliable and cheap technology, is an interesting option for both validation and genotyping. This method was introduced in 2002 (Reed et al. 2007) and it is based in the detection of the of the DNA melting temperature (Tm) which can display differences between several PCR products due to sequence variations (Wittwer et al. 2003; Reed et al. 2007). Even a single base change in the sequence may result in differences in these melting curves (Smith et al. 2010; Smith et al. 2012). As a result, both homozygote and heterozygote genotypes may be identified (Wittwer et al. 2003; Liew et al. 2004; Montgomery et al. 2007).

The main objective of this study was to identify and validate new polymorphic SNP using MPS reads and HRMA of two invasive species, the zebra mussel (*Dreissena polymorpha*, Pallas, 1771) and the Asian clam (*Corbicula fluminea*, Müller, 1774). 454 GS FLX reads of *D. polymorpha* and *C. fluminea* from previous studies (Peñarrubia L et al, 2015a; 2015b) were processed with CLC Genomics Workbench version 4.0 (http://www.clcbio.com/) following Shahin et al. (2012) and Seeb et al. (2011) recommendations. CLC Genomics Workbench software assembled 60,399 reads into 12,840 contigs (mean size = 424 bp; N50 = 469 bp; involving 5,446,443 bases), remaining 50,194 reads unassembled for *D. polymorpha*. In *C. fluminea*, 53,864 were assembled into 10,853 contigs (mean size = 402 bp; N50 = 458 bp; involving 4,363,423 bases) and 44,681 reads remained unassembled.

In *D. polymorpha*, 783 sequence variants were detected, of which 721 were Single Nucleotide Variants (SNVs) (356 transversions and 365 transitions), 20 were Multiple Nucleotide Variants (MNVs, two or more consecutive nucleotide changes), and 42 were Insertions-Deletions (InDels). In contrast, 446 sequence variants were detected in *C. fluminea*, of which 417 were SNV (188 transversions and 229 transitions), 11 were MNVs and 18 were InDels. SNP

annotation into contigs was performed using Geneious software, version 5.6 (Kearse et al. 2012). For all SNP, variant type, allele variants, sequence coverage, allele counts and percentage, and average quality were registered.

For HRMA, we used 15 individuals of each species following workflow designed by Smith et al. (2012). PCR primers were designed using Primer3 (Rozen and Skaletshy 2000) implemented in Geneious interface. Melting temperatures were estimated with uMELT^{SM} v2.0 (Dwight et al. 2011). BLASTN analysis (Altschul et al. 1990) was run to identify possible homologies producing non-specific amplifications in the PCR. A total of 86 SNPs were selected for validation to best predicted melting curves according Smith et al. (2012) recommendations. Thus, for 46 SNPs *in D. polymorpha* and 40 SNP in *C. fluminea* primers were designed, yielding amplicon sizes shorter than 65 bp (Table S1).

HRMA amplifications were conducted in 10 µL reactions containing 25–100 ng of genomic DNA, 1× Econotaq Plus Master Mix (Lucigen), 1× LCGreen+ fluorocrome (Idaho Technology), and 0.20 µM of each primer. Thermal cycling was performed on a LightCycler 480 Real-Time PCR system (Roche Diagnostics) with an initial denaturation of 10 min at 95°C followed by 35 cycles denaturing for 10 s at 95°C, annealing at 60°C for 30 s, and extension for 10 s at 72°C. Reactions were overlaid with 15 µL of mineral oil to prevent evaporation (Smith et al., 2010). Twenty-five HRMA data acquisitions were collected in a ramp rate of 0.02°C/s between 60 and 95°C. All melting curve patterns were analyzed using the LightCycler 480 Gene Scanning Software v. 1.5.0 SP1 (Roche Diagnostics). When a double-peak pattern resulted for all individuals, SNPs were genotyped in a set of 96 individuals to be corroborated (Figure S1). In *D. polymorpha*, 5 SNPs displayed the expected melting curves and were polymorphic among the analyzed individuals (Table 1). In *C. fluminea*, 4 SNP showed melting curves compatible with polymorphic SNPs (Table 1). Polymorphic SNPs were submitted to GenBank with accession numbers KT220181-86 and KT220188-90.

A constant heteroduplex melting pattern among the initial 15 individuals was found in 23 SNPs of *D. polymorpha* and 18 SNPs of *C. fluminea* (Table S1), and further genotyping in 96 individuals was found in all cases (Figure S1). This feature could be compatible with these regions being paralogous sequence variants (PSV) (Smith et al. 2005). This number of non-validated SNPs may be related to the fact that our MPS data sets were derived from a single individual of each species, belonging to invading populations with low levels of genetic

diversity (Blackburn et al. 2015). Alternatively, it could be possible that that PSV are a common trait of the genomes of mollusks, a taxa with cryptic repetitive DNA (McInerney et al. 2011).

Table 1. Name, nucleotide variant (*NV*), amplicon size (*As*), forward (*F*) and reverse (*R*) primers, Annealing temperature (*Ta*) for polymorphic SNPs in *D. polymorpha* and *C. fluminea*.

Species	Name	*NV*	*As*	Primer sequences (5' → 3')	*Ta* (°C)	GenBank Accesion
D. polymorpha	Dp292	A/T	60	*F:* TGCAACCGAGTTTACCAACGGCT *R:* TGCTGTTCAAATGAACCGGAGCAG	57	KT220181
	Dp367	T/G	60	*F:* TCGCCTTGCAAGTCTCGTGCT *R:* GCAATTGTTCTTGCAGTAATGTCCCGC	57	KT220182
	Dp452	T/G	60	*F:* GCTGCCTGAAACGTTCAGTGGT *R:* CCTCCGGGATCGGCCCACTT	57	KT220183
	Dp467	A/G	54	*F:* TGCGTGGAGCCTTTCCACCG *R:* TGGCAAGAACAAAGCAGACCGC	57	KT220184
	Dp501	C/T	55	*F:* GTGTGAAATCTTGAAAGCGCCTTGT *R:* GGCTGCTGGTAAATAAATGGGCTCCG	57	KT220185
C. fluminea	Cf46	C/G	54	*F:* CGAAAGCTGCGCATTTCTGCGA *R:* ACCTGCGGATGGATCATTACCGA	57	KT220186
	Cf132	T/G	59	*F:* TGTAGGCGGCCACCCCATGT *R:* GGTCTTCACTGACGGGCGGC	60	KT220188
	Cf190	T/A	60	*F:* AGCTTACAGTTTGCCCACTTACCTCT *R:* AGATGCGAATTGGCCCCGGT	57	KT220189
	Cf270	T/A	60	*F:* GTAATGTCCGTCTGCGTATCAGATTCA *R:* TGCCGGGGTGTCTTGTTTGTCG	57	KT220190

Acknowledgements

This research was carried out within the objectives of the research project CGL200909407 of the Ministry of Sciences and innovation (MICINN) of the Spanish Government. LP received economical support with a PhD fellowship of the Spanish MICINN with reference BES–2010037446. LP also received economical support of the Spanish MICINN for a scholar visitor fellowship with reference EEBB-I-12-05670.

Reference list

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410.

Blackburn TM, Lockwood JL, Cassey P (2015) The influence of numbers on invasion success. Mol Ecol 24:1942–1953.

Dwight Z, Palais R, Wittwer CT (2011) uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. Bioinformatics 27:1019–1020.

Everett MV, Grau ED, Seeb JE (2011) Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. Mol Ecol Resour 11:93–108.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Liew M, Pryor R, Palais R, Meadows C, Erali M, Lyon E, Wittwer C (2004) Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. Clin Chem 50:1156-1164.

Liu N, Chen L, Wang S, Oh C, Zhao H (2005) Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. BMC Genet 6:S26.

McInerney CE, Allcock AL, Johnson MP, Bailie DA, Prodöhl PA (2011) Comparative genomic analysis reveals species dependent complexities that explain difficulties with microsatellite marker development in molluscs. Heredity 106:78–87.

Montgomery J, Wittwer CT, Palais R, Zhou L (2007) Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. Nat Protoc 2:59-66.

Morin PA, Martien KK, Taylo BL (2009) Assessing statistical power of SNPs for population structure and conservation studies. Mol Ecol Resour 9:66-73.

Peñarrubia L, Araguas RM, Pla C, Sanz N, Viñas J, Vidal O. (2015). Identification of 246 microsatellites in the Asiatic clam (*Corbicula fluminea*). Conserv Genet Resour *7:*393-395.

Peñarrubia L, Sanz N, Pla C, Vidal O, Viñas J. (2015). Using Massive Parallel Sequencing for population genetics markers development, validation and application in the invasive bivalve zebra mussel (*Dreissena polymorpha*). PLoS ONE 10:e0120732.

Reed HG, Kent JO, Wittwer CT (2007) High-Resolution DNA melting analysis for simple and efficient molecular diagnostics. Pharmacogenomics 8:597-608.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386.

Seeb JE, Pascal CE, Grau DE, Seeb W, Templin WD, Harkins T, Roberts SB (2011) Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. Mol Ecol Resour 11:335-348.

Shahin A, van Gurp T, Peters SA, Visser RG, van Tuyl JM, Arens P (2012) SNP markers retrieval for a non-model species: a practical approach. BMC Res Notes 29:5-79.

Smith CT, Elfstrom CM, Seeb LW, Seeb JE (2005) Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Mol Ecol 14:4193–4203.

Smith BL, Lu CP, Alvarado-Bremer JR (2010). High-resolution melting analysis (HRMA): a highly sensitive inexpensive genotyping alternative for population studies. Mol Ecol Resour 10:193-196.

Smith BL, Lu CP, Alvarado Bremer JR (2012) Methodological streamlining of SNP discovery and genotyping via high-resolution melting analysis (HRMA) in non-model species. Mar Genomics 9:39-49.

Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. Clin Chem 49:853-860.

3.4

Genetic characterization of the invasive zebra mussel (*Dreissena polymorpha*) in the Iberian Peninsula

Peñarrubia L., Pla C., Viñas J., Vidal O. and Sanz N., 2016. **Genetic characterization of the invasive zebra mussel (*Dreissena polymorpha*) in the Iberian Peninsula.** *Hydrobiologia* **779**: 227–242.

Peñarrubia L., Pla C., Viñas J., Vidal O. and Sanz N. “Genetic characterization of the invasive zebra mussel (Dreissena polymorpha) in the Iberian Peninsula”. *Hydrobiologi*a. Vol. 779 (2016) : 227–242.

http://dx.doi.org/10.1007/s10750-016-2819-2

http://link.springer.com/article/10.1007/s10750-016-2819-2

Received: 16 February 2016

Revised: 5 May 2016

Accepted: 6 May 2016

Published online: 28 May 2016

Abstract

Asiatic clam (*Corbicula fluminea*) is one of the most invasive freshwater bivalves. Despite the ecologic and economic impacts of this species, there are only a few polymorphic microsatellites. In this study, we screened the genome from *C. fluminea* in search of microsatellite markers using massive parallel sequencing. We identified 246 new microsatellites involving di-, tri-, tetra- and pentanucleotide single repeats and different compound microsatellites, and we used a validation protocol to characterize a sample of those. Nine of them presented two or more alleles indicating a low level of variability.

Keywords

Asiatic clam

Corbicula fluminea

Microsatellites

Massively parallel sequencing

3.5

Genetic characterization of the Asian clam species complex (*Corbicula*) invasion in the Iberian Peninsula

Peñarrubia L., Pla C., Viñas J., Vidal O., Araguas R. M. and Sanz N., 2016. **Genetic characterization of the Asian clam species complex (*Corbicula*) invasion in the Iberian Peninsula.** *Hydrobiologia* DOI: 10.1007/s10750-016-2888-2.

Peñarrubia L., Pla C., Viñas J., Vidal O., Araguas R. M. and Sanz N. “Genetic characterization of the Asian clam species complex (Corbicula) invasion in the Iberian Peninsula. *Hydrobiologia*. Vol. 784, Issue 1 (2017 jan) : 349–365

http://dx.doi.org/10.1007/s10750-016-2888-2

http://link.springer.com/article/10.1007/s10750-016-2888-2

Received: 31 March 2016

Revised: 15 June 2016

Accepted: 17 June 2016

Published online: 1 July 2016

Abstract

The Asian clam (*Corbicula* sp.) is an invasive freshwater bivalve native to Asia, the Middle East, Australia, and Africa. It is now widely distributed around the world producing large ecological and economic impacts. Three well-described invasive lineages form a cryptic species complex with asexual reproduction based on androgenesis. In this study, we collected 175 individuals from different Iberian, European, and North American locations to genetically study *Corbicula* invasion in the Iberian Peninsula using COI and 28S genes. The use of mitochondrial and nuclear markers allows us to characterize both maternal and paternal inheritance from androgenetic *Corbicula* locations and to deal with the incongruences caused by egg parasitism. We identified 7 COI and 10 28S haplotypes that grouped individuals within the three invasive *Corbicula* lineages. Haplotype distribution of mitochondrial and nuclear markers detected genetic divergence between the Ebro Delta location and the rest of Iberian sites, suggesting that at least two invasion episodes occurred in the Iberian Peninsula. Haplotype distribution also suggested secondary contacts between Iberian and other European invaded regions. Additionally, results revealed that nuclear hybridization, a feature more widespread than previously reported, contributes to retain gene diversity in the *Corbicula* invasion.

Keywords

Corbicula; Iberian Peninsula; Invasive species; Mitochondrial COI; Nuclear 28S; Population structure

3.6

Validated methodology for quantifying infestation levels of dreissenid mussels in environmental DNA (eDNA) samples

Peñarrubia L., Alcaraz C., Bij de Vaate A., Sanz N., Pla C., Viñas J. and Vidal O., 2016. **Validated methodology for quantifying infestation levels of dreissenid mussels in environmental DNA (eDNA) samples.** *Submitted*.

Note: style format of this chapter is presented following the specifications of the JCR journal in which it has been submitted.

Validated methodology for quantifying infestation levels of dreissenid mussels in environmental DNA (eDNA) samples

Luis Peñarrubia[a], Carles Alcaraz[b], Abraham bij de Vaate[c], Nuria Sanz[a], Carles Pla[a], Oriol Vidal[a*], Jordi Viñas[a*].

a Laboratori d'Ictiologia Genètica. Department of Biology. Universitat de Girona, LEAR Building, Girona, 17071, Spain.
b IRTA Aquatic Ecosystems, Sant Carles de la Ràpita, 43540, Spain.
c Waterfauna Hydrobiologisch Adviesbureau, Oostrandpark 30, 8212 AP Lelystad, The Netherlands.

* Co-corresponding authors:

Jordi Viñas
Laboratori d'Ictiologia Genètica. Department of Biology. Universitat de Girona, LEAR Building room L-117, Girona, 17071, Spain. Tel.: +34 629409072.
E-mail address: jordi.vinas@udg.edu

Oriol Vidal
Laboratori d'Ictiologia Genètica. Department of Biology. Universitat de Girona, LEAR Building room L-115, Girona, 17071, Spain. Tel.: +34 629065158.
E-mail address: oriol.vidal@udg.edu

Abstract

The zebra mussel (*Dreissena polymorpha* Pallas, 1771) and the quagga mussel (*D. rostriformis* Deshayes, 1838) are successful invasive bivalves with substantial ecological and economic impacts in fresh water systems once they become established. Since their eradication is extremely difficult, their detection at an early stage is crucial to prevent spread. In this study, we optimized and validated a qPCR detection method based on the histone H2B gene to quantify combined infestation levels of zebra and quagga mussels in environmental DNA samples. Our results show specific dreissenid DNA present in filtered water samples for which microscopic diagnostic identification for larvae failed. Monitoring a large number of locations for invasive dreissenid species based on a highly specific environmental DNA qPCR assay may prove to be an essential tool for management and control plans focused on prevention of establishment of dreissenid mussels in new locations.

Keywords

Dreissenids, environmental DNA, Histone gene, Infestation level quantification, qPCR amplification, Quagga mussel, Zebra mussel.

Introduction

The zebra mussel, *Dreissena polymorpha* (Pallas, 1771) is a successful invasive bivalve native to the brackish estuaries and freshwaters systems of the Ponto-Caspian regions (Black, Caspian, and Azov Seas). It is considered one of the 100 world's worst invasive alien species (IUCN-ISSG)[1] possessing several biological life cycle features that favor its proliferation: rapid growth, early sexual maturity, and high fecundity of up to 1 million eggs per season[2,3]. Due to a planktonic larval stage and also facilitated by human-mediated activities such as larvae in ballast water[4], attachment of adults to hulls of commercial and recreational ships[5,6,7].

The zebra mussel was first time found on the Iberian Peninsula in the Ribarroja Reservoir (Ebro River) in 2001[8], about 100 years later than its first report in Western Europe[2]. This delayed invasion might have been due the Pyrenees mountain range serving as a barrier[9,10] to dispersal. Since then, this species has expanded along the Ebro River and adjacent basins in the Northeast region of the Iberian Peninsula[11,12]. Moreover, these effects are expected to increase with the imminent arrival of another invasive dreissenid, the quagga mussel (*D. rostriformis* Deshayes, 1838)[10,13,14]. Around 2004 the quagga mussel arrived in Western Europe[15,16,17] and has been constantly spreading into areas where zebra mussels previously invaded[18,19,20], and thus are predicted to arrive to the Iberian Peninsula in a near future unless effective prevention measures are enacted[10].

To face these challenges, the National Strategy for the Control of Zebra Mussel in Spain has focused efforts on the rapid detection of both species[21]. Plans for control and management[22,23] were developed to prevent further zebra mussel expansion since eradication of dreissenids is extremely difficult once they are established[4,24]. Therefore it is critical to detect their presence as early as possible, when measures of eradication could be effective[25,26].

The current methodology for detecting the presence of the zebra mussel is rather rudimentary. Adult individual detection is based on visual diagnostics, and veliger larvae are detected using microscopy[25]. The latter is labour intensive and time-consuming, and can result in false positives due to the similar appearance between dreissenid larvae and those of other macroinvertebrates[27]. Alternatively, molecular markers have proven useful in species identification diagnostics, but these previous studies do not intend sample quantification[4,28,29]. Thus, after optimization they offer a faster diagnosis without false positives but they do not allow the quantification of the dreissenid infestation. The combination of environmental DNA (eDNA) sampling is with a quantitative PCR (qPCR) method[30,31] may be optimized to permit quantification, as evaluated here.

In the present study, we developed and optimized a qPCR procedure to assess the level of dreissenid infestation using eDNA samples. This method is based on a single-copy gene used as a genetic marker from Massive Parallel Sequencing (MPS) output of zebra mussel genomic DNA[11]. This method could be applied to diagnose the dreissenid infestation level in locations at risk of invasion. Evaluation and quantification of these locations will be crucial to complete the management and control plans to prevent future dreissenid spread.

Tab**le 1.** Markers developed for dreissenid identification derived from MPS output and qPCR conduction. MPS contig reference, Blast Result, predicted gene, Forward (F) and Reverse (R) primer sequences, amplicon size in base pairs, and optimum annealing temperature (Ta) for PCR amplification, Melting temperature (Tm) in Real Time PCR, efficiency (E), coefficient of variation (CV), coefficient of regression (R^2), detection (DL) and quantification (QL) levels and GenBank accession number for four single-copy predicted genes selected for zebra mussel.*Significant BlastX E-value < 1E-50.

Single-copy gene calling step			PCR amplification step			qPCR amplification step						GenBank accession
MPS Contig reference	BlastX E- value	Predicted gene	Primer sequences (5' → 3')	Size (bp)	*Ta* (°C)	*Tm* (°C)	*E*	*CV*	R^2	*DL* (ng/µl)	*QL* (ng/µl)	
Contig000070	3.77E-88*	H3	*F:* GGTGACACGCTTGGCGTGGA *R:* GCCAGGAACCGTCGCCCTTC	229	60	-	-	-	-	-	-	JWHF01000070
Contig000076	5.31E-54*	H2B	*F:* CGCGCGCTCCACTGACAAGA *R:* CACCAGGCAGCAGGAGACGC	251	60	85.4692 ± 0.2057	1.9729	0.1541	0.9976	5E-4	5E-4	JWHF01000076
Contig000102	1.08E-27	H1	*F:* TCTTGGCGCCCGCCTTCTTG *R:* GTCAGTGCCGTCAACGCCCA	214	60	86.0933 ± 0.2612	1.9700	0.2283	0.9954	5E-3	5E-3	JWHF01000102
Contig000913	7.52E-23	MARS	*F:* AGTCCTCCCAGATTAGCCTGTGC *R:* AGATGTCGCGGTGGAGGGCT	277	65	80.8778 ± 2.6423	1.9102	0.1772	0.9940	5E-2	5E-1	JWHF01000913

Tab**le 2.** Location, coordinates, sampling date (S = spring, A = autumn), zebra mussel presence previously diagnosed by ACA, and molecular diagnostic results (+= positive; ND= non detected) for all samples analyzed. SEM: standard error of the mean.

Sample acquisitions			Visual diagnostic by ACA		Molecular diagnostic		
Location, reservoir	Latitude / Longitude	Sampling Date	larvae	adults	PCR	qPCR (ng/L ± SEM)	Target copies / µL (± SEM)
La Baells	42°08'N / 01°54'E	S: 08/05/2014	+	+	+	14.639 ± 0.773	5.32E+4 ± 2.81E+3
		A: 01/10/2014	+	+	+	5.901 ± 0.388	2.15E+4 ± 1.41E+3
Gaià	41°11'N / 01°19'E	S: 04/06/2014	ND	ND	ND	ND	ND
		A: 01/10/2014	ND	ND	+	204.771 ± 13.941	7.44E+5 ± 5.07E+4
La Llosa	42°05'N / 01°34'E	S: 21/05/2014	ND	ND	ND	ND	ND
		A: 29/09/2014	ND	ND	+	24.305 ± 1.081	8.84E+4 ± 3.93E+3
Sant Ponç	41°57'N / 01°36'E	S: 21/05/2014	ND	ND	+	5.267 ± 0.157	1.91E+4 ± 5.71E+2
		A: 29/09/2014	ND	ND	+	19.131 ± 2.304	6.95E+4 ± 8.37E+3
Boadella	42°20'N / 02°50'E	S: 14/05/2014	ND	ND	ND	4.456 ± 0.319	1.62E+4 ± 1.16E+3
		A: 17/09/2014	ND	ND	+	73.920 ± 5.502	2.69E+5 ± 2.00E+4
Banyoles Lake	42°07'N / 02°45'E	26/11/2013	ND	ND	ND	ND	ND

Results

Stage A: single-copy gene marker selection

Gene annotation, single-copy gene selection and PCR specificity

Analysis of MPS output[11] using Blast2GO software resulted in a total of 220 Gene Ontology (GO) terms that were tagged on 56 contigs (1.44% of the total 2,326)(see Supplementary Fig. S1 online). When single-copy genes were screened in contigs larger than 500 bp, just four were found in the BlastX analysis (Table 1). Of these genes, one corresponded to the methionyl-tRNA synthetase (MARS) gene with an E-value of 7.52E-23. The remaining three genes were different histone genes as follows: histone 1 (H1) with an E-value of 1.08E-27; histone 2B (H2B) with 5.31E-54; and histone 3 (H3) with having 3.77E-88 (Table 1). All four markers were verified for species specificity, and three (H1, H2B and MARS) presented a single and clear PCR amplicon for the two dreissenid species (see Supplementary Fig. S2 online).

Standard qPCR curve amplification

Following the species specificity test, we considered that three markers (H1, H2B and MARS) were suitable to be optimized for dreissenid DNA quantification by qPCR. After several attempts, only H1 and H2B markers had a dynamic range with a coefficient of variation (CV) below 0.3 in all concentrations tested (from 50 to 5E-3 ng/µL), characterized by a single peak in the melting curve (Table 1). H2B was selected due to real time amplification efficiency (E = 1.973), lower Blast E-value, and lower coefficient of variation among replicates (CV = 0.154). Adding the concentrations needed for the standard curve, the H2B gene showed both detection level (DL) and quantification level (QL) at 5E-4 ng/µL. In all cases, the sequence of the PCR products matched the reference sequence with a variation of 5 SNPs between the two dreissenid species (see Supplementary Fig. S2 online).

Stage B: Environmental quantification

PCR amplification

The Banyoles Lake eDNA sample confirmed the absence of dreissenid DNA resulting in negative PCR amplification. Positive dreissenid DNA presence was observed in the remaining five locations in at least one of the two temporal samples (Table 2, Fig. 1). In the spring sampling, three locations (La Baells, Sant Ponç and Boadella) were positive for dreissenid DNA, and all five locations were positive in the autumn sampling (Table 2, Fig. 1).

Infestation DNA level quantification

The DNA quantification comparisons between seasons were significantly different for all locations (Fig. 1). As a general pattern, the DNA levels were approximately seven times higher (t = -4.791; df = 34; P = 0.000; inset in Fig. 1) in autumn (70.081 ng/L; SEM = 14.258) than in spring (10.728 ng/L; SEM = 1.361). The Sant Ponç (t = -4.103; df = 7; P = 0.005) and Boadella (t

= -6.514; *df* = 5; *P* = 0.001) reservoirs followed this general pattern with significantly higher DNA values after the summer spawning period (Fig. 1). In contrast, La Baells Reservoir presented significantly lower DNA quantity for the autumn season (*t* = 10.105; *df* = 11; *P* 0.000). Moreover, the spring season comparison (*F* = 70.802; *df* = 14; *P* = 0.000) showed that La Baells Reservoir differed from the other two reservoirs (Sant Ponç and Boadella Reservoirs) with the presence of dreissenid DNA in this season. In contrast, the autumn season was more variable (*F* = 235.539; *df* = 30*; P* = 0.000) with four significantly differentiated groups. Only La LLosa and Sant Ponç Reservoirs presented similar quantification levels (Fig. 1).

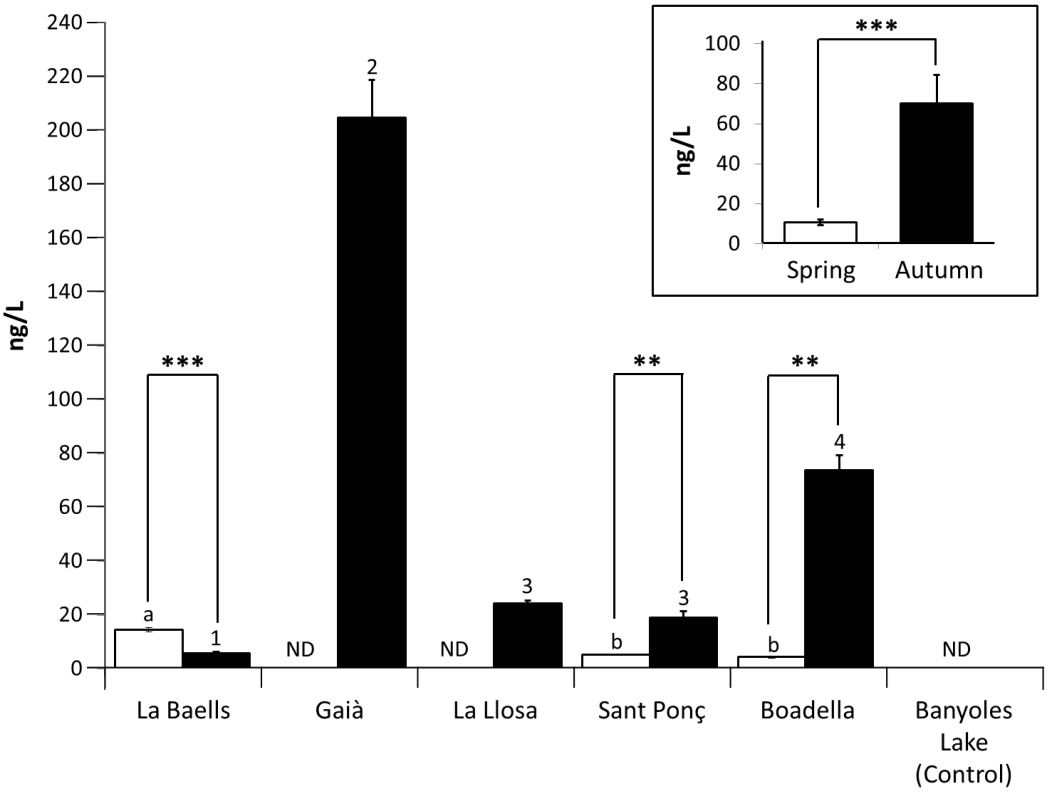


Figure 1. *DNA quantification level for each location. White and black bars correspond to spring and autumn sampling periods respectively. Right square represents the total quantification average among all locations for both sampling periods. Pairwise T-Student comparisons between sampling periods significance is represented by asterisks (*: P < 0.05; **: P < 0.001; ***: P < 0.0001). Letters and numbers correspond to statistical identity among spring and autumn comparisons respectively after ANOVA and subsequent Bonferroni post-hoc analysis. ND= non-detected.*

Discussion

Several previous studies have developed detection methods based on molecular markers for dreissenid identification in environmental samples[4,28,29]. However, to our knowledge, this study is the first published in which a method was developed for identification and quantification of

the infestation level of invasive dreissenid species. To achieve these results, we generated and validated a highly specific marker using qPCR quantification of the H2B nuclear single-copy gene. The selection of the maker (H2B single-copy gene) was realized by bioinformatics analyses of a previous MPS published output[11] with a laboratory validation of the usefulness of the marker. The choice of the marker was based on the high specificity for dreissenid species, which allows the discrimination of dreissenids species from endemic mollusk species. The marker also presented the best qPCR parameters compared to other putative markers (Table 1). Several other studies have targeted mitochondrial DNA based markers for species identification in environmental samples[32,33,34] instead of using single copy nuclear genes. The use of mitochondrial DNA is extremely useful for detecting the presence of a species, but has some disadvantages when besides detecting presence the protocol is designed for quantification[35,36].

Our qPCR methodology developed is not designed to provide the absolute number of larvae or adults in the water body, but it quantifies the amount of dreissenid DNA, which very likely informs about the levels of infestation. As we will discuss, the fact that we were able to detect the significant differences in seasonal fluctuations within dreissenid populations strongly supports this assessment. In this context, any source of further variation should be avoided. Thus, although the total amount of DNA (both nuclear and mitochondrial) may vary by several orders of magnitude during ontogeny[27,28,30], the number of mtDNA molecules also varies from cell to cell. The use of mtDNA markers is therefore not recommended as they would incorporate additional errors in quantifying the infestation[36].

This protocol outperforms an alternative quantification using the difference in the number of endpoint PCR replicates with positive detections, even in samples with low DNA presence[37]. Replicative endpoint PCR would experience the same problem of quantifying DNA instead of assessing the number of individuals. However, a single qPCR can yield a quantification of the amount of DNA[38,39]. Furthermore, replicating qPCRs, as it is done in replicative endpoint PCR methods should provide with a more precise quantification of the presence of dreissenid DNA. Thus, using the qPCR protocol described here, we can provide water managers with accurate data on the success of control measures in the early invasion stages. This information is crucial for eradications, since it has been demonstrated that effective control measures should be undertaken as early as possible[4,25,26].

The H2B single-copy gene marker does not discriminate between the closely related zebra and quagga mussels due to the low levels of interspecies differentiation of this marker between the two species (see Supplementary Fig. S2 and Fig. S3 online). The inability to separate zebra from quagga mussels is not a hindrance to implement the method developed in this study. Both dreissenid species have a related native range, similar life histories and morphology[3,40], and they are currently well expanded in Europe[10,15,19]The zebra mussel expansion typically occurs earlier [13,14,41] with an invasion lag time five times shorter than the quagga mussel[14,42,43]. Quagga mussels become established in water bodies where zebra mussels are present[7,10], and zebra mussel populations are gradually replaced by quagga mussels[13,43,44] with a complete replacement after nine of more years of coexistence[14]. Thus, the lack of discrimination between these two species and the possible incipient quagga mussel specimens in the area

studied is not a critical problem to implement the developed method, since regardless which of the two species are present the impact of the invasion is similar[3,13,14] or when species occurs concurrently, their effect is even greater[14]. Interestingly, as the first reported presence of the zebra mussel in the Iberian Peninsula was in 2001, we could suspect that the quagga mussel is already present in the Iberian Peninsula. However, the morphological analysis of more than 4,000 adult individuals from two Iberian locations (in the first and the last locations where zebra mussel was detected; see Supplementary Table S1) failed to identify quagga mussel individuals, which may indicate that the invasion of quagga mussel has not yet started or it is in its initial stages.

Additional advantages of our qPCR method developed in this study include: (1) faster results, (2) more cost-effective, (3) and more powerful resolution to detect presence of dreissenid infestation than the previous methods based on visual inspection and/or molecular markers. The traditional microscopic screening methods for veliger larvae detection[25,26] are labour intensive and time-consuming[40], with a relatively high potential for false positive results. In addition, our method outperforms previous molecular methods aimed at determining only dreissenid DNA presence[4,28,29]. To the best of our knowledge, this is the first method capable to differentiate specific dreissenid DNA levels from other bivalves in eDNA samples.

A goal of this study was to apply our method to actual environmental samples. Five locations were sampled within the recollection program of the Catalan Governmental control and management of zebra mussel expansion[22] (Table 2). Each location was analysed in two sampling periods as follows: prior to the spawning in spring period ; and after the spawning in autumn period[25]. When comparing the visual and molecular analyses results, the molecular analysis provided more positive results than the visual inspection. For instance, neither larvae nor adults were detected by traditional methods in the autumn samples from La Llosa and Boadella Reservoirs as well as in both temporal samples from the Sant Ponç Reservoir. In contrast, the PCR amplifications (subsequently confirmed by Sanger sequencing) were positive in all these samples. These discrepancies between analysis methods may be due to a low number of larvae present in the water column, which is not detected by the microscopical inspection. We should be cautious in the locations where we have failed to detect dreissenid DNA, as there is the possibility that the overall eDNA levels are poorly represented by the sampling design since eDNA concentration may vary between adjacent samples[45,46], with the risk of having dreissenid DNA (and thus infestation) in these negatively amplified locations. However, the sampling of our study is based on capturing free but also larvae DNA and thus having a more representative sampling consequence of the larvae movement and propagation[47]. On the contrary, the positive PCR results may indicate an incipient invasion of dreissenid, and these locations should be considered major objectives for the prevention directives by the governmental agencies.

A novel component of our qPCR method is the quantification of the dreissenid infestation level. As a general pattern, we obtained a significantly higher presence of dreissenid DNA in autumn than in spring (t = -4.791; df = 34; P = 0.000) (inset in Fig. 1). This is in concordance with the vital cycle of zebra mussels, in which major spawning is in summer[14,25] when the water temperature increases above 18°C[25]. Thus, significantly higher DNA amounts are

expected in autumn as a consequence of the higher concentrations of veligers in water samples. This pattern was shared in four of the five locations analysed (Fig. 1). In the Sant Ponç and Boadella Reservoirs, DNA presence was detected for both sampling seasons with significantly higher DNA quantity in the autumn sample, whereas dreissenid DNA presence was only detected in the autumn season in the Gaià and La LLosa Reservoirs (Fig. 1). La Baells Reservoir presented an opposite pattern with significantly higher DNA presence in the spring season (Fig. 1). The results from this reservoir were in concordance with the Catalan Water Agency (ACA) monitoring plan as it was the only reservoir with a confirmed presence of adult individuals along the year[22].

One surprising result was the high DNA quantity found in the autumn season for the Gaià Reservoir with no presence of DNA in the spring. A similar situation occurred in La Llosa samples albeit to a lesser extent (Table 2, Fig. 1). ACA only found the presence of larvae but no adults in Gaià in 2012. However, the presence of zebra mussels has not been observed in succeeding years. Based on these results, we suspect that these locations are probably having a flow of dreissenid input, with a high risk of establishment of the invasion in a short term.

In summary, to the best of our knowledge, we have developed the first method for specific detection and quantification of dreissenid DNA in environmental samples based on qPCR. This method outperforms previous methods based on visual and microscopic inspection, and it provides additional information than other molecular methods only based on the detection of the presence of dreissenid DNA. Application of our method allows early detection of dreissenid invasions and fast implementation of control measures.

Methods

Sequences from zebra mussel were obtained from a previous Massive Parallel Sequencing (MPS) study[11]. De novo assembly generated a total of 2,326 contigs (contig range size = 100 – 8,697 bp; mean contig size = 457 bp; N50 = 825 bp) and was submitted to the DDBJ/EMBL/GenBank Whole Genome Shotgun project when the contig size was longer than 500 nucleotides under accession number JWHF00000000.

Stage A: Single-copy marker selection

Homologous sequences of the contig sequences longer than 500 nucleotides were identified using Blast2GO software (version 2.4.2)[48]. The Non-Redundant (NR) NCBI protein database was searched using BlastX with a Cut-Off e-value of 1E-6, Cut-Off length of 30, and 20 Blast Hits. Subsequently, Gene Ontology (GO) terms were assigned according to the Gene Ontology Database with an e-value-Hit-Filter of 1E-6, annotation Cut-Off of 55, and GO Weight of 5. These results were used to predict single-copy genes using homologies with NCBI and Ensembl[49] databases. The PCR primers for selected genes were designed using Primer3[50] with default parameters.

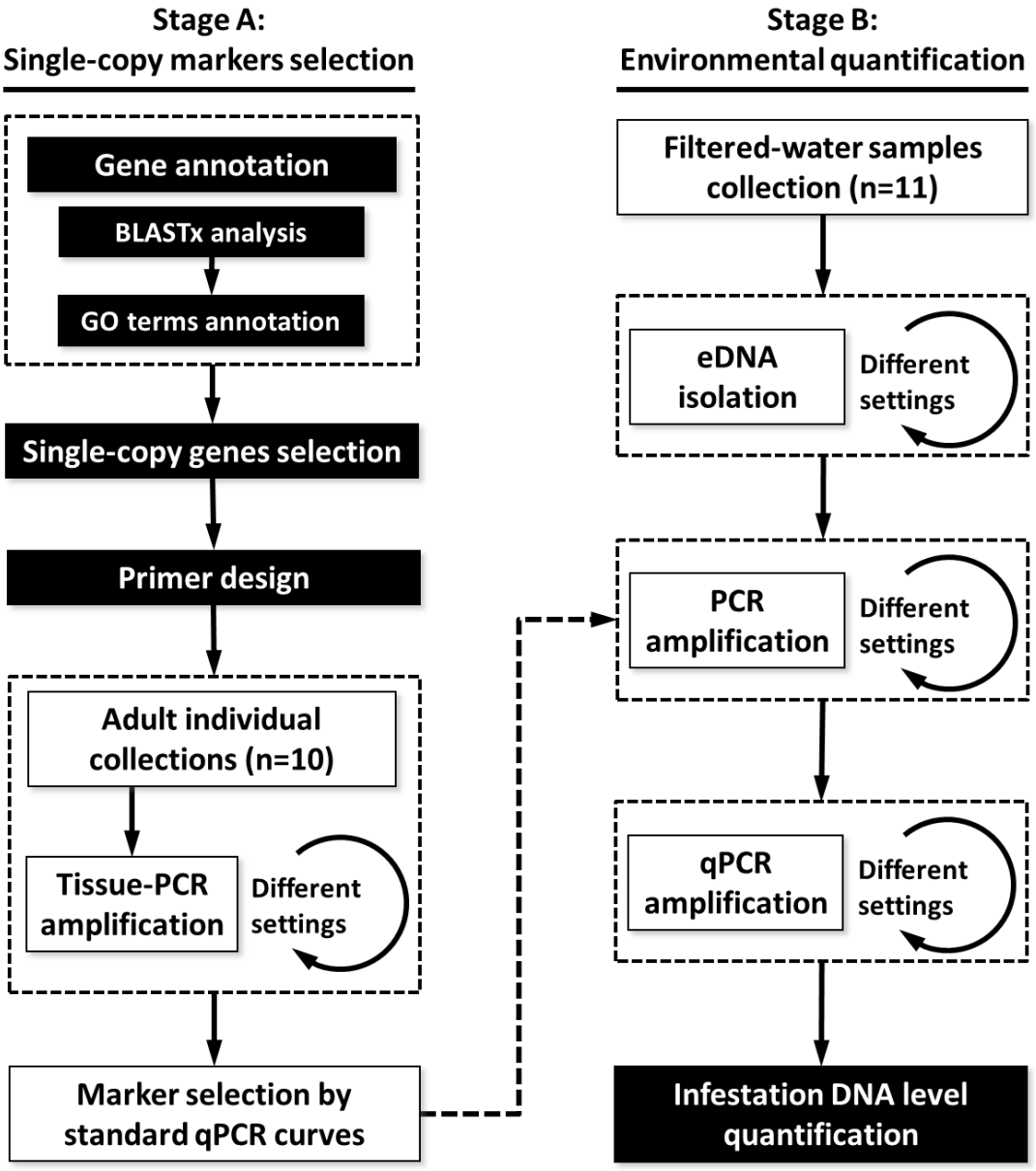


Figure 2. *Workflow of experimental design. Black steps: bioinformatic analysis, white steps: laboratory procedures.*

Adult individual collections for PCR development and validation

The PCR results were validated in three adult zebra mussels from the Aragón Imperial Canal (Ebro River), and the PCR specificity was verified against several mollusc species found in overlapping distribution with zebra mussel: one individual of Spengler's freshwater mussel (*Margaritifera auricularia*), two Asian clam (*Corbicula fluminea*) and one spike-topped apple snail (*Pomacea sp.*). Furthermore, two individuals of quagga mussel sampled in the Netherlands (52°42'N, 05°18'E) were also included. Whole bodies without shell were preserved in 70% ethanol until processed. DNA isolations were performed using the EZNA

Mollusk DNA kit (Omega Bio-Tek), and DNA was eluted in a volume of 200 µl. DNA quality and quantity was verified by agarose gel electrophoresis, Qubit v2.0 fluorometer (Life technologies) and NANODROP spectrophotometer (Thermo Fischer Scientific).

In addition, to infer the possible presence of quagga mussel in waters of the Iberian Peninsula, we collected more than 4,000 dreissenid adult individuals in 2013. All individuals were collected in Ribarroja reservoir (Ebro River; first cited record of the zebra mussel in the Iberian Peninsula[8]; n = 3,013) and La Baells (Llobregat River; last cited record in the Iberian Peninsula[23]; n = 1,230) (see Supplementary Table S1).

Tissue PCR amplification

All primer sets were tested by end-point PCR in adult tissue DNA using a 2720 Thermal Cycler (Applied Biosystems). PCR assays were set up in 30 µL reactions containing 25–100 ng of genomic DNA, 1X Buffer, 1.5 µM $MgCl_2$, 0.8 mM dNTPs, 0.2 µM of each primer, 2.5E-2 u/µL Taq polymerase (BIOLINE) and 3 µL of genomic DNA (approximately 100 ng). Thermal cycles consisted: an initial denaturing step of 3 min at 94°C; 35 cycles of denaturing at 94°C for 30 s, annealing at 50°C for 90 s, and extension at 72°C for 90 s; and a final extension period of 5 min at 72°C. The PCR annealing temperature was increased to 60°C if unspecific results were obtained. Negative controls were included in all PCR runs to ascertain the lack of cross-contamination. The PCR results were verified using 1.5% agarose gel electrophoresis. Finally, the primer pairs that produced a single, clean amplicon were selected for subsequent steps (Fig. 2).

Marker selection by standard qPCR curve amplification

The PCR reactions of single-copy genes that demonstrated specificity for dreissenids (zebra and quagga mussels) were used to generate standard curves for the qPCR using an ABI 7300 Real-Time PCR System (Applied Biosystems). First, DNA stock concentrations were normalized to 50 ng/µL and were mixed in a final volume of 200 µL with 100 ng/µL salmon sperm DNA (Invitrogen) as a DNA carrier to minimize loss of zebra mussel DNA in the aliquots. The standard curve was constructed by five consecutive 10-fold dilutions (range of dilutions from 50 to 5E-3 ng/µL). The quantitative qPCR mix was prepared in 20 µL volume reactions with 2 µL of each DNA dilution, 1× SyBR® Green PCR Master Mix (Applied Biosystems) and 0.20 µM of each primer. The amplification temperature profile consisted of an initial step at 50°C for 10 min; 40 cycles of 95°C for 15 s and annealing at 60°C for 1 min; and a dissociation curve consisting of 95°C for 15 s, 60°C for 30 s and 95°C for 15 s. Negative controls were also included to ascertain the lack of cross-contaminations. All samples were tested in triplicate, and the coefficient of variation (CV) was calculated in the triplicates of cycle threshold (CT) to determine the absence of technical and manipulation errors.

The quantification level (QL), defined as the minimum of quantifiable DNA in the standard curve, was determined as the lowest DNA concentration with positive amplification and a CV lower than 0.3. The detection level (DL) was determined as the lower concentration with positive amplification but a CV larger than 0.3. The DL indicates DNA presence but no reliability

in the quantification. The dynamic range of the standard curve for every single-copy gene marker was determined using DNA concentrations up to QL and linear correlation. The marker producing the best efficiency and linearity values (lowest CV, QL and DL) was the one selected for dreissenid DNA quantification in the environmental samples. Positive PCR amplifications were directly sequenced by Sanger sequencing to reconfirm identity.

Stage B: Environmental quantification of dreissenid DNA presence

Filtered water sample collection

Samples were obtained from five representative locations (See Table 1 and Supplementary Fig. S4) of different reservoirs in the Northeast Iberian Peninsula following the sampling protocol and procedure of the control and management of the zebra mussel expansion plan [22]. In addition, a sample from Banyoles Lake (Girona, Spain) with no potential risk of zebra mussel invasion due to its isolated geographical distribution was used as a negative control. The presence of adults and larvae in these locations has been tested periodically[22] using the official Spanish Government procedures[26]. The positive presence of larvae and adult individuals was only detected in La Baells Reservoir since 2011, and ACA also detected larvae but no adults in Gaià Reservoir only in 2012[22]. The remaining three reservoirs were visually negative for the presence of both adults and larvae (See summary of presence of larvae and/or adults in Table 1). For quantitative analysis, two 100 L water samples were collected at each location before (spring) and after (autumn) spawning periods (Table 1). All of the environmental samples were filtered using a 50 μm diameter mesh to target veliger larvae of *Dreissena* species, rehydrated with water and stored at -20°C until analysis.

Environmental DNA (eDNA) isolation

One of the main points of this method is optimizing the eDNA extraction for all 11 water samples (Table 1) obtaining the most DNA but the lowest presence of PCR inhibitors. After several attempts, the best method resulted using the DNA isolation kit for environmental samples called the FastDNA™ SPIN Kit for Soil (MP Biomedicals). All DNA isolations were validated by 1% agarose gel electrophoresis and quantified using a QUBIT v2.0 fluorometer.

PCR amplification and Sanger sequencing validation

The eDNA samples were validated by endpoint PCR using the best molecular marker for dreissenid specificity previously selected in Stage A. Gradient PCR with an annealing temperature ranging from 60 to 70°C and a Touch-Down PCR with decreasing annealing temperature from 70 to 60°C (-1°C/cycle in the first ten cycles) were performed in parallel to optimize PCR amplification. The PCR products were also sequenced using the Sanger method for validation.

Quantification of infestation

The dreissenid DNA from the 11 samples was quantified by qPCR in three consecutive 10-fold dilutions in triplicate. To infer the effect of inhibitors in the eDNA samples, a second qPCR analysis was performed with controlled contaminations (Spikes) using the same dilutions of environmental samples but with the addition of 2 µL of 2 ng/µL zebra mussel DNA extracted from adult tissue. In all cases, the qPCR mix and thermal cycles were performed following Stage A conditions. The quantification results were analysed using 7300 SDS v1.3.1 Software (Applied Biosystems). For positive amplification samples with no inhibitors, concentrations were determined using the standard curve previously developed. The values are presented as the mean concentration (ng/L) with the corresponding standard error of the mean (SEM). A qPCR amplification was considered positive when the following conditions were met: (1) at least one of the dilutions is in the dynamic range of the standard curve and has a cycle threshold (CT) at least six cycles earlier than the no template control; (2) a proportional correspondence among decimal template dilutions and CT amplifications; (3) replicates should present a coefficient of variation lower than 0.3; and 4) specificity must be verified by a single melting peak. Finally, we calculated the estimated number of target gene copies per microliter in positive amplified samples.

The data resulting from the quantification was tested for normality by a Kolmogorov-Smirnov test, and values were transformed when necessary (exponential transformation for comparisons among seasons within the same location; and square root transformation for comparisons among the seasons). The statistical analyses for concentration comparisons were performed using two-tailed Student *t*-tests or ANOVA tests with Bonferroni post-hoc tests. All statistical analyses were computed using the IBM SPSS Statistics package (v. 20.0; IBM Corp., USA).

Acknowledgements

We are indebted to different colleagues for zebra mussel sample collections: Concha Durán (Ebro Hydrografic Confederation, Spain) for the logistic advice; Anna Terrats and Carolina Solà (Catalan Water Agency from the Government of Catalonia, Spain) for the filtered-water environmental samples. We would thank to Mariona Palacios for her collaboration in the analysis into her Final Master Project. This research was carried out within the objectives of the research project CGL200909407 of the Spanish Ministerio de Ciencia e Innovación (MICINN). LP received economical support with a PhD fellowship of the Spanish MICINN with reference BES – 2010037446. CA held a postdoctoral fellowship from the Spanish National Institute for Agricultural and Food Research and Technology (DOC-INIA programme).

References

1. Lowe, S., Browne, M., Boudjelas, S. & De Poorter, M. 100 of the World's worst invasive alien species: a selection from the Global Invasive Species Database. *Aliens*. **12**, 1–12 (2000).
2. Astanei, I., Gosling, E., Wilson, J. & Powell, E. Genetic variability and phylogeography of the invasive zebra mussel, *Dreissena polymorpha* (Pallas). *Mol Ecol*. **14**, 1655–1666 (2005).

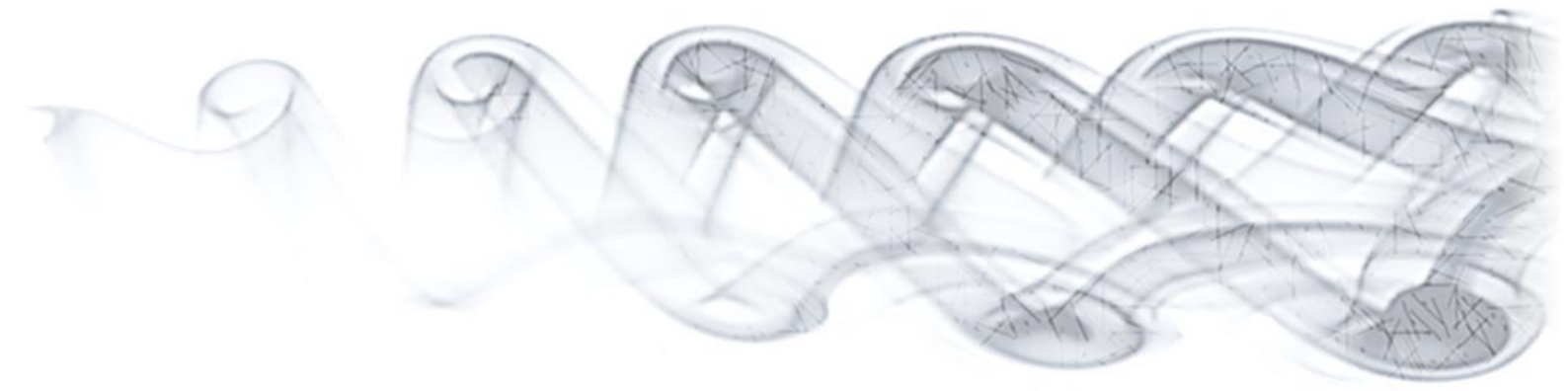
3. Stepien, C. A. *et al*. Evolutionary, biogeographic, and population genetic relationships of dreissenid mussels, with revision of component taxa in *Quagga and zebra mussel: biology, impacts, and control 2nd edn* (eds. Nalepa, T.F. & Schloesser, D.W.) 403-444 (CRC Press, 2014).
4. Egan, S. P. *et al*. Rapid molecular detection of invasive species in ballast and harbor water by integrating environmental DNA and light transmission spectroscopy. *Environ Sci Technol*. **49**, 4113−4121 (2015).
5. Johnson, L. E. & Carlton, J. T. Post-establishment spread in large-scale invasions: dispersal mechanisms of the zebra mussel *Dreissena polymorpha*. *Ecology*. **77**, 1686–1690 (1996).
6. Bij de Vaate, A., Jazdzewski, K., Ketelaars, H. A. M., Gollasch, S. & Van der Velde, G. Geographical patterns in range extension of Ponto-Caspian macroinvertebrate species in Europe. *Can. J. Fish Aquat Sci*. **59**, 1159–1174 (2002).
7. Pollux, B. J. A., van der Velde, G. & bij de Vaate, A. in *The Zebra Mussel in Europe* (eds. van der Velde G., Rajagopal S. & bij de Vaate A.) Ch. 4, 45-58 (Backhuys Publishers, Margraf Publishers, 2010).
8. Ruíz-Altaba, C., Jiménez, P. J. & López, M. A. El temido mejillón cebra empieza a invadir los ríos españoles desde el curso bajo del río Ebro. *Quercus*. **188**, 50-51 (2001).
9. Rajagopal, S. *et al*. Origin of Spanish invasion by the zebra mussel, *Dreissena polymorpha* (Pallas, 1771) revealed by amplified fragment length polymorphism (AFLP) fingerprinting. *Biol Invasions*. **11**, 2147-2159 (2009).
10. Bij de Vaate, A., van der Velde, G., Leuven, R. S. E. W. & Heiler, K. C. M. in *Quagga and zebra mussels: biology, impacts, and control 2nd edn* (eds Nalepa, T. F. & Schloesser D.) Ch. 6, 83-92 (CRC Press, 2013).
11. Peñarrubia, L., Sanz, N., Pla, C., Vidal, O. & Viñas, J. Using massive parallel sequencing for the development, validation, and application of population genetics markers in the invasive bivalve zebra mussel (*Dreissena polymorpha*). *PLoS ONE*. **10**, e0120732 (2015).
12. Peñarrubia, L., Vidal, O., Viñas, J., Pla, C. & Sanz, N. Genetic characterization of the invasive zebra mussel (*Dreissena polymorpha*) in the Iberian Peninsula. *Hydrobiologia*. DOI 10.1007/s10750-016-2819-2 (2016).
13. Ram, J. L., Karim, A. S., Banno, F. & Kashian, D. R. Invading the invaders: reproductive and other mechanisms mediating the displacement of zebra mussels by quagga mussels. *Int J. Invertebr Repr Dev*. **56**, 21-32 (2012).
14. Karatayev, A. Y., Burlakova, L. E. & Padilla, D. K. Zebra versus quagga mussels: a review of their spread, population dynamics, and ecosystem impacts. *Hydrobiologia*. **746**, 97–112 (2015).
15. Therriault, T. W., Orlova, M. I., Docker, M. F., MacIsaac, H. J. & Heath, D. D. Invasion genetics of a freshwater mussel (*Dreissena rostriformis bugensis*) in Eastern Europe: high gene flow and multiple introductions. *Heredity*. **95**, 16-23 (2005).
16. Molloy, D. P., bij de Vaate, A., Wilke, T. & Giamberini, L. Discovery of *Dreissena rostriformis bugensis* (Andrusov 1897) in Western Europe. *Biol Invasions*. **9**, 871-874 (2007).
17. Bij de Vaate, A. Some evidence for ballast water transport being the vector of the quagga mussel (*Dreissena rostriformis bugensis* Andrusov, 1897) introduction into Western Europe and subsequent upstream dispersal in the River Rhine. *Aquat Invasions*. **5**, 207-209 (2010).
18. Heiler, K. C. M. *et al*. Reconstruction of the early invasion history of the quagga mussel (*Dreissena rostriformis bugensis*) in Western Europe. *Aquat Invasions*. **8**, 53-57 (2013).
19. Bij de Vaate, A. & Beisel, J. N. Range expansion of the quagga mussel (*Dreissena rostriformis bugensis* Andrusov, 1897) in Western Europe: first observation from France. *Aquat Invasions*. **6**, 71-74 (2011).
20. Marescaux, J. *et al*. Unravelling the invasion pathways of the quagga mussel (*Dreissena rostriformis*) into Western Europe. *Biol Invasions*. **18**, 245-264 (2015).
21. MMARM (Ministerio Medio Ambiente, Rural y Marino) from Spanish Government. *Estrategia Nacional para el Control del Mejillón Cebra (*Dreissena polymorpha*) en España*. (2007) Available

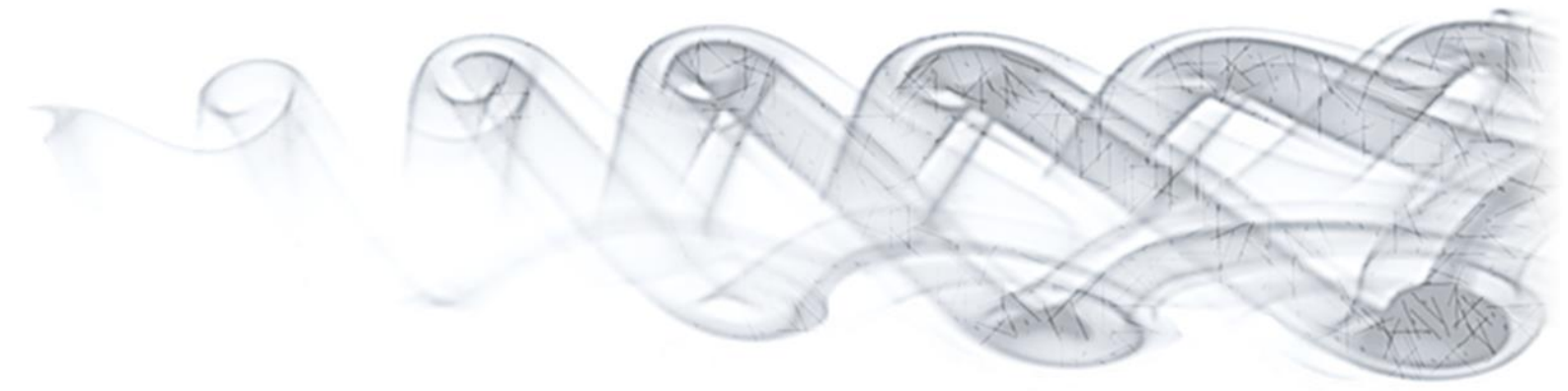
at: http://www.chebro.es/contenido.visualizar.do?idContenido=32951&idMenu=4240. (Accesed: 20th December 2013).

22. ACA (Catalan Water Agency). *El musclo zebrat. Estat de la presència a Catalunya – 2014*. (2015) Available at: http://aca-web.gencat.cat/aca/. (Accesed: 30th March 2015).
23. CHE (Confederación Hidrográfica del Ebro). *Monitorización de la presencia larvaria de Dreissena polymorpha en la Cuenca Hidrográfica del Ebro, campaña 2014*. (2015) Avaliable at: http://www.chebro.es/. (Accesed: 30th June 2015).
24. Sousa, R., Novais, A., Costa, R. & Strayer, D. Invasive bivalves in fresh waters: impacts from individuals to ecosystems and possible control strategies. *Hydrobiologia*. **735**, 233–251 (2014).
25. Lucy, F. Early life stages of *Dreissena polymorpha* (zebra mussel): the importance of long-term datasets in invasion ecology. *Aquat Invasions*. **1**, 171-182 (2006).
26. Durán, C., Lanao, M., Anadón, A. & Touyá, V. Management strategies for the zebra mussel invasion in the Ebro River basin. *Aquat invasions*. **5**, 309-316 (2010).
27. Baldwin, R. S. *et al*. A diagnostic molecular marker for zebra mussels (*Dreissena polymorpha*) and potentially co-occurring bivalves: mitochondrial COI. *Mol Mar Biol Biotechnol*. **5**, 9-14 (1996).
28. Frischer, M. E. *et al*. Specific amplification of the 18S rRNA gene as a method to detect zebra mussel (*Dreissena polymorpha*) larvae in plankton samples. *Hydrobiologia*. **487**, 33-44 (2002).
29. Ram, J. L. *et al*. Reproduction and genetic detection of veligers in changing *Dreissena* populations in the great Lakes. *Ecosphere*. **2**, 1-16 (2011).
30. Valdopalas, B., Bouma, J. V., Jackels, C. R. & Friedman, C. S. Application of real-time PCR for simultaneous identification and quantification of larval abalone. *J. Exp Mar Biol Ecol*. **334**, 219-228 (2006).
31. Quinteiro, J. *et al*. Quantification of manila clam *Ruditapes philippinarum* (Adams & Reeve, 1850) larvae based on SYBR Green Real-Time Polymerase Chain Reaction. *J. Shelfish Res*. **30**, 791-796 (2011).
32. Dejean, T. *et al*. Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog Lithobates catesbeianus. *J. Appl Ecol*. **49**, 953–959 (2012).
33. Pilliod, D. S., Goldberg, C. S., Arkle, R. S. & Waits, L. P. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Can J. Fish Aquat Sci.* **70**, 1123–1130 (2013).
34. Takahara, T., Minamoto, T. & Doi, H. Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS ONE*. **8**, e56584 (2013).
35. Thomsen, P. F. *et al*. Monitoring endangered freshwater biodiversity using environmental DNA. *Mol Ecol*. **21**, 2565–2573 (2012).
36. Thomsen, P. F. & Willerslev, E. Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biol Conserv.* **183**, 4–18 (2015).
37. Gál, A. B. *et al*. Comparison of real-time polymerase chain reaction and end-point polymerase chain reaction for the analysis of gene expression in preimplantation embryos. *Reprod Fert Develop.* **18**, 365-371 (2006).
38. Ellison, S. L. R., English, C. A., Burns, M. J. & Keer, J. T. Routes to improving the reliability of low level DNA analysis using real-time PCR. *BMC Biotechnol*. **6**, 33 (2006).
39. Bustin, S. A. *et al*. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*. **55**, 611-622 (2009).
40. Hoy, M. S., Kelly, K. & Rodriguez, R. J. Development of a molecular diagnostic system to discriminate *Dreissena polymorpha* (zebra mussel) and *Dreissena bugensis* (quagga mussel). *Mol Ecol Resour*. **10**, 190-192 (2010).

41. Burlakova, L. E. *et al*. Competitive replacement of invasive congeners may relax impact on native species: interactions among zebra, quagga, and native unionid mussels. *PLoS ONE*. **9**, e114926 (2014).
42. Karatayev, A. Y., Burlakova, L .E., Mastitsky, S. E., Padilla, D.K. & Mills, E. L. Contrasting rates of spread of two congeners, *Dreissena polymorpha* and *Dreissena rostriformis bugensis*, at different spatial scales. *J. Shellfish Res*. **30**, 923–931 (2011).
43. Marescaux, J. *et al*. Sympatric *Dreissena* species in the Meuse River: towards a dominance shift from zebra to quagga mussels. *Aquat invasions*. **10**, 287-298 (2015).
44. Matthews, J. *et al*. Rapid range expansion of the invasive quagga mussel in relation to zebra mussel presence in The Netherlands and Western Europe. *Biol Invasions*. **16**, 23-42 (2013).
45. Eichmiller, J. J., Przemyslaw, G. B. & Sorensen, P. W. The relationship between the distribution of common carp and their environmental DNA in a small lake. *PloS ONE*. **9**, e112611 (2014).
46. Newton, J., Sepulvedra, A., Sylvester, K. & Thum, R. A. Potential utility of environmental DNA for early detection of Eurasian watermilfoil (*Myriophyllum spicatum*). *J. Aquat Plant Manage*. **54**, 46-49 (2015).
47. Ackerman, J. D., Blair, S., Nichols, S. J. & Claudi, R. A review of the early life history of zebra mussels (*Dreissena polymorpha*): comparisons with marine bivalves. *Can J. Zool*. **72**, 1169-1177 (1994).
48. Conesa, A. *et al*. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. **21**, 3674-3676 (2005).
49. Cunningham, F. *et al*. Ensembl 2015. *Nucleic Acids Res*. **43**, D662-D669 (2015).
50. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*. **132**, 365-386 (2000).

4. DISCUSSION





4. DISCUSSION

4.1. Massively Parallel Sequencing in search of new molecular markers

The implementation of MPS has improved the identification of new molecular markers. Large databases generated by MPS methods can be bioinformatically screened to detect SSRs (Gardner *et al*. 2011) or SNPs (Everett *et al*. 2011). As previously mentioned, SSR and SNP markers are genetic markers commonly used in population genetics, and both provide valuable information for conservation strategies and management (Sunnucks 2000). Specifically, this doctoral thesis has seeked to identify and characterize new polymorphic SSRs in both zebra mussel and the Asian clam (chapters 1 and 2) using sequences obtained from 454 Pyrosequencing, to be used in subsequent population analyses in the Iberian Peninsula (Additional Figure 7.1.2). Additionally, the same 454 Pyrosequencing runs used for SSR developing have also been used to increase the number of available SNP markers in both species (chapter 3).

The search of molecular markers such as SSRs involves two main steps: (1) detection of candidate regions (in this case, tandem repeat structures) throughout the genome, and (2) validation with genotyping protocols. Because of this further validation, low coverage is not usually an issue, and it has been established that the average proportion of a 454 Pyrosequencing plate required to identify a set of 40 unique pre validated SSR loci in invertebrates is 7.25% (a 1/14 fraction) of a 454 plate (Gardner *et al*. 2011). Even in larger genomes this value is adequate. In the largest estimated haploid genome (the conifer *Callitris verrucosa*, Cupressaceae) included in an analysis of SSR characterization using 454 Pyrosequencing (Gardner *et al*. 2011), a 1/6 (16.67%) of a single plate was enough to obtain 117 pre validated loci with their corresponding primers. This same study includes information about successful conversion from selected loci to polymorphic loci (polymorphic yield), which is 62% in invertebrates. Thus, a 7.25% of a 454 plate should be enough to obtain a viable set of polymorphic SSRs. We therefore used 1/8 of a 454 plate (12.5%) for each species.

We were able to detect candidate regions containing tandem repeats using DNA from single individuals in our MPS libraries construction. Sequencing a single individual could generate a larger representation of the genome and more tandem repeat structures might be detected compared to from the use of pooled individuals, the total number of sequenced nucleotides being equal. Posteriorly, these single individual derived markers were analyzed in several specimens to detect polymorphism. However, the search of SNP in such MPS runs may yield a significant number of false positives (as well as false negatives) during the initial detection of base variation. As we will discuss, only heterozygous SNP were detected, and these may be related to genomic duplications and not to real polymorphisms.

4.1.1. SSRs

The sequencing of 12.5% of a 454 plate in zebra mussel generated 36,554,817 bases in 110,593 reads. Parallely, Asian clam sequencing resulted in the similar values of 29,474,243 bases in 98,534 reads. These yields in massive sequencing are in concordance with the normal range of

previous studies using 454 Pyrosequencing for isolating SSRs in other bivalves. For instance, a total of 63,715,316 bases (161,714 reads) were obtained in the Texas hornshell (*Popenaias popeii*, Lea, 1857) (Inoue *et al*. 2013), and 420,204,057 bases (859,313 reads) were also generated in the Chinese razor clam (*Sinonovacula constricta*, Lamarck, 1818) (Niu *et al*. 2013) using a whole 454 plate. Calculated yields in sequencing for these studies if they would have used the same proportion of the plate than us (12.5%) would be 31,857,658 bases (80,857 reads) and 52,525,507 bases (107,414 reads) respectively.

After the assembly, 3,885 zebra mussel and 3,347 Asian clam contigs were build and screened, with a total of 288 and 246 single tandem repeats, respectively. Finally, 93 of the 288 detected tandem repeats and 97 of the 246 tandem repeats were unique loci conformming the conditions for primer design, respectively. After PCR validation in 15 individuals from different locations, 14 out of 93 (15.05%) loci in zebra mussel and 9 out of 97 (9.28%) in the Asian clam were polymorphic. These values of polymorphism are lower than expected, considering the 62% of polymorphism yied reported in other invertebrates (Gardner *et al*. 2011).

However, our results are similar to those found in some previous MPS studies using 454 Pyrosequencing for SSR isolation in bivalve species, both in number of candidate tandem repeats and validated polymorphic loci. We identified one single tandem structure every 5,677 screened nucleotides in zebra mussel and one every 5,309 screened nucleotides in Asian clam (one repeat structure every 13 screened contig sequences in both cases). These results are in concordance with previous MPS studies using 454 Pyrosequencing for SSR discovering in the Korean mussel (*Mytilus coruscus*, Gould, 1861) (Kang *et al*. 2013), in the Antarctic bivalve (*Laternula elliptica*, King & Broderip, 1832) (Clark *et al*. 2010), or in the Chinese razor clam (*Sinonovacula constricta*, Lamarck, 1818) (Niu *et al*. 2013). These studies obtained one tandem repeat structure in 6,698, 6,085, and 6,527 bases respectively, representing one single repeat structure every 11-13 screened sequences. Our final polymorphic SSR sets (14 in the zebra mussel and nine in the Asian clam) and our polymorphic yield values (15.05% and 9.28% in the zebra mussel and the Asian clam) also were similar than previous studies. For instance, a final set of 18 polymorphic SSRs were finally described in the Korean mussel (Kang *et al*. 2013). Other studies in the Korean manila clam (*Ruditapes philippinarum*, Adams & Reeve, 1850) (Kim *et al*. 2014) and in the thick-shelled river mussel (*Unio crassus*, Retzius, 1788) (Sell *et al*. 2013) described 10 SSRs each. Polymorphic yield in these 3 species resulted in 10.23%, 16.13%, and 14.29%, respectively.

It is interesting to note that none of the other bivalve species for which SSR have been described are invasive, and thus the comparatively low polymorphic yield could be attributed to idiosyncratic genomic traits of these taxa, and not to invasiveness. Bivalves have large genomes (Saavedra and Bachere 2006) and a low frequency of SSRs (Zane *et al*. 2002). This feature may be decreasing the power of the approach, and a higher number of initial candidate sequences may be required to obtain a similar number of polymorphic SSRs than other invertebrate taxa (Gardner *et al*. 2011). Additionally, some studies suggest that PCR design for SSR amplification in mollusks may not be optimal because of the high incidence of cryptic repetitive DNA in their flanking regions (McInerney *et al*. 2011).

In a general way, MPS platforms have drastically increased the number of polymorphisms obtained per single sequencing run whiht a decreasing costs per base (Metzker 2010) while improvements in *de novo* assemblies and higher quality of both alignments and assembly strategies have increased final coverage per run and thus marker discovery. However, different MPS platforms have different length-coverage rates (Metzker 2010). Although 454 Pyrosequencing systems generate longer sequences compared to other platforms (Droege and Hill 2008), Illumina provides shorter sequences at a lower cost, therefore providing better coverage. Interestingly, the detection of tandem repeats for SSR discovery doesn't require high coverage in the sequencing step, although sufficient sequence length is needed for successful primer design in their flanking regions, and the use of 454 MPS is then recommended.

4.1.2. SNPs

Whereas for SSR capture larger assemblies are preferred, both for identification of the structural motif and for primer design, for SNP calling it is better to get better coverage for each base position. Thus, to identify SNP in our genomic data, we reassembled the 454 Pyrosequencing reads of chapters 1 and 2 in order to increase the number of sequences in the alignment. Following the criteria of Altmann *et al*. (2012) and Seeb *et al*. (2011), we considered a minimum of 40% of sequence overlap in contrast of a minimum of 40 bp criterion used for the SSR capture. A 40 bp criterion is much more restrictive because smaller reads will be trimmed and not included in the final assembly. Assemblies including smaller reads may contain contigs of lower quality, yet they are likely to increase coverage and to identify more variable positions (Seeb *et al*. 2011). For the zebra mussel, such assembly conditions increased the number of contigs from 3,885 (average size = 421 bp) to 12,840 (average size = 424 bp), and for the Asian clam the increase was from 3,347 (average size = 390 bp) to 10,853 (average size = 402 bp). Bioinformatic base variants detection resulted in 783 positions in the zebra mussel and 446 positions in the Asian clam, including single nucleotide variants (transitions and transversions), multiple nucleotide variants (changes of two or three consequtive bases), and Insertions/Deletions (InDels).

Next, we selected a subset of these variants for laboratory validation and polymorphism confirmation. In this case, validation was performed by High Resolution Melting Analysis (HRMA), to assess polymorphisms based on differences in DNA melting temperatures (*Tm*) of different sequence fragments (Reed *et al*. 2007). SNPs can be discerned because even a single base change may result in different melting curves (Smith *et al*. 2012). Thus, both homozygote and heterozygote genotypes present differently shaped melting curves (Liew *et al*. 2004). The main advantage of HRMA is its easy-to-use genotyping of nucleotide variants (McGlauflin *et al*. 2010). In order to increase the resolution among genotypes during HRMA, we excluded multiple nucleotide variants and InDels. We therefore included single variants with a non polymorphic 50 bp window, plus the flanking regions for primer design. Thus, amplicons of less than 65 bp were produced (Smith *et al*. 2012). A total of 46 (5.87% of total) base variants in the zebra mussel and 40 (8.97% of total) in the Asian clam met these criteria. After HRMA, a set of 5 and 4 base variants (10.87% and 10% of the analyzed loci, respectively) presented polymorphism.

Previous studies using 454 Pyrosequencing for SNP discovery in bivalves have better polymorphic yields. For instance, Gallardo-Escarate *et al*. (2015) also sequenced 1/8 of a 454 Pyrosequencing plate in the surf clam (*Mesodesma donacium*, Lamark, 1818). The authors obtained 10,178 contigs with average contig size of 581 bp. Such values are similar to our zebra mussel and Asian clam MPS outputs. However, they detected 2,594 variants, compared to the lower yield of 783 variants in zebra mussel or the 446 variants in Asian clam reported here. Interestingly, these authors pooled 15 surf clams to perform their MPS analysis, and the lower power of detection of our approach might be attributed to the design of our experiment, which focused on SSR detection and was based in one individual only. Using pooled samples to construct the DNA library increases the likelihood of finding base polymorphisms, because a better representation of the alleles is available.

During HRMA validation, we found in several loci a heterozygote pattern in all analyzed individuals. This anomalous heterozygosity rate may be related to the presence of paralogous sequence variants (PSVs). Thus, we would be detecting the occurrence of slightly different alleles on paralogous (duplicated) genes rather than real SNPs on homologue regions (Smith et al. 2005). These PSV could be a consequence of the high transpositional activity reported for the mollusk genomes (McInerney *et al*. 2011), and they are very difficult to overcome. In fact, even setting a stringency criterion at a 90% of sequence identity, would also assemble paralogues (Seeb *et al*. 2011). Again, increasing the number of individuals in library construction would neither avoid the PSV detection, but these false positives would be diluted with real SNPs, thus improving the number of real base variants initially detected and their polymorphic yield in SNPs validation.

If SNP are the targeted marker of the experimental design, there are other new methods that could improve the results, such as Restriction-site-Associated DNA Sequencing (RAD-seq) and Genotyping by Sequencing (GBS). Both methods avoid the random fragmentation of the DNA during the construction of the genomic library, a common first step of MPS protocols. This random fragmentation causes that different portions of the genome are sequenced in each individual, and the alignment is restricted to those parts in common. RAD-seq use restriction enzymes during this initial DNA fragmentation, and because the same fragments are generated in different individuals, the genomic overlap increases (Davey *et al*. 2011; Elshire *et al*. 2011). Furthermore, polymorphism detection and population analyses are performed simultaneously (Cristescu 2015).

4.2. Zebra mussel and Asian clam invasions and spread in the Iberian Peninsula

Invasions of the zebra mussel and the Asian clam in the Iberian Peninsula exhibit different histories and distributions. The first record of the Asian clam was documented in 1980 in the Tajo River estuary (Mouthon 1981). In contrast, the zebra mussel was probably introduced more recently, with a first report in the low reaches of the Ebro River, in Ribarroja Reservoir, in 2001 (Ruíz-Altaba *et al*. 2001). Since then, both species have been continuously spreading through the Iberian freshwater basins (Figure 1.2), following different pathways. Currently, whereas zebra mussel is mainly distributed in the Northeast of the Iberian Peninsula (Figure 1.2A), Asian clam has a more extended distribution and is found in all main freshwater

ecosystems (Figure 1.2B). Although both bivalve species share some biological and invasiveness factors, our genetic studies reported differences in their invasion history pattern in the Iberian Peninsula.

4.2.1. Genetic characterization of the zebra mussel invasion

The main result of our genetic characterization of Iberian locations (Additional Table 7.2.3) using SSR markers revealed that the zebra mussel invasion in the Iberian Peninsula could be considered a separate episode from the other European invasions. Our results partially disagree with previous studies using AFLPs, which locate the origin of the Iberian invasion in France (Rajagopal *et al*. 2009). Our analyses stablish the independent origin of the Iberian samples, and they group the northern French basins with the rest of European sites as reported by May *et al*. (2006). Both studies support the scenario of a single origin for all these north-of-the-Pyrenees non-native populations.

However, when including the South of France some discrepancies appear. Rajagopal *et al*. (2009) clustered the Southern France samples with the Iberian populations, in concordance with some results of Tarnowska *et al*. (2013) describing these Southern basins as presenting genetic divergence from the rest of French basins. Interestingly, though, this latter study found that the Southern French region shared some of the most abundant COI haplotypes found in the rest of French basins, and discarded the hypothesis of separate introduction episodes (Tarnowska *et al*. 2013). Therefore, it has been considered that zebra mussel in France conforms a single population, which we have included in our study through the Moselle river individuals.

Being this the scenario, the origin of the Iberian samples is then conflicted, as it is not related to the French population when using SSR, although AFLPs detect some kind of proximity between Iberian and Southern France individuals (Rajagopal *et al*. 2009). In general, SSR are considered to provide more reliable information than AFLP (Liu and Cordes 2004), so the hypothesis of an independent origin of the Iberian samples is well supported. The identification of the south of France as a mixture region could suggest that the Spanish and European invasions maintain a post-invasion gene flow.

Interestingly, our results could be compatible with a direct invasion from the native range. Nevertheless, previous studies (Müller *et al*. 2002; Gelembuik *et al*. 2006), after analyzing the native populations in the Black-Ponto-Caspian Sea basins, concluded that fluctuations and instability of genetic diversity in the native zebra mussel populations generated a mixture of different genetic populations in the native range. As a consequence, the exact identification of the original invasion sources in this species could be difficult and significantly affected by stochastic sampling effects, even with more extensive analysis of the native range. As we discussed in chapter 4, further sampling may be helpful to discover the source of the Spanish invasion. However, the original population may have already changed significantly, thus making very difficult to make assessments.

High genetic homogeneity among Iberian sites revealed that all Iberian locations belong to the same genetic population. First invasion in 2001 in the middle course of the Ebro River (Ruíz-Altaba *et al*. 2001) seems to be produced by a large and diverse set of individuals that recovered quickly from the founder event. Since then, the zebra mussel has spread along the river and has reached neighboring basins, following a similar step-wise model than in France (Tarnowska *et al*. 2013). This dispersion has likely been carried out through the usual strategies of the zebra mussel including attachment to hulls of commercial and recreational ships, as well as veliger larvae surviving in ballast water (Johnson and Carlton 1996). Interestingly, one hypothesis about this independent introduction in the Iberian Peninsula is the direct transport from the native range by tourism related activities and/or transport of exotic species. In fact, globalization has facilitated the spread of exotic species from the East to the Western Europe including other macroinvertebrate freshwater feeding invaders, such as the freshwater hydroid (*Cordylophora caspia*, Pallas, 1771), the gravel snail (*Lithoglyphus naticoides*, Pfeiffer, 1828), or the freshwater mite (*Halacarellus hyrcanus*, Viets, 1928) as revised in Bij de Vaate *et al*. (2002) or in Havel *et al*. (2015).

In general, populations stablishing in a new area are usually originated from a small number of individuals, and suffer a reduction of genetic diversity (Sakai *et al*. 2001). However, initial settlements of zebra mussel are usually originated from individuals belonging to more than one location and/or from repeated translocations (Brown and Stepien 2010). As a consequence, these species are supposed to avoid loss of genetic diversity and bottleneck effects (Blackburn *et al*. 2015). Accordingly, our genetic analyses reported similar genetic variability values in native and introduced zebra mussel populations. However, our results suggested that the zebra mussel invasion in the Iberian Peninsula has been produced by a single event; therefore we could attribute these high genetic diversity values to a large number of individuals initially introduced and not to repeated translocations.

It has been suggested that the high transpositional activity present in bivalve genomes also could explain their rapid adaptation to new habitats (McInerney *et al*. 2011). Stress-induced changes could alter the action of these genes and thus promote structural variation, which, in turn may facilitate the rapid adaptation observed in new environments (Stapley *et al*. 2015). These genetic variations could generate the high morphological plasticity of zebra mussel and provide the basis for genetic adaptations in new habitats (Stepien *et al*. 2014). Specifically, our genetic results suggest that an incipient differentiation is already occurring in the Llobregat River. The introduction in this river was first documented in 2011 (chapter 4), ten years after the first introduction in the Ebro River. Therefore, this short period of time could be sufficient to genetically differenciate populations in zebra mussel invasions (Stockwell and Ashley 2004; Dlugosch and Parker 2008).

4.2.2. Genetic characterization of the Asian clam invasion

The relationship among the different Asian clam lineages is difficult to assess in phylogenetic analyses because this species reportedly displays asexual reproduction by androgenesis in invaded areas (Pigneur *et al*. 2012). Thus, the genetic characterization may yield different results depending on the molecular markers of choice. Furthermore, nuclear SSR and SNP

markers optimized in the present thesis (chapters 2 and 3 respectively) are not numerous enough and/or not polymorphic enough to attain a reliable description of androgenetic populations. In this sense, the best combination of molecular markers seems to be a set of one mitochondrial and one nuclear marker (COI and 28S genes, respectively) (Lee *et al*. 2005; Hedtke *et al*. 2008). Together they provide information of both paternal (nuclear) and maternal (mitochondrial) inheritance.

Our population genetic study of the Iberian locations (Additional Table 7.2.4) showed for the first time that Iberian *Corbicula* populations are a mixture of the three most representative invasive lineages (chapter 5). These three Asian clam lineages are commonly present in other invaded European locations (Pigneur *et al*. 2014a), and the most common lineage within the Iberian locations (RA lineage) corresponded to the most abundant lineage throughout Europe and North America (Hedtke *et al*. 2008; Pigneur *et al*. 2014a). Since the first invasion in the estuaries of the western Tajo River in 1980 (Moutthon 1981), this Asian clam lineage seems to have extended east to the remaining Iberian basins. Moreover, our results indicate that an independent colonization episode occurred in the estuaries of the Ebro River, where the first reported Asian clam were found in 1996 (López and Altaba 1997). Accordingly, the Ebro River harbors a genetically distinct population. This second episode has spread upstream, along the Ebro River, maybe because of the crawling capacity of the clam larva at first stages of development (Karatayev *et al*. 2005). Both invasion episodes converge in the middle Ebro River, where overlapped mitochondrial-nuclear haplotypes from eastern and western colonizations are found. According to our results, the adjacent basins in the East of the Iberian Peninsula (like the Ter River) may have been invaded from this middle Ebro River section.

Additionally, our results revealed high number of 28S nuclear heterozygotes, with alleles from divergent lineages in the same individuals. This feature likely confirms a predominant use of androgenetic reproduction (with egg parasitism involving two lineages) in these populations. Interestingly, heterosis in founder populations has been linked to rapid adaptation (Drake 2006). This heterosis, rather than selection of increased genetic diversity, would be responsible for the increased fitness and the major success on the colonization of hybrid individuals.

The results also suggested secondary contacts between Iberian Asian clam populations and other European sites (chapter 5). This admixture, if confirmed, could increase genetic variation related to ecologically important traits or to novel phenotypes. Then adaptation to new environments could be facilitated (Colautti and Lau 2015). Additionally, the invasive success of the androgenetic Asian clam lineages may be linked to their asexual mode of reproduction (Pigneur *et al*. 2012), which we have most likely detected in the Iberian populations. A single androgenetic individual can therefore establish new Asian clam populations.

4.2.3. Genetics of invasion

Although the zebra mussel and the Asian clam are frequently described using the same mechanisms of invasions, in the Iberian Peninsula we have detected some differences. Thus, the zebra mussel invasion seems to have been produced by a single introduction episode, whereas the Asian clam invasion involves multiple episodes.

However, both zebra mussel (chapter 4) and Asian clam (chapter 5) showed a genetic diversity similar to the one of their native populations. The relevance of genetic variation in invasive species is not completely clear, and even though its loss is considered a difficulty for the survival of the population (Sakai *et al*. 2001), different invasive populations may present normal values of genetic diversity. Population dynamics may strongly influence these variations, and if bottlenecks are moderate (when large numbers of individuals are transported) and/or population growth rapid, the genetic diversity may remain unaltered (Roman and Darling 2007; Suarez and Tsutsui 2008).

In general, in zebra mussel and Asian clam, the similar values of genetic diversity between native and invasive Iberian populations can be attributed to (1) large amount of individuals in the first step of the infestation, as both species can be easily carried along by human transport activities allowing opportunistic transport (Sousa *et al*. 2014) and/or (2) the release of large numbers of offspring (Astanei *et al*. 2005; Sousa *et al*. 2008). In any case, the relatively high values we have found in the Iberian populations may have been playing an important role for the invasion success (Stockwell and Ashley 2004).

Moreover, multiple introductions may be important sources of genetic variation necessary for adaptation to new environments and thus can be critical to a successful establishment of introduced species (Dlugosch and Parker 2008; Suarez and Tsutsui 2008). When multiple introductions of a species occur, introduced populations may attain higher levels of genetic diversity than native populations. Individuals in these populations may possess alleles in combinations that do not exist in the native range, thus increasing the likelihood of novel epistatic interactions or the expression of new phenotypes (Suarez and Tsutsui 2008; Lau and Terhorst 2015). This maybe the situation in the Iberian Asian clam populations, and even in the zebra mussel: although we have identified a single origin for the Iberian populations, the source area could be already harboring new combinations of adaptive alleles due to previous contacts between divergent populations.

Many invasive species reach high densities in the invaded range, and these large populations may adapt faster to changing environments by increasing the probability of adaptive mutations arising (Colautti and Lau 2015). Thus, the high densities achieved by the zebra mussel and the Asian clam species could be a fundamental feature for its invasion success, at least if this success is related to fast genetic change and adaptation. In fact, both zebra mussel and Asian clam species have adapted to periodically disturbed habitats (McMahon 2002). Although the SSR markers we have used are considered to be neutral, the incipient differentiation we have detected in the zebra mussel of the Llobregat River could be an indication of adaptive genetic variations (Stockwell and Ashley 2004).

4.3. Molecular protocols for detection

Eradication of invasive bivalves is extremely difficul to achieve once the population is established, and thus a fast detection method is crucial to stop the spread. As we have already detailed, current Spanish plans for control and management including early detection are only implemented for zebra mussel.

Therefore, we focused on developing a genetic method to improve the sensibility and the specificity of routine zebra mussel detection systems based on qPCR (chapter 6). To do so, we identified a dreissenid specific DNA sequence, the Histone-2B gene that was amplified by Real Time PCR. Although quantification of larvae should be possible, we have not been able to construct a standard curve to cross DNA amplification results with larvae count, and it is not clear if the differences we have reported between locations are related to different levels of infestation and/or yearly oscillations. Further analyses to check both sampling and season biases should be required. Moreover, a DNA extraction method yielding environmental free DNA (not only DNA from plankton samples) could also improve our results, as this DNA might be less sensitive to these variations. Previous studies developing other molecular detection methods don't address these problems, because they either use conventional PCR (Frischer *et al*. 2002; Egan *et al*. 2015) or do not sample intending quantification (Ram *et al*. 2011). Even though our genetic protocol has limitations, it has been able to detect zebra mussel presence in waterbodies where visual protocols yielded negative results. Then, this probably reflects better detection sensibility. When visual methods fail to detect larvae in waterbodies where molecular methods do, either an incipient invasion or a location with characteristics to "resist" complete invasion could be hypothesized. In both cases, these sites should be considered major objectives of analysis by the governmental agencies.

In contrast, there are no molecular protocols for detection of Asian clam. Because Asian clam larvae are not floating in plankton (Hedtke *et al.* 2008), any attempt should be focused on environmental DNA, which is going to be the only floating trace of the invasion. A Real-Time PCR protocol with the capacity of quantification would be of great interest, as clam colonies are much less evident than aggregations of zebra mussel. Our MPS output could be used to identify single copy nuclear markers suitable for such design.

4.4. Future challenges in control and management

4.4.1. Zebra mussel control plans

The high economic impacts produced by the zebra mussel have concentrated all the measures included in the Spanish control and management plans. The relatively late invasion in Spain has allowed the government to design monitoring strategies to prevent large scale expansion across the country. These strategies are developed through the National Strategy of Zebra Mussel Control in Spain (MMARM 2007), a program involving all the Spanish Hydrographic Confederations. This National Strategy includes integrated ecological and cost-effective management plans, such as informative campaigns, the construction of disinfection stations, closing of uncontrolled accesses, and research on the species and its interaction with the environment in affected and in unaffected areas (Durán *et al*. 2010). Because zebra mussel distribution is mainly found in the Northeast of the Iberian Peninsula (Figure 1.2A), most of the control plans belong to the Ebro Hydrographic Confederation (CHE) (http://www.chebro.es/) and the Catalan Water Agency (ACA) (http://aca-web.gencat.cat/aca/appmanager/aca/aca/) (Abel-Abellán 2010, ACA 2015, CHE 2015). Another important governmental agency is the Basque Water Agency (URA) (www.uragentzia.euskadi.net), which has developed an Action Plan for the Zebra Mussel Expansion Control in the Basque Country, in the Northeast of Spain

(URA 2013). Moreover, because zebra mussel has also been punctually detected in the Guadalquivir River, the Guadalquivir Hydrographic Confederation (CHG) (http://www.chguadalquivir.es/) has focused on preventing further invasions along the southern basins. The CHG is now developing a monitoring plan for detecting zebra mussel DNA (http://www.chguadalquivir.es/).

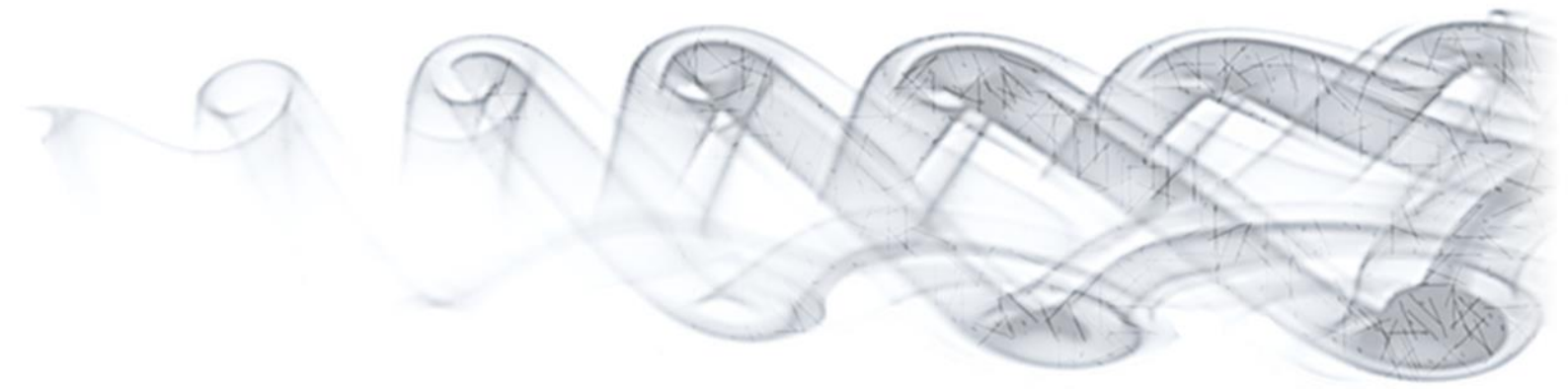
Our results may be indicating that the implementation of the National Strategy of Zebra Mussel Control in Spain in 2007 was partially successful, and the zebra mussel spread has become mostly controlled. All the Iberian locations belong to the same population, demonstrating that no further introductions have occurred in the Iberian basins (chapter 4). Zebra mussel spread seems to be by passive diffusion among Northeast Iberian Peninsula Rivers. However, a large-scale introduction event may have happened in the Guadalquivir River. Thus, additional measures may be necessary to avoid further expansion of this species within the country. Our detection protocol shows freshwater ecosystems with no reported zebra mussels but with detection of dreissenid DNA (chapter 6), indicating that the invasion region may be larger that was initially expected.

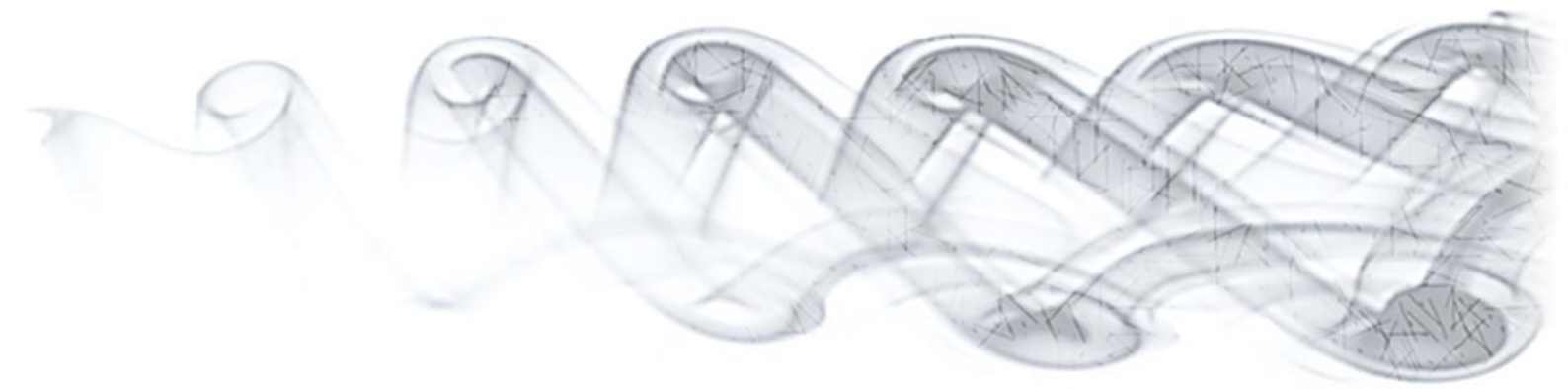
Based on the invasion history among the Ebro river basin, we can hypothesize that the recent introduction in the Guadalquivir River (or any river in the initial stages of invasion) will progress with the same pattern: it seems likely that zebra mussel will expand throughout the whole basin by passive diffusion following the step-wise spread model. In addition, the presence of zebra mussel in additional freshwater bodies than Ebro River puts on risk all other river basins of the Iberian Peninsula since all new infested locations can act as source populations of invasive species. Thus, it would be interesting to implement an efficient detection system for monitoring the as more water bodies as possible. The genetic methodology optimized in this thesis could also be used for detecting specifically dreissenid DNA into monitored water samples, and prevention efforts could be focused on sites where dreissenid DNA was detected by the first time and/or in sites when dreissenid presence is increasing.

4.4.2. Asian clam control plans

Asian clam monitoring remains secondary in priority within the Spanish plans for constraining invasive species. The different River Confederations have only developed strategies to detect adults (CHG 2014; CHS 2014), but no management plans to prevent future spread are currently stablished. Fortunately, it seems that strategies carried out to control the zebra mussel contain the Asian clam. When the National Strategy of Zebra Mussel Control was implemented in 2007, the Asian clam had already been present in several Iberian basins for almost 30 years. Interestingly, based on our compilation of the Asian clam presence in the Iberian Peninsula (Figure 1.2B), its dispersion after 2007 has been mainly controlled, and no more large-scale introduction episodes have occurred in the Western area. However, some Eastern basins adjacent of the Ebro River have been invaded (Figure 1.2B). It could be useful develop similar genetic methodologies to monitor locations with high risk of future invasions.

5. CONCLUSIONS





6. CONCLUSIONS

From objective 1: Development of new genetic markers using MPS technologies for the zebra mussel and the Asian clam (**chapters 1, 2 and 3**).

1. We have validated new sets of 14 polymorphic SSR markers for the zebra mussel and 9 SSRs for the Asian clam.

2. We also have characterized 5 new polymorphic SNP markers for the zebra mussel and 4 for the Asian clam using High Resolution Melting Analysis.

From objective 2: Genetic characterization of Iberian populations to understand their invasion history in the Iberian Peninsula (**chapters 4 and 5**).

3. The zebra mussel invasion in the Iberian Peninsula involves a single and recent invasion episode, with a step-wise expansion model during the colonization of the Northeast Iberian Peninsula from the initial introduction in the middle Ebro River stream.

4. The introduction of the zebra mussel in the Iberian Peninsula may come from an independent colonization event not related to other European sites, disagreeing with previous studies situating its origin on French populations.

5. The Asian clam invasion in the Iberian Peninsula involves the three major lineages described in the invaded range of this species.

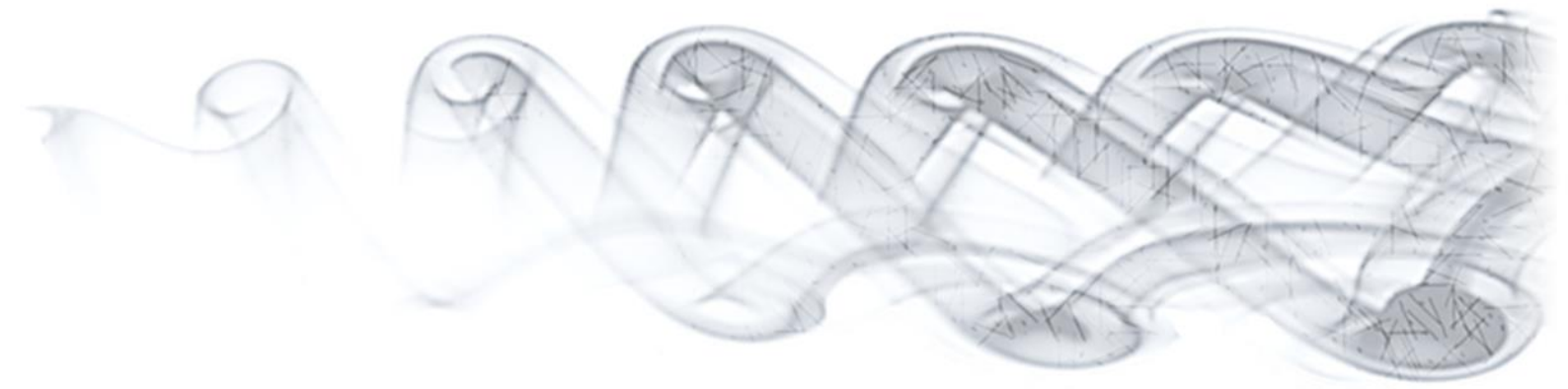
6. In Asian clam, based on both nuclear and mitochondrial analyses, at least two colonization episodes have occurred in the Iberian Peninsula, each one derived from different invasive androgenetic lineages. The first one involves the main area of distribution of the Asian clam in the Iberian Peninsula, and the second episode started in the eastern Peninsula at the low Ebro River section, spreading upstream throughout the Ebro River.

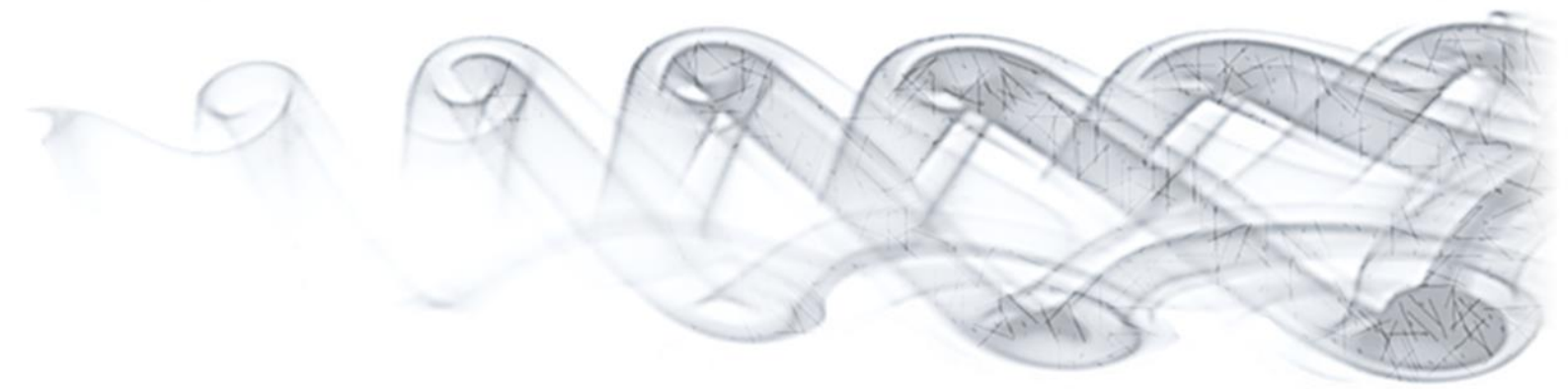
7. In Asian clam, both Iberian colonization pathways converged in the middle Ebro River, where haplotypes from both episodes are present. Populations from adjacent Eastern basins like Ter River seem to be originated from this middle Ebro River section.

From objective 3: Development a genetic method for monitoring the zebra mussel detection in environmental samples (**chapter 6**).

8. We have developed and optimized a genetic method based on Real Time PCR for detecting dreissenid presence in water samples.

6. REFERENCES





Abel-Abellán I., 2010. Seguimiento de la estrategia Nacional Contra el Mejillón Cebra. Available at: http://biodiversia.es/media/Biblioteca/Seguimiento_de_la_Estrategia_Nacional_contra_el_mejillon_cebra.pdf. (Accessed: 30th January 2016).

ACA (Catalan Water Agency), 2015. El musclo zebrat. Estat de la presència a Catalunya – 2014. Available at: http://aca-web.gencat.cat/aca/. (Accessed: 30th March 2015).

Albrecht C., Schulthei R., Kevrekids T., Streit B. and Wilke T., 2007. Invaders or endemics? Molecular phylogenetics, biogeography and systematics of *Dreissena* in the Balkans. *Freshwater Biology* **52**: 1525-1536.

Allendorf F. W. and Luikart G., 2009. Genetic variation in natural populations: DNA. In: Conservation and the genetics of populations. Allendorf F. W. and Luikart G. (eds.). Willey-Blackwell publishing Ltd. Malden, Massachusetts. pp: 63-90.

Allendorf F. W. and Lundquist L. L., 2003. Introduction: population biology, evolution, and control of invasive species. *Conservation Biology* **17**: 24-30.

Altmann A., Weber P., Bader D., Preuss M., Binder E. B. and Müller-Myhsok B., 2012. A beginners guide to SNP calling from high-throughput DNA-sequencing data. *Human Genetics* **10**:1541-1554.

Ansorge W. J., 2009. Next-generation DNA sequencing techniques. *New Biotechnology* **25**: 195-203.

Araujo R., 2004. Los bivalvos dulceacuícolas de La Rioja. *Zubía* **22**: 29-39.

Araujo R., Moreno D. and Ramos M. A., 1993. The asiatic clam *Corbicula fluminea* (Müller, 1774) (Bivalvia: Corbiculidae) in Europe. *American Malacological Bulletin* **10**: 39-49.

Arif I. A. and Khan H. A., 2009. Molecular markers for biodiversity analysis of wildlife animals: a brief review. *Animal Biodiversity and Conservation* **32**: 10-17.

Arif I. A., Khan H. A., Bahkali A. H., Homaidan A. A. A., Farhan A. H., Sadoon M. A. and Shobrak M., 2011. DNA marker technology for wildlife conservation. *Saudi Journal of Biological Sciences* **18**: 219-225.

Astanei I., Gosling E., Wilson J. and Powell E., 2005. Genetic variability and phylogeography of the invasive zebra mussel, *Dreissena polymorpha* (Pallas). *Molecular Ecology* **14**: 1655-1666.

Avise J. C., 1994. Molecular tools. In: Molecular markers: Natural history and evolution. Avise J. C. (ed.). Chapman & Hall, New York. pp: 44-91.

Avise J. C., 2000. Phylogeography: The history and formation of species. Avise J. C. (ed.). Harvard University Press, Cambridge, M. A.

Baldwin B. S., Black M., Sanjur O., Gustafson R., Lutz R. A. and Vrijenhoek R. C., 1996. A diagnostic molecular marker for zebra mussels (*Dreissena polymorpha*) and potentially co-occurring bivales: mitochondrial COI. *Molecular Marine Biology and Biotechnology* **5**: 9-14.

Bernatchez L. and Duchesne P., 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? *Canadian Journal of Fisheries and Aquatic Sciences* **57**: 1-12.

Blackburn T. M., Lockwood J. L. and Cassey P., 2015. The influence of numbers on invasion success. *Molecular Ecology* **24**: 1942-1953.

Bij de Vaate A., Jazdzewski K., Ketelaars H. A. M., Gollasch S. and Van der Velde G., 2002. Geographical patterns in range extension of Ponto-Caspian macroinvertebrate species in Europe. *Canadian Journal of Fisheries and Aquatic Sciences* **59**: 1159–1174.

Bij de Vaate A., Van der Velde G., Leuven R. S. E. W. and Heiler K. C. M., 2013. Spread of the Quagga Mussel (*Dreissena rostriformis bugensis*) in Western Europe. In: Quagga and zebra mussels: biology, impacts, and control. Nalepa, T. F. and Schloesser D. (eds.). CRC Press, Boca Ratón. pp: 83-92.

Botstein D., White R. L., Skolnick M. and Davis R. W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**: 314–331.

Brown J. E. and Stepien C. A., 2010. Population genetic history of the dreissenid mussel invasions: expansion patterns across North America. *Biological Invasions* **12**: 3687-3710.

CHE (Ebro Hydrographic Confederation), 2015. Monitorización de la presencia larvaria de *Dreissena polymorpha* en la Cuenca Hidrográfica del Ebro, campaña 2014. Avaliable at: http://www.chebro.es/. (Accessed: 30th June 2015).

CHG (Guadalquivir Hydrographic Confederation), 2014. La Almeja Asiática. Evitar su expansión está en nuestras manos. Available at: http://www.chguadalquivir.es/. (Accessed: 15th March 2015).

CHG (Guadalquivir Hydrographic Confederation), 2015. Mejillón cebra. Evitar su expansión está en nuestras manos. Available at: http://www.chguadalquivir.es/. (Accessed: 30th January 2016).

CHJ (Júcar Hydrographic Confederation), 2015. Seguimiento de la evolución de la población de mejillón cebra (*Dreissena polymorpha*) en la cuenca del río Júcar. Informe anual resultados 2015. Available at: http://www.chj.es/. (Accessed: 30th January 2016).

CHS (Segura Hydrographic Confederation), 2014. La CHS confirma la presencia de ejemplares de almeja asiática en la vega media del Segura. Available at: www.chsegura.es/. (Accessed: 30th March 2015).

Chistiakov D. A., Hellemans B. and Volckaret F. A. M., 2006. Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture* **255**: 1–29.

Clark M. S., Thorne M. A., Vieira F. A., Cardoso J. C., Power D. M. and Peck L. S., 2010. Insights into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing. *BMC Genomics* **11**: 362.

Colautti R. I. and Lau J. A., 2015. Contemporany evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology* **24**: 1999-2017.

Cristescu M. E., 2015. Genetic reconstructions of invasion history. *Molecular Ecology* **24**: 2212–2225.

Cushman S. A., 2015. Pushing the envelope in genetic analysis of species invasion. *Molecular Ecology* **24**: 259-262.

Davey J. W., Hohenlohe P. A., Etter P. D., Boone J. Q., Catchen J. M. and Blaxter M. L., 2011. Genome-wide genetic marker discovery and genotyping using Next-Generation Sequencing. *Nature Reviews. Genetics* **12**: 499-510.

De Olivera A., Holyoak G. A. and Holyoak D. T., 2010. Additional records of allien freshwater mollusca in Portugal [Materiais para o estudio da malacofauna Não-Marina de Portugal]. *Noticiario de la Sociedad Española de Malacologia* **54**: 41-45.

Defaveri J., Viitaniemi H., Leder E. and Merilä J., 2013. Characterizing genic and nongenic molecular markers: Comparison of microsatellites and SNPs. *Molecular Ecology Resources* **13**: 377-392.

Descy J. P., Everbecq E., Gosselain V., Viroux L. and Smitz S., 2003. Modelling the impact of benthic filter-feeders on the composition and biomass of river plankton. *Freshwater Biology* **48**: 404–417.

DeWoody J. A. and Avise J. C., 2000. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology* **56**: 461–473.

Dlugosch K. M. and Parker I. M., 2008. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* **17**: 431-449.

Drake J. M., 2006. Heterosis, the catapult effect and establishment success of a colonizing bird. *Biology letters* **2**: 304–307

Droege M. and Hill B., 2008. The Genome Sequencer FLXTM System—Longer reads, more applications, straight forward bioinformatics and more complete data sets. *Journal of Biotechnology* **136**: 3–10.

Durán C. and Anadón A., 2008. The zebra mussel invasion in Spain and navigation rules. *Aquatic Invasions* **3**: 315-324.

Durán C., Lanao M., Anadón A. and Touyá V., 2010. Management strategies for the zebra mussel invasion in the Ebro River basin. *Aquatic invasions* **5**: 309-316.

Durán C., Lanao M., Pérez y Pérez L., Chica C., Anadón A. and Touya V., 2012. Estimación de los costes de la invasión del mejillón cebra en la cuenca del Ebro (período 2005-2009). *Limnetica* **31**: 213-230.

Egan S. P., Grey E., Olds B., Feder J. L., Ruggiero S. T., Tanner C. E. and Lodge D. M., 2015. Rapid molecular detection of invasive species in ballast and harbor water by integrating environmental DNA and light transmission spectroscopy. *Environmental Science and Technology* **49**: 4113−4121.

Elshire R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K., Buckler E. S. and Mitchel S. E., 2011. A robust, simple genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS ONE* **5**: e19379.

Everett M. V., Grau E. D. and Seeb J. E., 2011. Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. *Molecular Ecology Resources* **11**: 93–108.

Excoffier L. and Heckel G., 2006. Computer programs for population genetics data analysis: a survival guide. *Nature Reviews. Genetics* **7**: 745-758.

Frischer M. E., Hansen A. S., Wyllie J. A., Wimbush J., Murray J. and Nierzwicki-Bauer S. A., 2002. Specific amplification of the 18S rRNA gene as a method to detect zebra mussel (*Dreissena polymorpha*) larvae in plankton samples. *Hydrobiologia* **487**: 33-44.

Gallardo-Escarate C., Valenzuela-Muñoz V., Nuñez-Acuña G. and Haye P., 2015. SNP discovery and gene annotation in the surf clam *Mesodesma donacium*. *Aquaculture Research* **46**: 1175–1187.

Galtier N., Nabholz B., Glémin S. and Hurst G. D., 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology* **22**: 4541-4550.

Galtin M. R., Shoup D. E. and Long J. M., 2013. Invasive zebra mussels (*Dreissena polymorpha*) and Asian clams (*Corbicula fluminea*) survive gut passage of migratory fish species: implications for dispersal. *Biological Invasions* **15**: 1195-1200.

Gardner M. G., Fitch A. J., Bertozzi T. and Lowe A. J., 2011. Rise of the machines—recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources* **11**: 1093–1101.

Gelembuik G. W., May G. E. and Lee C. E., 2006. Phylogeography and systematics of zebra mussels and related species. *Molecular Ecology* **15**: 1033–1050.

Guichoux E., Lagache L., Wagner S., Chaumeil P., Leger P., Lepais O., Lepoittevin C., Malausa T., Revardel E., Salin F. and Petit R. J., 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**: 591–611.

Hakenkamp C. C., Ribblett S. G., Palmer M. A., Swan C. M., Reid J. W. and Goodison M. R., 2001. The impact of an introduced bivalve (*Corbicula fluminea*) on the benthos of a sandy stream. *Freshwater Biology* **46**: 491-501.

Handley L. L. J., Estoup A., Evans D. M., Thomas C. E., Lombaert E., Facon B., Aebi A. and Roy H. E., 2011. Ecological genetics of invasive alien species. *Biological Control* **56**: 409-428.

Havel J. E., Kovalenko K. E., Thomaz S. M., Amalfitano S. and Kats L. B., 2015. Aquatic invasive species: challenges for the future. *Hydrobiologia* **750**: 147–170.

Hedtke S. M., Stanger-Hall K., Baker R. J. and Hillis D. M., 2008. All-male asexuality: origin and maintenance of androgenesis in the Asian clam *Corbicula*. *Evolution* **62**: 1119-1136.

Helyar S. J., Hemmer-Hansen J., Bekkevold D., Taylor M. I., Ogden R., Limborg M. T., Cariani A., Maes G. E., Diopere E., Carvalho G. R. and Nielsen E. E., 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**: 123-136.

Hoelzel R. A., 1998. Molecular genetic analysis of populations, a practical approach. Hoelzel R. A. (ed.). B. D. Hames, University of Leeds, Leeds.

Howard J. K. and Cuffey K., 2006. The functional role of native freshwater mussels in the fluvial benthic environment. *Freshwater Biology* **51**: 460–474.

Huang W. and Marth G., 2011. EagleView: A genome assembly viewer for Next-Generation Sequencing technologies. *Genome Research* **18**: 1538-1543.

Inoue K., Lang B. K. and Berg D. J., 2013. Development and characterization of 20 polymorphic microsatellite markers for the Texas hornshell, *Popenaias popeii* (Bivalvia: Unionidae), through next-generation sequencing. *Conservation Genetics Resources* **5**:195–198.

Johnson L. E. and Carlton J. T., 1996. Post-establishment spread in large-scale invasions: dispersal mechanisms of the zebra mussel *Dreissena polymorpha*. *Ecology* **77**: 1686–1690.

Kalinowski S. T., 2002. How many alleles per locus should be used to estimate genetic distances? *Heredity* **88:** 62–65.

Kang J. H., Kim Y. K., Park J. Y., Noh E. S., Jeong J. E., Lee Y. S. and Choi T. J., 2013. Development of microsatellite markers for a hard-shelled mussel, *Mytilus coruscus*, and cross-species transfer. *Genetics and Molecular Research* **12**: 4009-4017.

Karatayev A. Y., Burlakova L. E. and Padilla D. K., 2005. Contrasting distribution and impacts of two freshwater exotic suspension feeders, *Dreissena polymorpha* and *Corbicula fluminea*. In: The comparative roles of suspension-feeders in ecosystems. Dame R. F. and Olenin S. (eds.). Springer, The Netherlands. pp: 239-262.

Karatayev A. Y., Burlakova L. E. and Padilla D. K., 2015. Zebra versus quagga mussels: a review of their spread, population dynamics, and ecosystem impacts. *Hydrobiologia* **746**: 97–112.

Karatayev A. Y., Padilla D. K., Minchin D., Boltovskoy D. and Burlakova L. E., 2007. Changes in global economies and trade: the potential spread of exotic freshwater bivalves. *Biological Invasions* **9**: 161–180.

Kelkar Y. D., Strubczewski N., Hile S. E., Chiaromonte F., Ecket K. A. and Makova K. D., 2010. What is a microsatellite: A computational and experimental definition based upon repeat mutational behavior at A/T and GT/AC repeats. *Genome Biology and Evolution* **2**: 620-635.

Kim E. M., An H. S., Kang J. H., An C. M., Dong C. M., Hong Y. K. and Park J. Y., 2014. New polymorphic microsatellite markers for the Korean manila clam (*Ruditapes philippinarum*) and their application to wild populations. *Genetics and Molecular Research* **13**: 8163-8173.

Kwok P. Y. and Chen X., 2003. Detection of Single Nucleotide Polymorphisms. *Current Issues in Molecular Biology* **5**: 43-60.

Lau J. A. and Terhorst C. P., 2015. Causes and consequences of failed adaptation to biological invasions: the role of ecological constraints. *Molecular Ecology* **24**: 1987-1998.

Lee T., Siripattrawan S., Ituarte C. F. and Ó Foighil D., 2005. Invasion of the clonal clams: Corbicula lineages in the New World. *American Malacology Bulletin* **20**: 113-122.

Lewis K. M., Feder J. L. and Lamberti G. A., 2000. Population genetics of the zebra mussel *Dreissena polymorpha* (Pallas): local allozyme differentiation within midwestern lakes and streams. *Canadian Journal of Fisheries and Aquatic Sciences* **57**: 637–643.

Liew M., Pryor R., Palais R., Meadows C., Erali M., Lyon E. and Wittwer C. T., 2004. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clinical Chemistry* **50**: 1156-1164.

Liu N., Chen L., Wang S., Oh C. and Zhao H., 2005. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics* **6**: S26.

Liu Z. J. and Cordes J. F., 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**: 1 –37.

Lois S., 2010. New records of *Corbicula fluminea* (Müller, 1774) in Galicia (Northwest of the Iberian Peninsula): Mero, Sil and Deva rivers. *Aquatic Invasions* **5**: S17–S20.

Lopes-Lima M., Teixera A., Froufe E., Lopes A., Varandas S. and Sousa R., 2014. Biology and conservation of freshwater bivalves: past, present and future perspectives. *Hydrobiologia* **735**: 1–13.

López M. A. and Altaba C. R., 1997. Presencia de *Corbicula fluminea* (Müller, 1774) (Bivalvia, Corbiculidae) al Delta de L´Ebre. *Bulleti del Parc Natural Delta de L´Ebre* **10**: 20-22.

Lucy F. E., 2006. Early life stages of *Dreissena polymorpha* (zebra mussel): the importance of long-term datasets in invasion ecology. *Aquatic Invasions* **1**: 171-182.

Lucy F. E., Karatayev A. Y. and Burlakova L., 2012. Predictions for the spread, population density, and impacts of *Corbicula fluminea* in Ireland. *Aquatic Invasions* **7**: 465-474.

Malmqvist B., 2002. Aquatic invertebrates in riverine landscapes. *Freshwater Biology* **47**: 679-694.

Mardsen J. E., Spidle A. and May B., 1995. Genetic similarity among zebra mussel populations within North America and Europe. *Canadian Journal of Fisheries and Aquatic Sciences* **52**: 836-847.

Mardis E. R., 2008a. Next-Generation DNA Sequencing methods. *Annual Review of Genomics and Human Genetics* **9**: 387–402.

Mardis E. R., 2008b. The impact of Next-Generation Sequencing technology on genetics. *Trends in Genetics* **24**: 133-141.

Margulies M., Egholm M., Altman W. E., Attiya S., Bader J. S., Bemben L. A., Berka J., Braverman M. S., Chen Y. J., Chen Z., Dewell S. B., Du L., Fierro J. M., Gomes X. V., Godwin B. C., He W., Helgesen S., Ho C. H., Irzyk G. P., Jando S. C., Alenquer M. L. I., Jarvie T. P., Jirarge K. B., Kim J. B., Knight J. R., Lanza J. R., Leamon J. H., Lefkowitz S. M., Lei M., Li J., Lohman K. L., Lu H., Makhijani V. B., McDade K. E., McKenna M. P., Myers E. W., Nickerson E., Nobile J. R., Plant R., Puc B. P., Ronan M. T., Roth G. T., Sarkis G. J., Simons J. F., Simpson J. W., Srinivasan M., Tartaro K. R., Tomasz A., Vogt K. A., Volkmer G. A., Wang S. H., Wang Y., Weiner M. P., Yu P., Begley R. F. and Rothberg J. M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.

May G. E., Gelembuik G. W., Panov V. E., Orlova M. I. and Lee C. E., 2006. Molecular ecology of zebra mussel invasions. *Molecular Ecology* **15**: 1021-1031.

McGlauflin M. T., Smith M. J., Wang J. T., Young S. F., Chen Y., Lee Y. C., Pascal C., Seeb L. W., Stevens J. and Seeb J. E., 2010. High-Resolution Melting Analysis for the discovery of novel Single-Nucleotide Polymorphisms in rainbow and cutthroat trout for species iIdentification. *Transactions of the American Fisheries Society* **39**: 676-684.

McInerney C. E., Allcock A. L., Johnson M. P., Bailie D. A. and Prodohl P. A., 2011. Comparative genomic analysis reveals species-dependent complexities that explain difficulties with microsatellite marker development in mollusks. *Heredity* **106**: 78–87.

McMahon R. F., 1982. The occurrence and spread of the introduced asiatic freshwater clam, *Corbicula fluminea* (Müller), in North-America – 1924-1982. *Nautilus* **96**: 134–141.

McMahon R. F., 2002. Evolutionary and physiological adaptations of aquatic invasive animals: r selection versus resistance. *Canadian Journal of Fisheries and Aquatic Sciences* **59**: 1235–1244.

Metzker M. L., 2010. Sequencing technologies —the next generation. *Nature Reviews. Genetics* **11**: 31-46.

MMARM (Ministerio Medio Ambiente, Rural y Marino), 2007. Estrategia Nacional para el Control del Mejillón Cebra (*Dreissena polymorpha*) en España. Available at: http://www.chebro.es/contenido.visualizar.do?idContenido=32951&idMenu=4240. (Accessed: 20th December 2013).

Morais P., Teodósio J., Reis J., Chícharo, M. A. and Chícharo L., 2009. The Asian clam *Corbicula fluminea* (Müller, 1774) in the Guadiana River Basin (southwestern Iberian Peninsula): Setting the record straight. *Aquatic Invasions* **4**: 681-684.

Morin P. A., Luikart G. and Wayne R. K., 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19**: 208-216.

Morin P. A., Martien K. K. and Taylo B. L., 2009. Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources* **9**: 66-73.

Mouthon J., 1981. Sur la présence en France et au Portugal de *Corbicula* (Bivalvia, Corbiculidae) originaire d'Asie. *Basteria* **45**: 109-116.

Müller J. C., Hidde D. and Seitz A., 2002. Canal construction destroys the barrier between major European invasion lineages of the zebra mussel. *Proceedings of the Royal Society* **269**: 1139–1142.

Nagel K. O., 1989. Ein weiterer Fundort von *Corbicula fluminalis* (MÜLLER 1774) (Mollusca: Bivalvia) in Portugal. *Mitteilungen der Deutschen Malakozoologischen Gesellschaft* **44/45**: 17.

Navarro A., Sánchez-Fontenla J., Cordero D., Faria M., Pena J. B., Saavedra C., Blazquez M., Ruiz O., Urena R., Torreblanca A., Barata C. and Pina B., 2013. Genetic and phenoptypic differentiation of zebra mussel populations colonizing Spanish river basins. *Ecotoxicology* **22**: 915-928.

Navarro J. and Ugalde M., 2008. *Corbicula fluminea* (Müller, 1774) (Bivalvia, Veneroidea) También localizada en Tierra de Campos (Palencia). *Noticiario de la Sociedad Española de Malacologia* **49**: 36-39.

Nei M., Tajima F. and Tateno Y., 1983. Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* **19**: 153–170.

Nielsen R., Paul J. S., Albrechtsen A. and Song Y. S., 2011. Genotype and SNP calling from Next-Generation Sequencing data. *Nature Reviews. Genetics* **12**: 443-451.

Niu D., Wang L., Sun F., Liu Z. and Li J., 2013. Development of molecular resources for an intertidal clam, *Sinonovacula constricta*, using 454 transcriptome sequencing. *PLoS ONE* **8**: e67456.

Nowrousian M., 2010. Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. *Eukaryotic Cell* **9**: 1300-10.

Oscoz J., Agorreta A., Durán C. and Larraz M. L., 2006. Aportaciones al conocimiento de algunos bivalvos dulceacuícolas en la cuenca del Ebro. *Naturaleza Aragonesa* **16**: 27-36.

Oscoz J., Larraz M. L., Tomás P., Pardos M. and Durán C., 2008. Nuevas citas de almeja asiática (*Corbicula fluminea* (Müller, 1774) (Mollusca, Bivalvia) en ríos de Navarra. *Noticiario de la Sociedad Española de Malacologia* **50**: 42–43.

Oscoz J., Tomás P. and Durán C., 2010. Review and new records of non-indigenous freshwater invertebrates in the Ebro River basin (Northeast Spain). *Aquatic Invasions* **5**: 263-284.

Park J. K. and Kim W., 2003. Two *Corbicula* (Corbiculidae: Bivalvia) mitochondrial lineages are widely distributed in Asian freshwater environment. *Molecular Phylogenetics and Evolution* **29**: 529-539.

Pfenninger M., Reinhardt F. and Streit B., 2002. Evidence for cryptic hybridization between different evolutionary lineages of the invasive clam genus *Corbicula* (Veneroida, Bivalvia). *Journal of Evolutionary Biology* **15**: 818-829.

Pérez-Bote J. L. and Fernández J., 2008. First record of the Asian clam *Corbicula fluminea* (Müller, 1774) in the Guadiana River Basin (southwestern Iberian Peninsula). *Aquatic Invasions* **3**: 87-90.

Pérez-Quintero J. C., 2008. Revision of the distribution of *Corbicula fluminea* (Müller 1744) in the Iberian Peninsula. *Aquatic Invasions* **3**: 355–358.

Pigneur L. M., Etoundi E., Aldridge D. C., Marescaux J., Yasuda N. and Van Doninck K., 2014a. Genetic uniformity and long-distance clonal dispersal in the invasive androgenetic *Corbicula* clams. *Molecular Ecology* **23**: 5102-5116.

Pigneur L. M., Falisse E., Roland K., Everbecq E., Deliège J. F., Smitz J. S., Van Doninck K. and Descy J. P., 2014b. Impact of invasive Asian clams, *Corbicula spp*., on a large river ecosystem. *Freshwater Biology* **59**: 573-583.

Pigneur L. M., Hedtke S., Etoundi E. and Van Doninck K., 2012. Androgenesis: a review through the study of the selfish shellfish *Corbicula spp*. *Heredity* **108**: 581–591.

Pigneur L. M., Marescaux J., Roland K., Etoundi E., Descy J. P. and Van Doninck K., 2011. Phylogeny and androgenesis in the invasive *Corbicula* clams (Bivalvia, Corbiculidae) in Western-Europe. *BMC Evolutionary Biology* **11**: 147.

Pimentel D., Zuniga D. R. and Morrison D., 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics* **52**: 273-288.

Rajagopal S., Pollux B. J. A., Peters J. L, Cremers G., Moon-van der Staay S. Y., Van Alen T., Eygensteyn J., Van Hoek A., Palau A., Bij de Vaate A. and Van der Velde G., 2009. Origin of Spanish invasion by the zebra mussel, *Dreissena polymorpha* (Pallas, 1771) revealed by amplified fragment length polymorphism (AFLP) fingerprinting. *Biological Invasions* **11**: 2147-2159.

Ram J. L., Karim A. S., Acharya P., Jagtap P., Purohit S. and Kashian D. R., 2011. Reproduction and genetic detection of veligers in changing *Dreissena* populations in the great Lakes. *Ecosphere* **2**: 1-16.

Ratan A., Zhang Y., Hayes V. M., Schuster S. C. and Miller W., 2010. Calling SNPs without a reference sequence. *BMC Bioinformatics* **11**: 130.

Reed H. G., Kent J. O. and Wittwer C. T., 2007. High-Resolution DNA melting analysis for simple and efficient molecular diagnostics. *Pharmacogenomics* **8**: 597-608.

Ricciardi A., 2003. Predicting the impacts of an introduced species from its invasion history: an empirical approach applied to zebra mussel invasions. *Freshwater Biology* **48**: 972-981.

Ricciardi A., Whoriskey F. G. and Rasmussen J. B., 1995. Predicting the intensity and impact of *Dreissena* infestation on native unionid bivalves from *Dreissena* field density. *Canadian Journal of Fisheries and Aquatic Sciences* **52**: 1449-1461.

Rios-Jiménez F., 2010. Nuevas citas de dispersión de la almeja asiatica *Corbicula fluminea* (Müller 1774) en Andalucia y algunos datos de interés cronológico y bibliográfico. *Noticiario de la Sociedad Española de Malacologia* **53**: 42-43.

Roman J. and Darling J. A., 2007. Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology and Evolution* **22**: 454-464.

Romiguier J., Gayral P., Ballenghien M., Bernard A., Cahais V., Chenuil A., Chiari Y., Dernat R., Duret L., Faivre N., Loire E., Lourenco J. M., Nabholz B., Roux C., Tsagkogeorga G., Weber A. A. T., Weinert L. A., Belkhir K., Bierne N., Glémin S. and Galtier N., 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**: 261-263.

Rothberg J. M., Hinz W., Rearick T. M., Schutz J., Mileski W., Davey M., Leamon J. H., Johnson K., Milgrew M. J., Edwards M., Hoon J., Simons J. F., Marran D., Myers J. W., Davidson J. F., Branting A., Nobile J. R., Puc B. P., Light D., Clark T. A., Huber M., Branciforte J. T., Stoner I. B., Cawley S. E., Lyons M., Fu Y., Homer N., Sedova M., Miao X., Reed B., Sabina J., Feierstein E., Schorn M., Alanjary M., Dimalanta E., Dressman D., Kasinskas R., Sokolsky T., Fidanza J. A., Namsaraev E., McKernan K. J., Williams A., Roth G. T. and Bustillo J., 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**: 348-352.

Roux J. L. and Wieczorek A. M., 2008. Molecular systematics and population genetics of biological invasions: towards a better understanding of invasive species management. *Annals of Applied Biology* **154**: 1-17.

Rueda J., Gómez E., Benavent J. M., Collado F., Peña C., Rabasa J. M., Sanz V. and Sebastián A., 2012. First record of invasive exotic bivalve *Corbicula fluminea* (Müller, 1774) (Bivalvia, Corbiculidae) in the Júcar River basin (East Iberian Peninsula). *Boletín de la Real Sociedad Española de Historia Natural, Actas* **109**: 21-22.

Ruíz-Altaba C., Jiménez P. J. and López M. A., 2001. El temido mejillón cebra empieza a invadir los ríos españoles desde el curso bajo del río Ebro. *Quercus* **188**: 50-51.

Saavedra C. and Bachere E., 2006. Bivalve genomics. *Aquaculture* **256**: 1-14.

Sakai A. K., Allendorf F. W., Holt J. S., Lodge D. M., Molofsky J., With K. A., Baughman S., Cabin R. J., Cohen J. E., Ellstrand N. C., McCauley D. E., O'Neil P., Parker I. M., Thompson J. N. and Weller S. G., 2001. The population Biology of invasive species. *Annual Review of Ecology, Evolution, and Systematics* **32**: 305-332.

Seeb J. E., Pascal C. E., Grau E. D., Seeb L. W., Templin W. D., Harkins T. and Roberts S. B., 2011. Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. *Molecular Ecology Resources* **11**: 335–348.

Sell J., Mioduchowska M., Kaczmarczyk A. and Szymańczak R., 2013. Identification and characterization of the first microsatellite loci for the thick-shelled river mussel *Unio crassus* (Bivalvia: Unionidae). *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology* **319**: 113-116.

Shanin A., Van Gurp T., Peters S. A., Visser R. G. F., Van Tuyl J. M. and Arens P., 2012. SNP markers retrieval for a non-model species: a practical approach. *BMC Research Notes* **5**: 79.

Shendure J. and Ji H., 2008. Next-generation DNA sequencing. *Nature Biotechnology* **26**: 1135-1145.

Smith C. T., Elfstrom C. M., Seeb L. W. and Seeb J. E., 2005. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. *Molecular Ecology* **14**: 4193–4203.

Smith B. L., Lu C. P. and Alvarado Bremer J. R., 2012. Methodological streamlining of SNP discovery and genotyping via high-resolution melting analysis (HRMA) in non-model species. *Marine Genomics* **9**: 39-49.

Sousa R., Novais A., Costa R. and Strayer D. L., 2014. Invasive bivalves in fresh waters: impacts from individuals to ecosystems and possible control strategies. *Hydrobiologia* **735**: 233–251.

Sousa R., Antunes C. and Guilhermino L., 2007. Species composition and monthly variation of the molluscan fauna in the freshwater subtidal area of the River Minho estuary. *Estuarine Coastal and Shelf Science* **75**: 90-100.

Sousa R., Antunes C. and Guilhermino L., 2008. Ecology of the invasive Asian clam *Corbicula fluminea* (Müller, 1774) in aquatic ecosystems: an overview. *Annales de Limnologie -International Journal of Limnology* **44**: 85-94.

Spooner D. and Vaughn C. C., 2006. Context-dependent effects of freshwater mussels on stream benthic communities. *Freshwater Biology* **51**: 1016-1024.

Stapley J., Santure A. W. and Dennis S. R., 2015. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular Ecology* **24**: 2241-2252.

Stepien C. A., Brown J. E., Neilson M. E. and Tumeo M. A., 2005. Genetic diversity of invasive species in the Great Lakes versus their Eurasian source populations: insights for risk analysis. *Risk Analysis* **25**: 1043–1060.

Stepien C. A., Grigorovich I. A., Gray M. A., Sullivan T. J., Yerga-Woolwine S. and Kalayci G., 2014. Evolutionary, biogeographic, and population genetic relationships of dreissenid mussels, with revision of component taxa. In: Quagga and zebra mussels: biology, impacts, and control. Nalepa, T. F. and Schloesser D. (eds.). CRC Press, Boca Ratón. pp: 403-444.

Stepien C. A., Taylor C. D. and Dabrowska K. A., 2002. Genetic variability and phylogeographic patterns of a nonindigenous species invasion: a comparison of exotic versus native zebra and quagga mussel populations. *Journal of Evolutionary Biology* **15**: 314–328.

Stockwell C. A. and Ashley M. V., 2004. Diversity-Rapid adaptation and conservation. *Conservation Biology* **18**: 272-273.

Strayer D. L., Hattala K. A. and Kahnle A. W., 2004. Effects of an invasive bivalve (*Dreissena polymorpha*) on fish in the Hudson River estuary. *Canadian Journal of Fisheries and Aquatic Sciences* **61**: 924–941.

Strayer D. L. and Malcom H. M., 2007. Effects of zebra mussels (*Dreissena polymorpha*) on native bivalves: the beginning of the end or the end of the beginning? *Journal of the North American Benthological Society* **26**: 111–122.

Strayer D. L. and Smith L. C., 1993. Distribution of the zebra mussel (*Dreissena polymorpha*) in estuaries and brackish waters. In: Quagga and zebra mussels: biology, impacts, and control. Nalepa, T. F. and Schloesser D. (eds.). CRC Press, Boca Ratón. pp: 715-727.

Suarez A. and Tsutsui N. D., 2008. The evolutionary consequences of biological invasions. *Molecular Ecology* **17**: 351-360.

Sunnucks P., 2000. Efficient genetic markers for population biology. *Trends in Ecology and Evolution* **15**: 199-203.

Tarnowska K., Daguin-Thiebaut C., Pain-Devin S. and Viard F., 2013. Nuclear and mitochondrial genetic variability of an old invader, *Dreissena polymorpha* (Bivalvia), in French river basins. *Biological Invasions* **15**: 2547–2561.

Tautz D. and Schlötterer C. S., 1994. Simple sequences. *Current Opinion in Genetics and Development* **4**: 832–837.

Tjensvoll K., Hodneland K., Nilsen F. and Nylund A., 2005. Genetic characterization of the mitochondrial DNA from *Lepeophtheirus salmonis* (Crustacea; Copepoda). A new gene organization revealed. *Gene* **353**: 218-230.

Twyman R. M., 2005. Single Nucleotide Polymorphism (SNP) genotyping techniques—an overview. *Encyclopedia of Diagnostic Genomics and Proteomics* **2**: 1202-1207.

URA (Vasc Water Agency), 2013. Plan de Acción para el Control de la expansión del mejillón cebra en la CAPV 2013-2015. Cimera Estudios Aplicados, S. A. para la Agencia Vasca del Agua. Available at: www.uragentzia.euskadi.net. (Accessed: 30th January 2014).

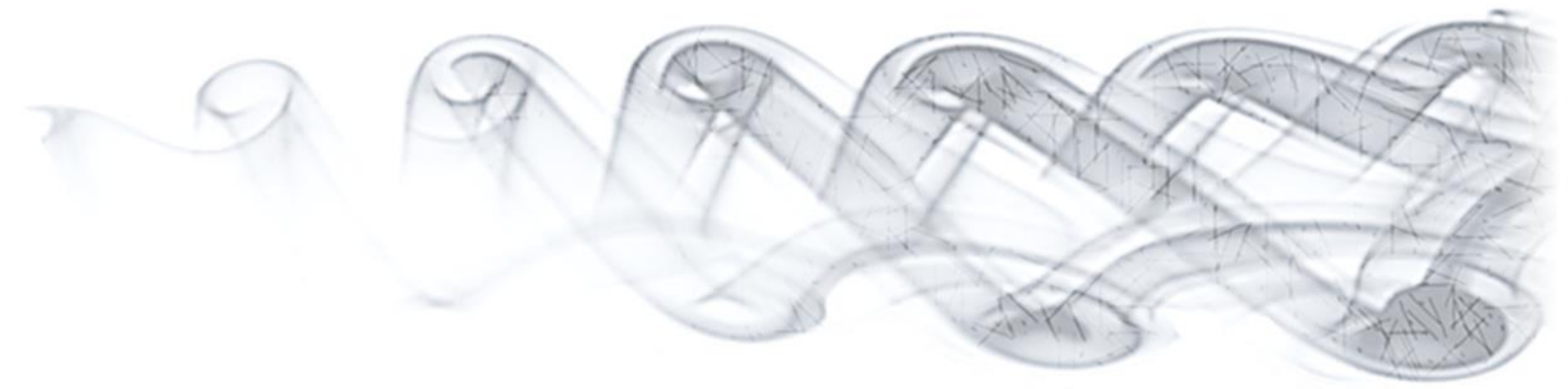
Vaughn C. C. and Hakenkamp C. C., 2001. The functional role of burrowing bivalves in freshwater ecosystems. *Freshwater Biology* **46:** 1431-1446.

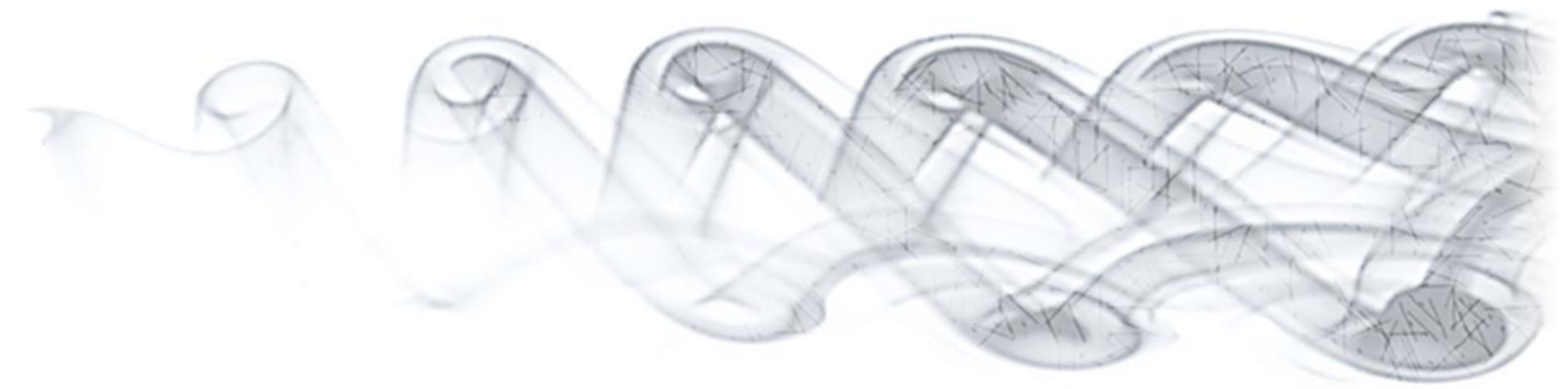
Zalapa J. E., Cuevas H., Zhu H., Steffan S., Senalik D., Zeldin E., McCown B., Harbut R. and Simon P., 2012. Using Next-generation Sequencing approaches to isolate Simple Sequence Repeats (SSR) loci in plant sciences. *American Journal of Botany* **2**: 193-208.

Zane L., Bargelloni L. and Patarnello T., 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**: 1–16.

Zhan A., Bao Z., Hu X., Lu W. and Hu J., 2008. Methods comparison for microsatellite marker development: different isolation methods, different yield efficiency. *Journal of Ocean University of China* **2**: 161–165.

7. ADDITIONAL INFORMATION





7.1

Additional Figures

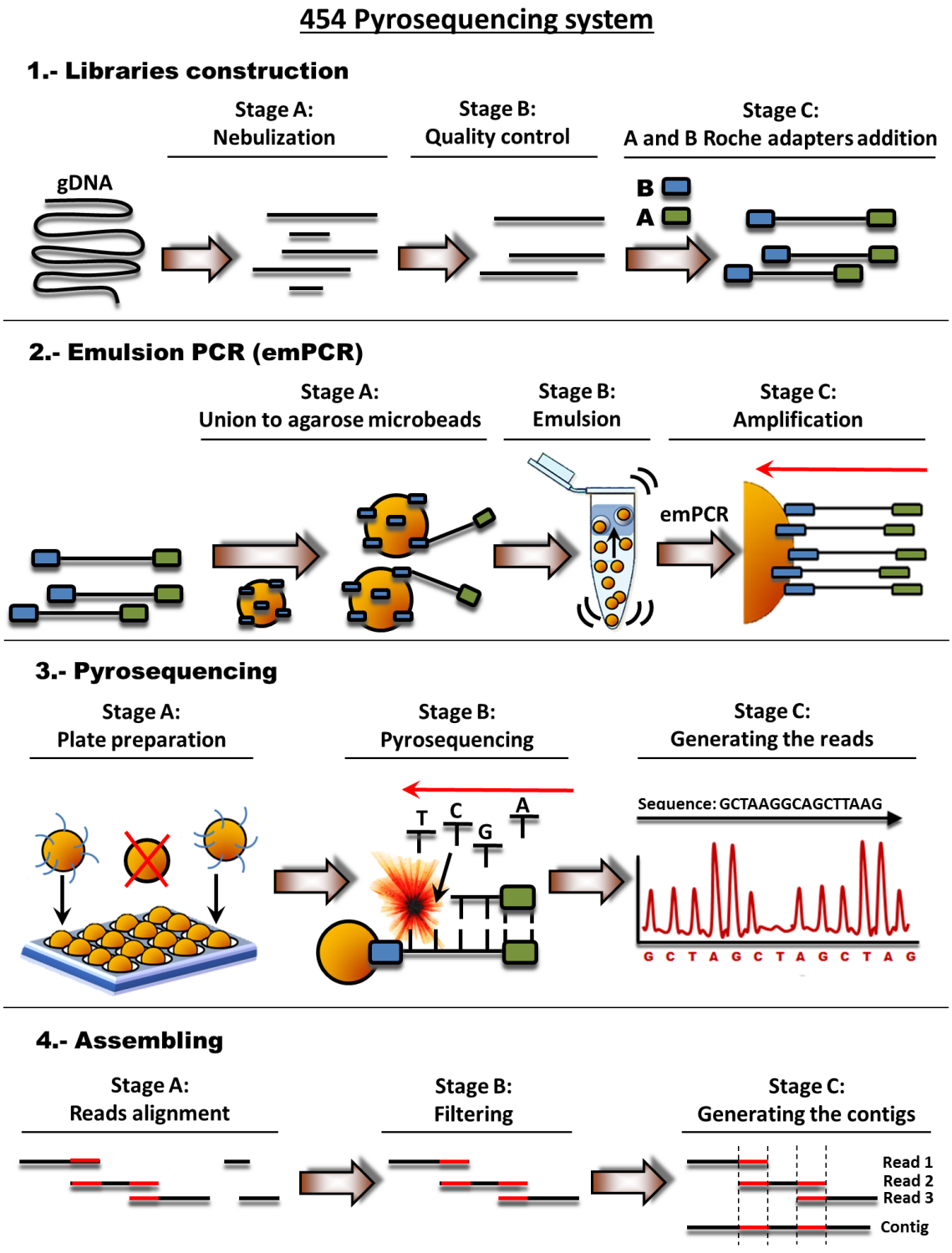


Figure 7.1.1. 454 Pyrosequencing framework. After libraries construction (1), emPCR (2) followed by Pyrosequencing (3) steps generate thousands of reads. These reads can be aligned and assembled (4) to obtain larger contigs.

Research workflow of the thesis

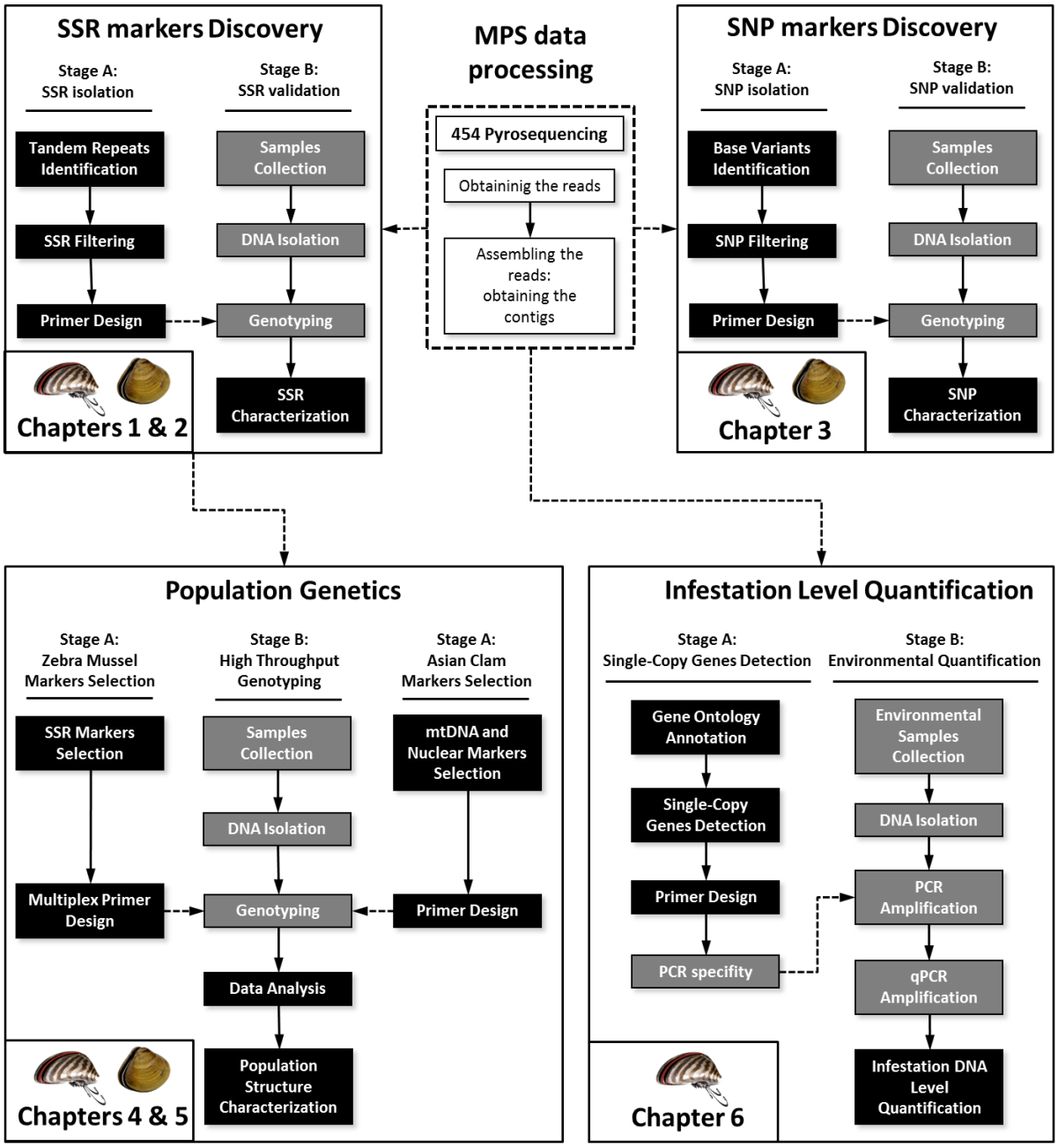


Figure 7.1.2. Workflow of the PhD thesis. 454 Pyrosequencing has been used for discovering new sets of SSRs (chapers 1 and 2) and SNPs (chapter 3) in the zebra mussel and the Asian clam. Subsequently, we used the best molecular markers to characterize the Iberian populations in both species (chapters 4 and 5). Finally, we also used the 454 Pyrosequencing for detecting single-copy genes in zebra mussel to optimize a protocol based on qPCR for the quantification of the infestation level of dreissenids in water samples (chapter 6).

7.2

Additional Tables

Table 7.2.1. Invasion history of the zebra mussel in the Iberian Peninsula.

Basin	River	Location	Latitude / Longitude	First record	Reference
Ebro River basin	Ebro River	1.- Sobron Reservoir	42º46'N 03º08'W	2006	Durán and Anadón 2008
	Ebro River	2.- Miranda Ebro	42º41'N 02º57'W	2006-2010	Oscoz et al. 2010; CHE 2015
	Ebro River	3.- Ircio River	42º40'N 02º54'W	2006-2010	Oscoz et al. 2010; CHE 2015
	Ebro River	4.- St. Vicente	42º34'N 02º46'W	2006-2010	Oscoz et al. 2010; CHE 2015
	Linares River	5.- Mendavia	42º27'N 02º12'W	2006-2010	Oscoz et al. 2010; CHE 2015
	Aragón Imperial Canal	6.- Zaragoza	41º37'N 00º51'W	2006-2010	Oscoz et al. 2010; CHE 2015
	Ebro River	7.- Ribarroja Reservoir	41º15'N 00º26'E	2001	Ruíz-Altaba et al. 2001
	Ebro River	8.- Ebro's Delta	40º43'N 00º50'E	2006-2010	Oscoz et al. 2010; CHE 2014
	Zadorra River	9.- Urivarri-Gamboa Reservoir	42º93'N 02º57'W	2012	CHE 2015
	Gallego River	10.- La Sotonera Reservoir	42º11'N 00º68'W	2013	CHE 2015
	Jalón River	11.- La Tranquera Resrvoir	41º24'N 01º78'W	2006	CHE 2015
Catalan basins	Llobregat River	12.- La Baells Reservoir	42º08'N 01º54'E	2011	ACA 2015
Jucar River basin	Mijares River	13.- Sitjar Reservoir	40º01'N 00º14'W	2005	Navarro et al. 2013
	Jucar River	14.- Forata Reservoir	39º34'N 00º87'W	2006	CHJ,2015
Guadalquivir River basin	Cacín River	15.- Bermejales Reservoir	37º28'N 04º33'W	2009	CHG 2015
	Genil River	16.- Iznájar Reservoir	36º99'N 03º89'W	2011	CHG 2015

Table 7.2.2. Invasion history of the Asian clam in the Iberian Peninsula.

Basin	River	Location	Latitude / Longitude	First record	Reference
Galician basins	Mero River	1.- Cambre	43º17'N 08º21'W	2009	Lois 2010
Minho-Sil Rivers basin	Minho River	2.- Minho's Delta	43º57'N 08º44'W	1989	Araujo *et al.* 1993
	Minho River	3.- Bridge Valença	42º01'N 08º38'W	2004	Sousa *et al*. 2007
	Sil River	4.- San Martiño	42º23'N 07º05'W	2009	Lois 2010
Duero River basin	Duero River	5.- Duero's Delta	41º08'N 08º35'W	1988	Nagel 1989
	Duero River	6.- Regua	41º09'N 07º47'W	1989	Araujo *et al.* 1993
	Castilla Canal	7.- Amusco, Palencia	42º19'N 04º49'W	2007	Navarro and Ugalde 2008
Tajo River basin	Tajo River	8.- Tajo's Delta	38º55'N 09º00'W	1980	Mouthon 1981
	Ocreza River	9.- Beira Baixa	39º84'N 07º65'W	2009	De Olivera *et al*. 2010
Guadiana River basin	Arroyo Grande	10.- Piedras River	37º18'N 07º12'W	1999	Pérez-Quintero 2008
	Rivera Grande	11.- Rivera Grande	37º27'N 07º27'W	1988	Morais *et al*. 2009
	Guadiana River	12.- Montijo Reservoir	38º55'N 06º25'W	2006	Pérez-Bote and Fernández 2008
Guadalquivir River basin	Guadalquivir River	13.- Sevilla	37º22'N 05º59'W	1993	Pérez-Quintero 2008
	Guadalquivir River	14.- Guadalcacin Reservoir	36º63'N 05º64'W	2009	Rios-Jiménez 2009
	Guadalquivir River	15.- Whole Guadalquivir range	w/d	2000-2015	CHG 2014
Ebro River basin	Arga River	16.- Peralta	42º34'N 01º79'W	2004	Araujo 2004
	Ebro River	17.- Pina de Ebro	41º37'N 00º51'W	2008	Oscoz *et al*. 2008
	Ebro River	18.- Ribarroja Reservoir	41º16'N 00º30'E	2002	Oscoz et al. 2006
	Ebro River	19.- Ebro's Delta	40º42'N 00º42'E	1997	López and Altaba 1997
Catalan basins	Ter River	20.- Torroella de Montgri	42º02'N 03º07'E	2012	ACA 2015
Jucar River basin	Jucar River	21.- Tous Reservoir	39º13'N 00º65'E	2011	Rueda *et al*. 2012
Segura River basin	Segura River	22.- Cieza	38º22'N 01º41'E	2013	CHS 2014

w/d: without data.

Table 7.2.3. Zebra and quagga mussel adult samples. Country, basin, latitude and longitude, number of collected individuals (Tissue), number of DNA isolations (DNA), and sampling date are detailed for each location.

Species	Country	Basin	Location	Latitude / Longitude	Tissue	DNA	Sampling date
Zebra mussel	Spain	Ebro River	Soborn Reservoir, Burgos	42°45'36" N / 03°07'37" W	160	60	21/04/2010
			Miranda de Ebro, Burgos	42°40'54" N / 02°56'57" W	52	52	27/10/2011
			Ircio, Burgos	42°39'11" N / 02°53'40" W	52	52	27/10/2011
			Linares River, Mendavia, Navarra	42°27'00" N / 02°12'00" W	12	12	27/10/2011
			St. Vicente de la Sonsierra, La Rioja	42°33'36" N / 02°45'43" W	52	52	27/10/2011
			Canal Imperial de Aragón, Zaragoza	41°36'43" N / 00°51'28" W	81	81	23/07/2012
			Ribarroja Reservoir, Tarragona	41°14'42" N / 00°25'46" E	181	81	15/04/2011
			Ribarroja Badia Tucana, Tarragona	41°14'33" N / 00°25'02" E	162	62	15/04/2011
			Ebro's Delta, Tarragona	40°42'36" N / 00°49'37" E	60	60	05/03/2010
		Llobregat River	La Baells Reservoir, Barcelona	42°07'39" N / 01°53'35" E	190	86	26/09/2011
		Mijares River	Sitjar Reservoir, Castellón	40°00'35" N / 00°13'54"W	52	52	12/07/2012
	France	Rhone River	Lac des Eaus Bleues, Lyon	45°4825" N / 04°56'10 E	60	60	13/03/2013
		Mosselle River	Metz (Saulcy), Lorena	49°07'07" N / 06°09'41" E	60	60	02/07/2013
			Sierk les Bains, Lorena	49°26'37" N / 06°21'07" E	60	60	02/07/2013
	Italy	Lugano Lake	Lugano Lake, Lombardy	45°59'39" N / 08°58'03" E	31	31	20/01/2013
		Trasimeno Lake	Trasimeno Lake, Umbria	43°08'43" N / 12°06'12" E	72	72	01/08/2013
	UK	Cadney Reservoir	Cadney Reservoir, Nortk Lincolnshire	53°31'42" N / 00°28'25" W	60	60	25/05/2013
		Manchester Ship Canal	Salfold Quays, Greater Manchester	53°28'18" N / 02°17'44" W	28	28	15/01/2013
		Bridgewater Canal	Bridgewater Canal, Greater Manchester	53°29'15" N / 02°26'11" W	29	29	15/01/2013
	Romania	Danube River	Galati, Moldavia	45°24'47" N / 28°02'43" W	48	48	28/10/2008
	USA	Red River	Texoma Lake, Denison, Texas	33°54'17" N / 96°42'11" W	175	75	24/09/2013
		Trinity River	Ray Roberts lake, Denton, Texas	33°24'10" N / 97°02'05" W	160	60	24/09/2013
Quagga mussel	The Neetherlands	Ijsselmeer Lake	Enkhuizen, North Holland	52°42'05" N / 05°18'51" W	57	57	18/06/2013

Table 7.2.4. Asian clam adult samples. Country, basin, latitude and longitude, number of collected individuals (Tissue), number of DNA isolations (DNA), and sampling date are detailed for each location.

Species	Country	Basin	Location	Latitude / Longitude	tissue	DNA	Sampling date
Asian clam	Spain	Minho River	Concello de Tui, Pontevedra	42°02'43" N / 08°38'31" W	50	50	21/09/1994
		Mero River	Cambre, A Coruña	43°17'23" N / 08°21'13" W	160	60	09/08/2012
		Guadiana River	Rivera Grande, Huelva	37°27'30" N / 07°27'06" W	148	60	15/05/2013
		Ebro River	Alagón, Zaragoza	41°46'52" N / 01°05'58 W	21	21	14/08/2012
			Canal Imperial de Aragón, Zaragoza	41°36'43" N / 00°51'28" W	178	78	23/07/2012
			Ribarroja Reservoir, Tarragona	41°14'42" N / 00°25'46" E	51	51	10/07/2012
			Playa de Xerta, Tarragona	40°53'59" N / 00°29'11" E	55	55	10/07/2012
			Ebro's Delta, Tarragona	40°42'36" N / 00°49'37" E	178	78	08/06/2011
		Ter River	Torroella de Montgri, Girona	42°02'07" N / 03°07'32" E	170	70	03/06/2012
	France	Rhone River	Saint Clair du Rhone, Lyon	45°25'02" N / 04°44'38" E	60	60	13/03/2013
	Romania	Danube River	Giorgeni, Ialomita Country	44°45'10" N / 27°52'27" W	24	24	15/08/2007
	USA	Sabine River	Deweyville, Texas	30°17'15" N / 93°42'45" W	180	80	17/11/2012
		Trinity River	Dayton, Texas	30°04'18" N / 94°49'32" W	22	22	14/10/2013
		Brazos River	Sealy, Texas	29°48'43" N / 96°06'13" W	129	72	14/11/2013
		Guadalupe River	Gonzales, Texas	29°29'48" N / 97°27'20" W	110	81	22/01/2013

7.3

Data accession availability

MPS outputs accession numbers

Both 454 GS FLX read sets were submitted to the NCBI *Sequence Read Archive* (SRA). The zebra mussel run corresponded to *BioSample* accession number **SAMN03257522** and *BioProject* accession **PRJNA269790**. SRA file containing all reads was under accession number **SRP051009** with the experiment accesion **SRX803533** and run **SRR1702044**. Parallely, the Asian clam run corresponded to *BioSample* accession number **SAMN04633495** and *BioProject* accession **PRJNA318156**. SRA file containing all reads was under accession number **SRP073154** with the experiment accession **SRX1692559** and run **SRR3360742**. De novo assembly was also submitted for both species in *DDBJ/EMBL/GenBank Whole Genome Shotgun* project when contigs size was longer than 500 nucleotides. Zebra mussel assembly was submitted under accession number **JWHF00000000**, whereas Asian clam assembly was under accession number **LWHL00000000**.

SSRs accession numbers

All SSRs developed in the present thesis were also submitted to GenBank. For the zebra mussel (**chapter 1**), 81 validated SSRs were submitted. Of them, 67 were monomorphic with accession numbers **KP274952–KP275018**, and 14 were polymorphic microsatellite sequences with accessions **JQ812984–JQ812997**. In case of the Asian clam (**chapter 2**), all 246 identified microsatellites were submitted. The nine polymorphic SSRs corresponded to accessions **KF730321-KF730329**, whereas the 49 monomorphic SSRs corresponded to accessions **KF744990-KF745038**. The remaining 188 identified SSRs were also submitted with accession numbers **KM589068-KM589255**.

SNPs accession numbers

Only polymorphic SNPs of the two species (**chapter 3**) were submitted to GenBank. In this sense, five SNPs in zebra mussel with accessions **KT220181-KT220185**, and four SNPs for the Asian clam with accessions **KT220186-KT220190**.

Nuclear and mitochondrial haplotypes described in the Asian clam

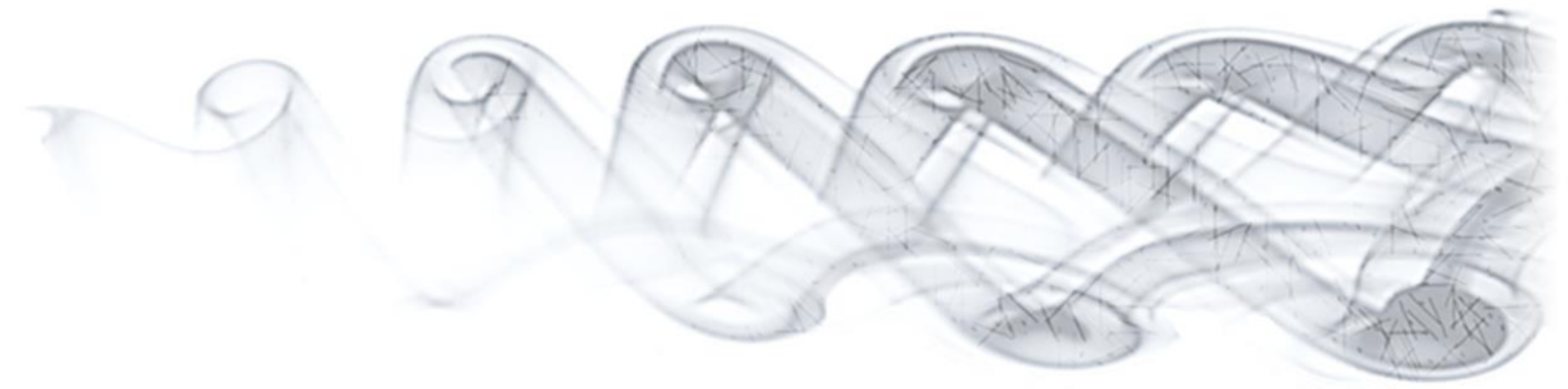
Sequences obtained in the Asian clam genetic characterization (**chapter 5**) were deposited in GenBank. Mitochondrial haplotypes corresponding to seven COI gene sequences

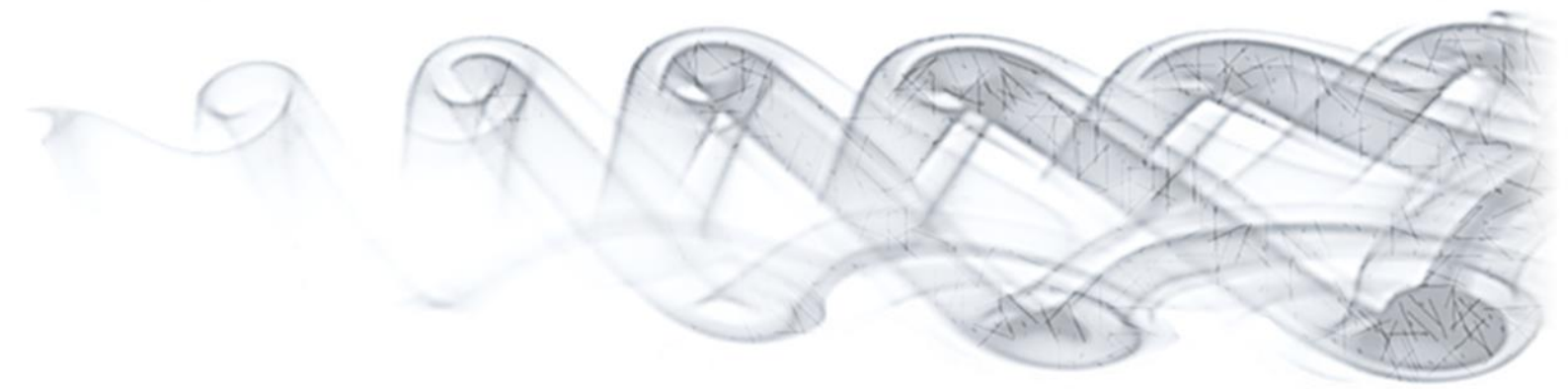
corresponded to accession numbers **KT373819-KT373825**. In addition, a total of ten nuclear haplotypes for the 28S gene corresponded to accesions **KT373826-KT373835**. Additionally, less frequent 28S haplotypes found during the cloning step (not included in the analyses) were also submitted with accession numbers **KT373836-KT373841**.

Single-copy genes described in the zebra mussel

Single-copy genes in the zebra mussel were also characterized from the MPS output (**chapter 6**). In this sense, we kept the accession numbers from the assembly submission, and four single-copy genes were characterized and used in subsequent genetic procedures: Histone 3 (Accession number **JWHF01000070**), Histone 2B (Accession **JWHF01000076**), Histone 1 (Accession **JWHF01000102**), and Methionyl-tRNA Synthetase (Accesion **JWHF01000913**).

8. SUPPORTING INFORMATION FROM CHAPTERS





8.1

Supporting information from Chapter 1:

Using Massive Parallel Sequencing for the development, validation, and application of population genetics markers in the invasive bivalve zebra mussel (*Dreissena polymorpha*)

Table S1. Samples analyzed in this study with their genetic variability statistics. *n*, number of samples; *A*, number of alleles average per locus); *Ar*, allelic richness; *Hs*, gene diversity.

Study number	Study site	Basin	Coordenates		Date of collection	*n*	*A*	*Ar*	*Hs*
			Latitude, N	Longitude, E					
1	Sobron Reservoir	Ebro River basin	42° 46’	-3° 07’	2010	8	48 (3.43)	2.14	0.53
2	Canal Imperial Aragón	Ebro River basin	41° 36’	-0° 51’	2011	8	50 (3.57)	2.33	0.61
3	Ribarroja Badia-Tucana Wharf	Ebro River basin	41° 14'	0° 25'	2011	8	46 (3.29)	2.16	0.54
4	Ribarroja Reservoir	Ebro River basin	41° 14’	0° 25’	2011	8	46 (3.29)	2.29	0.60
5	Delta of Ebro River	Ebro River basin	40° 42’	0° 49’	2010	8	50 (3.57)	2.31	0.60
6	La Baells Reservoir	Llobregat River basin	42° 07’	1° 53’	2012	8	43 (3.07)	2.06	0.50
	TOTAL/AVERGE					48	67 (4.79)	2.21	0.56

Table S2. Number (and percentage) of microsatellite types based on the size of the repeat motif.

% of each motif Repeat Size	SSR isolation and primer design		
	SSR identified (%)	Potentially amplifically loci (%)	Loci primers designed (%)
Dinucleotide	19 (6.60)	8 (7.77)	8 (8.60)
Trinucleotide	109 (37.85)	41 (39.81)	39 (41.94)
Tetranucleotide	114 (39.58)	45 (43.69)	43 (46.24)
Pentanucleotide	46 (15.97)	9 (8.74)	4 (4.30)
Hexanucleotide	-	-	-
TOTAL	288	103	93

Table S3. Description of the 93 validated microsatellites. Repeat motif, teoric size in base pairs (bp), forward (*F*) and reverse (*R*) (5'– 3') primer sequences and GenBank accession number.

Locus	Repeat motif	Teoric Size (bp)	Primers Forward (*F*) and Reverse (*R*) 5' → 3'	Validation	GenBank Accession
Dp1	$[ATA]_{21}$	262	*F:* GGATTTTTCTCCCGTGGAAT *R:* CGGTAGCGTTCTCTTCACAA	Polymorphic	JQ812984
Dp2	$[TGA]_{17}$	409	*F:* GCTACCGGAGCTCAACCTAA *R:* ACGTCGAACCCTGTCAAAAA	Polymorphic	JQ812985
Dp3	$[AT]_{12}$	105	*F:* TGTTAACTTGTCCATGCTATTCG *R:* ATCTCAGGGGTCACAACAGG	Validated	KP274952
Dp4	$[CTA]_{59}$ 2 $[CCA]_{25}$	373	*F:* CCCATATTTACTGCCAGTGC *R:* GCAGCCATTGTGTGAATACG	Not amplified	-
Dp5	$[TG]_{10}$	148	*F:* TTTCTTTTGGGAAGTCGACAC *R:* ATTAGTTGCCCGGTTTCCAT	Validated	KP274953
Dp6	$[GGCT]_{11}$	256	*F:* TAACCAGGTTTTCCGAAGGA *R:* GGCCATTTTTGACCTTTGAA	Validated	KP274954
Dp7	$[TAA]_{11}$ 4 $[ATT]_5$ 48 $[TAT]_{11}$ 71 $[ATA]_{14}$	417	*F:* GGAATACCGGGTGCTGTAGA *R:* GACGTGCGTCACAATAGGTG	Polymorphic	JQ812986
Dp8	[ATC]10 [TCG]7	303	*F:* TCTATGCTCGCCAAAACGTA *R:* CGGCGAAACAGTTACACAAC	Validated	KP274955
Dp9	$[CTA]_{10}$ 36 $[CTA]_{19}$	357	*F:* AATATGGGCTTCCATCCTGA *R:* TGGTTATAGTGGTAGTGGTAGTGG	Validated	KP274956
Dp10	$[TGA]_5$ 37 $[TGA]_{10}$ 36 $[TGA]_{14}$	415	*F:* TTTGGGATATTTTGGCTGCT *R:* TCCGATAAACATCGCTGTC	Validated	KP274957
Dp11	$[CAA]_{22}$ $[CAG]_{19}$ 79 $[CTA]_8$ 66 $[CTA]_{23}$ $[CTG]_5$	436	*F:* ATGACTTGTAGAAATTGCTTTGAA *R:* GAACACCTTACACTTACAAATAGTCAA	Validated	KP274958
Dp12	$[ATT]_{16}$	254	*F:* CGATATACAATTGCGCCAAA *R:* CAAAATAGGAATTTTGAGTTTGTGAA	Validated	KP274959
Dp13	$[CA]_{15}$	224	*F:* AACACAGGCGCATAACACAC *R:* CATAAGCGCACTTCCGTACA	Validated	KP274960
Dp14	$[ACA]_8$	139	*F:* AAATTTAAATATAAACACGCATTGTG *R:* AAAACAAAATCATACGCACTATTGA	Validated	KP274961
Dp15	$[ATG]_{56}$	300	*F:* TTTGTTTTTACGTGGCAGCA *R:* TGAAACAAGGAAATGAGGACCTA	Not amplified	-
Dp16	$[GAT]_{10}$	212	*F:* ACAATACTCCGCTCCGAATC *R:* CGACGCGCGATATTTTATTA	Validated	KP274962
Dp17	$[TGCG]_7$	150	*F:* TGCTCCCTCCCTTCCTTAAT *R:* AAAAATGTGGTAGCCCTAGATCC	Validated	KP274963
Dp18	$[GTCC]_{13}$	104	*F:* AGAGGCGGAGGGATATTGTT *R:* TATGGCTCCGGACACAAAAG	Validated	KP274964
Dp19	$[AT]_{17}$	377	*F:* TGGGATATTAATTTGGCTTCAA *R:* CAATTATGGCTGACCAAGCA	Validated	KP274965
Dp20	$[ATC]_{24}$	224	*F:* CGAAAAACATATCACGCTTTCA *R:* TCGTGTTTTGCTGGTTCTGA	Validated	KP274966
Dp21	$[TCCG]_{11}$	416	*F:* TTATCCCCACGCTTTTTGAA *R:* GAAATTTTTGGTGGGTGTGG	Not amplified	-
Dp22	$[TA]_{23}$	246	*F:* TGCATGAGTGGTGTCAATGT *R:* GTCCTCTGCACAAACCATGA	Validated	KP274967
Dp23	$[TA]_{18}$	285	*F:* TAGTCGCATGCATGAGTGGT *R:* GGGCATATGTCATCCTACGG	Validated	KP274968
Dp24	$[TA]_{11}$	286	*F:* CCCAGACTGCCCATATCCT *R:* CACTGGAACACCAGTCATGC	Validated	KP274969
Dp25	$[CTG]_{43}$	341	*F:* TGCAATCTTGAGAATTGTGG	Not amplified	-

			R: CACAACCAGCGGTAGCATTA		
Dp26	$[CGGA]_8$	117	*F:* TTCAATGAAATCCGCCAAAG *R:* GACCTAATTATCCCCGCCATA	Validated	KP274970
Dp27	$[TAG]_{18}$	102	*F:* ATGTGCATAGTTTGGGCAGA *R:* CATATTCTACTGCCAGCACCA	Validated	KP274971
Dp28	$[TGA]_{13}$	140	*F:* TAGTAGTGACGCTGCTGCTG *R:* CTTTCATTCACAATTTCCACTTT	Validated	KP274972
Dp29	$[CCGC]_9$	119	*F:* TTTGGTGAAGATCGGATGAA *R:* CGGGACGTATTATGTGAAACC	Validated	KP274973
Dp30	$[TTG]_7$ 58 $[GTT]_6$	218	*F:* GCGTTGGTGTTGTGTACGTC *R:* CTGAGCATCTCACCGTCAAA	Polymorphic	JQ812987
Dp31	$[ATT]_{13}$	267	*F:* CGAGTTTCTTGCACGTTTCA *R:* TGTTATTTTAAGAAGGCCACATTG	Polymorphic	JQ812988
Dp32	$[ATC]_{14}$ 24 $[TCG]_{12}$	348	*F:* CATCGTTATGGTCGTCATCG *R:* TTTTAGCGACAAATTGACTTGG	Validated	KP274974
Dp33	$[CAT]_6$ $[TCT]_8$	447	*F:* ATCATCATCGTCGTCGTCGT *R:* CACAACGGTCCGAAGTCATA	Validated	KP274975
Dp34	$[CAT]_{19}$	405	*F:* GCTAAGGGCGACAGTGTTGT *R:* AACACAATGATGCTGCTGCT	Validated	KP274976
Dp35	$[GTCG]_9$	382	*F:* ACTGGACAGCAACCAGCTTT *R:* TCTGACAACCACCTGAAGGA	Validated	KP274977
Dp36	$[GAAG]_{10}$	200	*F:* CTTGACCTTGACCCCATTGA *R:* AATAAAATTTTGGCGGAGCA	Validated	KP274978
Dp37	$[ATG]_9$	277	*F:* TTCCCATCTTAAGTGCATTGTG *R:* TTGGGTCATAGGGAGTCTGG	Validated	KP274979
Dp38	$[TCCG]_{12}$	391	*F:* TCGTCCAAATGACCAAATGT *R:* GCTTTTAGTGTAGGAGGAGATAGTGG	Validated	KP274980
Dp39	$[GGCG]_{11}$	389	*F:* GACGTCATGGTTCTGAATGG *R:* CCGGACAAGCTCATTTATGG	Polymorphic	JQ812989
Dp40	$[TTAA]_7$	110	*F:* GGTCAGAAGATTGGCCTCAA *R:* TCGAAAGGGGGCATAAAAAT	Not amplified	-
Dp41	$[TCTG]_8$	113	*F:* AGGGGTGGGGCATTTTTAT *R:* AGAGCGGACACGAAAAGTGT	Validated	KP274981
Dp42	$[GTTG]_9$	239	*F:* TCGCTTAACCTGACCAGTGA *R:* CCAAATATCAAGTTGCCTATCTTCA	Polymorphic	JQ812990
Dp43	$[TTA]_{10}$	240	*F:* TTGCTCATGATGAAATATGATGT *R:* ATGCGTTTCACTTTGGCATC	Polymorphic	JQ812991
Dp44	$[GACC]_8$	148	*F:* CCCCAAGCGTCTTGAGTATC *R:* TCCTGCCAAGCATGTATGAG	Polymorphic	JQ812992
Dp45	$[TGT]_8$	412	*F:* GCATAACGGCTAAGTTGGTTTT *R:* TGCTTCGTACTCAGCGATTT	Validated	KP274982
Dp46	$[ACCG]_6$	380	*F:* AGGGGTCATCTGCAAGTCAT *R:* TGGGGGATATGTGTGTTGTT	Validated	KP274983
Dp47	$[GAGC]_6$	434	*F:* AGGGTGCTGCATTCTTATCG *R:* GCTTGTCCAACGCAGCTATT	Not amplified	-
Dp48	$[TAAA]_5$	448	*F:* TATCCCGGCCCTATAGGAAA *R:* GACAGCTCCCTCATCTTTGC	Validated	KP274984
Dp49	$[AATA]_5$	440	*F:* TGTCAATGGCGAACAGAGAA *R:* GATTGGCACAAACCTGAAAAA	Validated	KP274985
Dp50	$[TTA]_6$	335	*F:* GCGCTCGATGTCAATGATTA *R:* CGCGTGATTTAACAATGTCG	Validated	KP274986
Dp51	$[AAAT]_5$	301	*F:* CACAACTTTTGATTTTTGACACC *R:* ATGCTGTCTGCTGGCGTTAT	Validated	KP274987
Dp52	$[AAG]_5$	208	*F:* GGCCGTTTTAAATCGTGAAA *R:* TGTTTGTTTCCCCTTTCTGG	Validated	KP274988
Dp53	$[TTATA]_6$	205	*F:* TAGTGCATGCTGGGTAGCAC	Validated	KP274989

			R: TGGCCTTCCAGGTAAATACG		
Dp54	$[ATC]_5$	287	*F:* TGTCATCGTAAACCCAGTCG *R:* AATGGGCGTAACATTTGCAT	Validated	KP274990
Dp55	$[ATACA]_6$	261	*F:* AGGGTTACACGCACAGGTTC *R:* CCTACAAAACGCGCACATAC	Validated	KP274991
Dp56	$[CCGT]_5$	279	*F:* CGCAAATGTTCTATTGAGCGTA *R:* CTAAGGTATGGCTCCGGACA	Validated	KP274992
Dp57	$[AAC]_5$	271	*F:* TCCCCCGTAACCAACTAACA *R:* TCACTCTGGATCAGCAGACG	Not amplified	-
Dp58	$[ACACC]_7$	205	*F:* AACGACACCGCATCACTCTA *R:* TGGTGTGGTGTTTATTGGAA	Validated	KP274993
Dp59	$[CA]_8$	278	*F:* AGGGTAGGGTCAACCACAACT *R:* CAAAATAGTATGCGTCTCGGAAT	Not amplified	-
Dp60	$[ATT]_6$	138	*F:* CATGACACAACTTTCCCCTTA *R:* CCGCTACATACCGATTGACA	Validated	KP274994
Dp61	$[TCCG]_6$	121	*F:* CGGAGGGGACTTATGGTTTT *R:* TGCCATCTATCCATGTTCCA	Validated	KP274995
Dp62	$[ACAG]_5$	141	*F:* AGGGATGCTCCACATGAAAT *R:* CTCGGGGGTGAGCTATTGT	Validated	KP274996
Dp63	$[TGT]_5$	125	*F:* TTGACGCCAACAATAACAGTTT *R:* ACTGGTGGTGGGAACGATAA	Validated	KP274997
D64	$[TTG]_6$	149	*F:* TGCCAGTAGTGACGTTGGAG *R:* ACAGGTCATGAACCCTCCTC	Validated	KP274998
Dp65	$[TTA]_5$	141	*F:* TCCTAACATCGATCTGTTCCAA *R:* CGGCTGCTAAGGGGTAAAGT	Validated	KP274999
Dp66	$[TGT]_5$	141	*F:* GCCGTGGTTCACCTTTTT *R:* AGCTAATACGCCCATTGCAT	Validated	KP275000
Dp67	$[GTTG]_7$	150	*F:* GAACATTGTTTTTGCCCTGTC *R:* CCACTCAAAATGTGCAGCTC	Validated	KP275001
Dp68	$[TGTTC]_5$	287	*F:* TGCTACACACCGTATTTGCTG *R:* ACACGTGGATGGTGTGAAGA	Polymorphic	JQ812993
Dp69	$[GACA]_7$	257	*F:* GGGGTCAGATCAAAATTCCA *R:* TGTCCAGCATCGTCAAAAAG	Validated	KP275002
Dp70	$[TTG]_5$	212	*F:* GCAGCGACATCAGACAACAT *R:* TTTCAGCTGTTTTATTGGTATGAA	Validated	KP275003
Dp71	$[ACAA]_6$	358	*F:* CTGCCATCAGCATCTGAAAA *R:* TGACTGGAGTTGAGCACAGG	Validated	KP275004
Dp72	$[GGTA]_8$	383	*F:* TGCACACACATCTTGACCTG *R:* GCTGAAGGCACAACATTTGA	Polymorphic	JQ812994
Dp73	$[ACTG]_5$	217	*F:* GGATGACCTTCACCTTCACC *R:* ATCGCCATTGCTTTTTATGC	Validated	KP275005
Dp74	$[CGTC]_9$	361	*F:* ATCCCCTCAAGACGTTTCCT *R:* ACCATACCGGTGGCATAAAA	Polymorphic	JQ812995
Dp75	$[ATTT]_6$	336	*F:* AAAAACTAGGGTAGGTAGGTACGG *R:* GCACAGCATTGATGTTCTTCT	Not amplified	-
Dp76	$[GACG]_9$	312	*F:* AAATCCGCCAAAGAACTTCC *R:* AAAAGGGGATTTGCCGTAAC	Validated	KP275006
Dp77	$[AAAT]_7$	259	*F:* GGAAGACCTACCTACGTTCAACC *R:* GGTGATGGCCATGATGATTA	Validated	KP275007
Dp78	$[AACC]_6$	100	*F:* AGTTATTGCAAATGTTAAAGTTGGTG *R:* CCACTAAAGTATTGGGGGACA	Validated	KP275008
Dp79	$[AAC]_5$	100	*F:* TGGTGTAAGTTAGCCATGCAA *R:* TTTGTAGCTTAATTGCCGAAA	Validated	KP275009
Dp80	$[ACA]_5$	113	*F:* AGCGAACCATAGCTCGTACC *R:* GCTTAAATCAAACGATTCAAGTGT	Validated	KP275010
Dp81	$[GTCT]_6$	138	*F:* TGGCAGTCTCACTGTTCACC	Validated	KP275011

			R: TCAAAAACGGCTAGAAAACTCC		
Dp82	$[GCAC]_6$	106	*F:* CAGAATTGGACCAGTAGTTTAGGA *R:* CCACGCTAATGTCATGGTCA	Validated	KP275012
Dp83	$[ACGG]_7$	294	*F:* TGAGAAATAGCCCGGACAAA *R:* GATATCCGACCTCCGTTCAA	Not amplified	-
Dp84	$[CGTC]_{11}$	364	*F:* TGTAGACCGTTTGAAACTTGC *R:* AAGATCGGACAAGAAATGTGG	Validated	KP275013
Dp85	$[TCA]_{39}$	259	*F:* TTTGGGGTAAAATGCCTGAC *R:* GATGATGATGGTGGTGTTGG	Validated	KP275014
Dp86	$[CGTC]_5$	312	*F:* GCAAAGGGAGAAAACTGCAC *R:* CACTGTCACCGTCGCACATA	Polymorphic	JQ812996
Dp87	$[AACC]_7$	302	*F:* GATGACCTTGACCTTTCACCA *R:* TGGGCATATGTCATCCTATGG	Not amplified	-
Dp88	$[TTTA]_7$	238	*F:* ATAGGCGCTCGATGTCAATG *R:* CAATAAGCGAGAGGGGTGAC	Validated	KP275015
Dp89	$[CGTC]_8$	275	*F:* TTTTCACACAGCAGCCAAAG *R:* TGAGAAATAGCCCGGACAAA	Polymorphic	JQ812997
Dp90	$[GGAC]_6$	216	*F:* GTTTTTGACCCAGCATGACC *R:* CAAAGCGTATAGGGGGCATA	Validated	KP275016
Dp91	$[ATA]_8$	243	*F:* AAAGCGCACTTCCGTACACT *R:* CGTACATGCACAACACAACG	Not amplified	-
Dp92	$[GACG]_6$	241	*F:* GAAATTGCTATCTGAAGCGACA *R:* GGTCATGCAGGGTCAAAAAC	Validated	KP275017
Dp93	$[AGAC]_7$	138	*F:* TTTTGCATGATAAAAGTTCGTGA *R:* CCGGTATGTTTTCCACCACT	Validated	KP275018

Table S4. Allele frequencies for the 14 polymorphic microsatellite loci in the six locations. Refer to Table S1 for Sample numbers.

Locus	Allele	Ebro River Basin					Llobregat River Basin
		1	2	3	4	5	6
Dp1	299	0.1	0.2	0	0.25	0.429	0.125
	308	0.3	0.2	0.5	0.083	0.071	0.125
	314	0.2	0.2	0	0.25	0.429	0.125
	317	0.4	0.2	0.5	0.25	0.071	0.125
	323	0	0.2	0	0.167	0	0.5
Dp2	411	0	0	0	0	0.062	0
	429	0	0	0	0	0	0.1
	432	0	0	0	0.062	0.062	0
	435	0	0	0.375	0.312	0.188	0
	438	0	0	0.125	0	0	0
	441	1	0.812	0.375	0.375	0.188	0.5
	444	0	0	0.062	0.25	0.25	0.4
	453	0	0.188	0	0	0.25	0
	459	0	0	0.062	0	0	0
Dp7	430	0.438	0.625	0.125	0.688	0.375	0.214
	436	0.562	0.375	0.875	0.312	0.625	0.786
Dp30	233	0.312	0.25	0.125	0.125	0.125	0.357
	236	0.5	0.438	0.562	0.562	0.75	0.286
	242	0	0	0	0.188	0	0.286
	245	0.062	0.125	0.125	0	0	0.071
	251	0.125	0.125	0.188	0.125	0.125	0
	254	0	0.062	0	0	0	0
Dp31	250	0.125	0.062	0	0.125	0.062	0
	253	0.062	0.062	0	0.062	0.188	0.143
	262	0.125	0.062	0.312	0.375	0.25	0.214
	277	0	0	0.062	0	0	0
	280	0.062	0.125	0.125	0.188	0.062	0
	283	0.312	0.188	0.188	0.062	0.25	0.5
	295	0.062	0	0.062	0	0.125	0
	307	0.062	0.062	0	0	0	0
	310	0.062	0	0.062	0	0	0.071
	313	0.062	0.062	0	0	0	0
	325	0.062	0.312	0.125	0.188	0.062	0.071
	328	0	0.062	0	0	0	0
	337	0	0	0.062	0	0	0
Dp39	404	0.312	0.375	0.25	0.312	0.312	0.062
	408	0.188	0.125	0.312	0.375	0.188	0.75
	412	0.5	0.5	0.438	0.312	0.5	0.188
Dp42	261	0.125	0.25	0.375	0.25	0.312	0.188
	265	0.125	0.312	0.125	0.125	0.312	0.125
	277	0.75	0.438	0.5	0.625	0.375	0.688
Dp43	242	0	0	0	0	0	0.071
	260	0.812	0.625	0.75	0.938	0.688	0.857
	263	0	0	0	0	0.125	0
	269	0.062	0.125	0.188	0.062	0.188	0.071
	278	0.125	0.25	0.062	0	0	0
Dp44	156	0	0.125	0	0	0.062	0
	160	0.188	0.062	0.125	0.312	0.125	0.071
	164	0.062	0.312	0.562	0.5	0.438	0.357
	168	0	0	0	0	0.188	0
	172	0.75	0.5	0.312	0.188	0.188	0.571
Dp68	310	0.875	0.875	0.938	0.812	0.938	0.857
	315	0.125	0.125	0.062	0.125	0.062	0.143
	320	0	0	0	0.062	0	0
Dp72	397	0.562	0.25	0.25	0.438	0.188	0
	401	0.438	0.75	0.75	0.562	0.812	1

Dp74	351	0.125	0.125	0.125	0.188	0.188	0.083
	371	0.625	0.688	0.438	0.5	0.562	0.75
	379	0.25	0.188	0.438	0.312	0.25	0.167
Dp86	324	0.375	0.25	0.375	0.5	0.5	0.562
	328	0.188	0.25	0.438	0.188	0.188	0.25
	332	0.062	0.125	0.062	0.188	0.125	0.125
	336	0.375	0.375	0.125	0.125	0.188	0.062
Dp89	279	0.062	0	0	0	0.062	0
	291	0.25	0.438	0.188	0.562	0.375	0.5
	299	0.625	0.5	0.812	0.438	0.562	0.5
	303	0.062	0.062	0	0	0	0

Table S5. Pairwise F_{ST} values among populations. Above diagonal: pairwise F_{ST} values. Below diagonal: P-values obtained after 300 permutations. Refer to Table S1 for sample numbers.

	Ebro River Basin					Llobregat River Basin
	1	2	3	4	5	6
1		-0.009	0.062	0.0539	0.0783	0.0985
2	0.84333		0.0451	0.0147	0.0146	0.0622
3	0.16333	0.05		0.0354	0.0276	0.0683
4	0.03667	0.1	0.26		-0.0013	0.0586
5	0.05667	0.23	0.10333	0.73667		0.0619
6	0.05	0.03	0.16	0.16333	0.15	

Table S6. Number of microsatellite markers used in the different genetic analyses. N: number of the study. *Cites from 79 to 89 belong to study 78.

Application	N	Reference	SSR used
1. LINKAGE MAP	1	Woram RA, McGowan C, Stout JA, Gharbi K, Ferguson MM, Hoyheim B, Davidson EA, Davidson WS, Rexroad C, Danzmann RG (2004) A genetic linkage map for Arctic char (*Salvelinus alpinus*): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents. Genome 47: 304-315.	184
	2	Moen T, Hoyheim B, Munck H, Gomez-Raya L (2004) A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. Anim Genet 35: 81-92.	31
	3	Waldbieser GC, Bosworth BG, Nonneman DJ, Wolters WR (2001) A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*. Genetics 158: 727-734.	262
	4	Ohara E, Nishimura T, Nagakura Y, Sakamoto T, Mushiake K, Okamoto N (2005) Genetic linkage map of two yellowtails (*Seriola quinqueradiata* and *Seriola lalandi*). Aquaculture 244: 41-48.	297
	5	Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM (2001) The genetic architecture of divergence between threespine stickleback species. Nature 414: 901-905.	281
	6	Kocher TD, Lee WJ, Sobolewska H, Penman D, McAndrew B (1998) A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). Genetics 148: 1225-1232.	59
	7	Agresti JJ, Seki S, Cnaani A, Poompuang S, Hallerman EM, Umiel N, Hulata G, Gall GAE, May B (2000) Breeding new strains of tilapia: development of an artificial center of origin and linkage map based on AFLP and microsatellite loci. Aquaculture 185: 43-56.	63
	8	Lee BY, Lee WJ, Streelman JT, Carleton KL, Howe AE, Hulata G, Slettan A, Stern JE, Terai Y, Kocher TD (2005) A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). Genetics 170: 237-244.	525
	9	Sakamoto T, Danzmann RG, Gharbi K, Howard P, Ozaki A, Khoo SK, Woram RA, Okamoto N, Ferguson MM, Holm LE, Guyomard R, Hoyheim B (2000) A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. Genetics 155: 1331-1345.	191
	10	Nichols KM, Young WP, Danzmann RG, Robison BD, Rexroad C, Noakes M, Phillips RB, Bentzen P, Spies I, Knudsen K, Allendorf FW, Cunningham BM, Brunelli J, Zhang H, Ristow S, Drew R, Brown KH, Wheeler PA, Thorgaard GH (2003) A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). Anim Genet 34: 102-115.	226
	11	Chistiakov DA, Hellemans B, Haley CS, Law AS, Tsigenopoulos CS, Kotoulas G, Bertotto D, Libertini A, Volckaert FA (2005) A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L. Genetics 170: 1821-1826.	162
	12	Inami M, Hatanaka A, Mitsuboshi T, Yamada S, Tateishi A, Fukuda H, Sakamoto T (2005) A microsatellite linkage map of red sea bream (*Pagrus major*) and mapping of QTL markers associated with resistance to red sea bream iridovirus (RSIV). Plant & Animal Genomes XIII Conference .	136
	13	Walter RB, Rains JD, Russell JE, Guerra TM, Daniels C, Johnston DA, Kumar J, Wheeler A, Kelnar K, Khanolkar VA, Williams EL, Hornecker JL, Hollek L, Mamerow MM, Pedroza A, Kazianis S (2004) A microsatellite genetic linkage map for *Xiphophorus*. Genetics 168: 363-	256

		372.	
	14	Knapik EW, Goodman A, Atkinson OS, Roberts CT, Shiozawa M, Sim CU, Weksler-Zangen S, Trolliet MR, Futrell C, Innes BA, Koike G, McLaughlin MG, Pierre L, Simon JS, Vilallonga E, Roy M, Chiang PW, Fishman MC, Driever W, Jacob HJ (1996) A reference cross DNA panel for zebrafish (*Danio rerio*) anchored with simple sequence length polymorphisms. Development 123: 451-460.	102
	15	Knapik EW, Goodman A, Ekker M, Chevrette M, Delgado J, Neuhauss S, Shimoda N, Driever W, Fishman MC, Jacob HJ (1998) A microsatellite genetic linkage map for zebrafish (*Danio rerio*). Nat Genet 18: 338-343.	705
	16	Woods IG, Kelly PD, Chu F, Ngo-Hazelett P, Yan YL, Huang H, Postlethwait JH, Talbot WS (2000) A comparative map of the zebrafish genome. Genome Res 10: 1903-1914.	616
	17	Shimoda N, Knapik EW, Ziniti J, Sim C, Yamada E, Kaplan S, Jackson D, de SF, Jacob H, Fishman MC (1999) Zebrafish genetic map with 2000 microsatellite markers. Genomics 58: 219-232.	2000
	18	Singer A, Perlman H, Yan Y, Walker C, Corley-Smith G, Brandhorst B, Postlethwait J (2002) Sex-specific recombination rates in zebrafish (*Danio rerio*). Genetics 160: 649-657.	141
	19	Coimbra MRM, Kobayashi K, Koretsugu S, Hasegawa O, Ohara E, Ozaki A (2003) A genetic linkage map of the Japanese flounder, *Paralichtys olivaceus*. Aquaculture 220: 203-218.	183
2. INDIVIDUAL DNA IDENTIFICATION AND PARENTAGE ASSIGMENT	20	Miller ML, Kapuscinski AR (1996) Microsatellite DNA markers reveal new levels of genetic variation in northern pike. TAFS 125: 971-977.	4
	21	Larsen PF, Hansen MM, Nielsen EE, Jensen LF, Loeschcke V (2005) Stocking impact and temporal stability of genetic composition in a brackish northern pike population (*Esox lucius* L.), assessed using microsatellite DNA analysis of historical and contemporary samples. Heredity (Edinb) 95: 136-143.	8
	22	Hansen MM, Ruzzante DE, Nielsen EE, Bekkevold D, Mensberg KL (2002) Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (*Salmo trutta*) populations. Mol Ecol 11: 2523-2535.	9
	23	Zierdt H, Hummel S, Herrmann B (1996) Amplification of human short tandem repeats from medieval teeth and bone samples. Hum Biol 68: 185-199.	1
	24	Bentzen P, Olsen JB, McLean JE, Seamons TR, Quinn TP (2001) Kinship analysis of Pacific salmon: insights into mating, homing, and timing of reproduction. J Hered 92: 127-136.	14
	25	Castro J, Bouza C, Presa P, Pino-Querido A, Riaza A, Ferreiro L, Sanchez L, Martinez LM (2004) Potential sources of error in parentage assessment of turbot (*Scophthalmus maximus*) using microsatellite loci. Aquaculture 242: 119-135.	12
	26	Doyle RW, Perez-Enriquez R, Takagi M, Taniguchi N (2001) Selective recovery of founder genetic diversity in aquacultural broodstocks and captive, endangered fish populations. Genetica 111: 291-304.	5
	27	Guinand B, Scribner KT, Page KS, Burnham-Curtis MK (2003) Genetic variation over space and time: analyses of extinct and 28remnant lake trout populations in the Upper Great Lakes. Proc Biol Sci 270: 425-433.	5
	28	Hauser L, Adcock GJ, Smith PJ, Ramirez JH, Carvalho GR (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). Proc Natl Acad Sci U S A 99: 11742-11747.	7
	29	Herbinger CM, Doyle RW, Pitman ER, Paquet D, Mesa KA, Morris DB,	4

		Wright JM, Cook D (1995) DNA fingerprintbased analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. Aquaculture 137: 245-256.	
	30	Hutchinson WF, van OC, Rogers SI, Carvalho GR (2003) Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*). Proc Biol Sci 270: 2125-2132.	3
	31	King TL, Kalinowski ST, Schill WB, Spidle AP, Lubinski BA (2001) Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. Mol Ecol 10: 807-821.	12
	32	Neff BD (2001) Genetic paternity analysis and breeding success in bluegill sunfish (*Lepomis macrochirus*). J Hered 92: 111-119.	11
	33	Norris AT, Bradley DG, Cunningham EP (2000) Parentage and relatedness determination in farmed Atlantic salmon (*Salmo salar*) using microsatellite markers. Aquaculture 182: 73-83.	8
	34	Waldbieser GC, Wolters WR (1999) Application of polymorphic microsatellite loci in a channel catfish, *Ictalurus punctatus*, breeding program.J. World Maric. Soc 30: 256-262.	13
3. PHYLOGENY, POPULATION AND CONSERVATION GENETICS	35	Harr B, Kauer M, Schlotterer C (2002) Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. Proc Natl Acad Sci U S A 99: 12949-12954.	1
	36	Streelman JT, Zardoya R, Meyer A, Karl SA (1998) Multilocus phylogeny of cichlid fishes (*Pisces: Perciformes*): evolutionary comparison of microsatellite and single-copy nuclear loci. Mol Biol Evol 15: 798-808.	1
	37	Zardoya R, Vollmer DM, Craddock C, Streelman JT, Karl S, Meyer A (1996) Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (*Pisces: Perciformes*). Proc Biol Sci 263: 1589-1598.	1
	38	Shaw PW, Turan C, Wright JM, O'Connell M, Carvalho GR (1999) Microsatellite DNA analysis of population structure in Atlantic herring (*Clupea harengus*), with direct comparison to allozyme and mtDNA RFLP analyses. Heredity (Edinb) 83 (Pt 4): 490-499.	4
	39	McCartney MA, Acevedo J, Heredia C, Rico C, Quenoville B, Bermingham E, McMillan WO (2003) Genetic mosaic in a marine species flock. Mol Ecol 12: 2963-2973.	5
	40	Stamford MD, Taylor EB (2004) Phylogeographical lineages of Arctic grayling (Thymallus arcticus) in North America: divergence, origins and affinities with Eurasian *Thymallus*. Mol Ecol 13: 1533-1549.	5
	41	Corujo M, Blanco G, Vazquez E, Sanchez JA (2004) Genetic structure of northwestern Spanish brown trout (*Salmo trutta* L.) populations, differences between microsatellite and allozyme loci. Hereditas 141: 258-271.	6
	42	Dannewitz J, Maes GE, Johansson L, Wickstrom H, Volckaert FA, Jarvi T (2005) Panmixia in the European eel: a matter of time. Proc Biol Sci 272: 1129-1137.	6
	43	Heath DD, Pollard S, Herbinger C (2001) Genetic structure and relationships among steelhead trout (*Oncorhynchus mykiss*) populations in British Columbia. Heredity (Edinb) 86: 618-627.	6
	44	Hendry AP, Wenburg JK, Bentzen P, Volk EC, Quinn TP (2000) Rapid evolution of reproductive isolation in the wild: evidence from introduced salmon. Science 290: 516-519.	6
	45	Lemaire C, Versini JJ, Bonhomme F (2005) Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. J Evol Biol 18: 70-80.	6
	46	Pujolar JM, Maes GE, Vancoillie C, Volckaert FA (2005) Growth rate	6

	correlates to individual heterozygosity in the European eel, *Anguilla anguilla* L. Evolution 59: 189-199.	
47	Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, Pericic M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioglu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan JP, Chaventre A, Ferak V, Furedi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzic R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Semino O (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe. Am J Hum Genet 75: 128-137.	6
48	Ruzzante DE, Taggart CT, Cook D (1999) A review of the evidence for genetic structure of cod (*Gadus morhua*) populations in the NW Atlantic and population affinities of larval cod off Newfoundland and the Gulf of St. Lawrence. Fish Res 43: 79-97.	6
49	Gold JR, Burridge CP, Turner TF (2001) A modified stepping-stone model of population structure in red drum, Sciaenops ocellatus (*Sciaenidae*), from the northern Gulf of Mexico. Genetica 111: 305-317.	7
50	Reusch TB, Wegner KM, Kalbe M (2001) Rapid genetic divergence in postglacial populations of threespine stickleback (*Gasterosteus aculeatus*): the role of habitat type, drainage and geographical proximity. Mol Ecol 10: 2435-2445.	7
51	Fillatre EK, Etherton P, Heath DD (2003) Bimodal run distribution in a northern population of sockeye salmon (*Oncorhynchus nerka*): life history and genetic analysis on a temporal scale. Mol Ecol 12: 1793-1805.	8
52	Nielsen EE, Nielsen PH, Meldrup D, Hansen MM (2004) Genetic population structure of turbot (*Scophthalmus maximus* L.) supports the presence of multiple hybrid zones for marine fishes in the transition zone between the Baltic Sea and the North Sea. Mol Ecol 13: 585-595.	8
53	Spencer CC, Neigel JE, Leberg PL (2000) Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. Mol Ecol 9: 1517-1528.	8
22	Hansen MM, Ruzzante DE, Nielsen EE, Bekkevold D, Mensberg KL (2002) Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (*Salmo trutta*) populations. Mol Ecol 11: 2523-2535.	9
54	Vasemagi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in atlantic salmon (*Salmo salar* L.). Mol Biol Evol 22: 1067-1076.	9
55	Brown KM, Baltazar GA, Hamilton MB (2005) Reconciling nuclear microsatellite and mitochondrial marker estimates of population structure: breeding population structure of Chesapeake Bay striped bass (*Morone saxatilis*). Heredity (Edinb) 94: 606-615.	10
56	Triantafyllidis A, Krieg F, Cottin C, Abatzopoulos TJ, Triantaphyllidis C, Guyomard R (2002) Genetic structure and phylogeography of European catfish (*Silurus glanis*) populations. Mol Ecol 11: 1039-1055.	10
57	Ramstad KM, Woody CA, Sage GK, Allendorf FW (2004) Founding events influence genetic population structure of sockeye salmon (*Oncorhynchus nerka*) in Lake Clark, Alaska. Mol Ecol 13: 277-290.	11

	58	Gum B, Gross R, Kuehn R (2005) Mitochondrial and nuclear DNA phylogeography of European grayling (*Thymallus thymallus*): evidence for secondary contact zones in central Europe. Mol Ecol 14: 1707-1725.	12
	59	O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between F and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. Mol Ecol 13: 1799-1814.	14
	60	Bérubé M, Urban J, Dizon AE, Brownell RL, Palsboll PJ (2002) Genetic identification of a small and highly isolated population of a fin whales (*Balaenoptera physalus*) in the Sea of Cortez, Mexico. Conserv Genet 3: 183-190.	16
	61	Koskinen MT, Nilsson J, Veselov AJ, Potutkin AG, Ranta E, Primmer CR (2002) Microsatellite data resolve phylogeographic patterns in European grayling, *Thymallus thymallus*, Salmonidae. Heredity (Edinb) 88: 391-401.	17
	62	Goldstein DB, Roemer GW, Smith DA, Reich DE, Bergman A, Wayne RK (1999) The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics 151: 797-801.	19
	63	Nair S, Williams JT, Brockman A, Paiphun L, Mayxay M, Newton PN, Guthmann JP, Smithuis FM, Hien TT, White NJ, Nosten F, Anderson TJ (2003) A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. Mol Biol Evol 20: 1526-1536.	33
4. MOLECULAR EPIDEMOLOGY AND PATHOLOGY	64	Khoo SK, Ozaki A, nakamura F, arakawa T, Ishimoto S, Nicolov R, Sakamoto T, Akatsu T, Mochuziki M, Denda L, Okamoto N (2004) Identification of a novel chromosomal region associated with infectious hemapoietic necrosis (INH) resistance in rainbow trout *Oncorhynchus mykiss.* Fish Pathol 39: 95-101.	4
	65	Ozaki A, Sakamoto T, Khoo S, Nakamura K, Coimbra MR, Akutsu T, Okamoto N (2001) Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). Mol Genet Genomics 265: 23-31.	51
	12	Inami M, Hatanaka A, Mitsuboshi T, Yamada S, Tateishi A, Fukuda H, Sakamoto T (2005) A microsatellite linkage map of red sea bream (*Pagrus major*) and mapping of QTL markers associated with resistance to red sea bream iridovirus (RSIV). Plant & Animal Genomes XIII Conference.	136
	66	Cnaani A, Zilberman N, Tinman S, Hulata G, Ron M (2004) Genome-scan analysis for quantitative trait loci in an F2 tilapia hybrid. Mol Genet Genomics 272: 162-172.	40
	67	Fornoni A, Wang Y, Lenz O, Striker LJ, Striker GE (2002) Association of a decreased number of d(CA) repeats in the matrix metalloproteinase-9 promoter with glomerulosclerosis susceptibility in mice. J Am Soc Nephrol 13: 2068-2076.	1
	68	Gibson G (2002) A genetic attack on the defense complex. Bioessays 24: 487-489.	227
	69	Grimholt U, Drablos F, Jorgensen SM, Hoyheim B, Stet RJ (2002) The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. Immunogenetics 54: 570-581.	1
	70	Saleem Q, Anand A, Jain S, Brahmachari SK (2001) The polyglutamine motif is highly conserved at the Clock locus in various organisms and is	1

	not polymorphic in humans. Hum Genet 109: 136-142.	
71	Schartl M, Wilde B, Hornung U (1998) Triplet repeat variability in the signal peptide sequence of the Xmrk receptor tyrosine kinase gene in *Xiphophorus* fish. Gene 224: 17-21.	1
72	Wang XB, Kakoulidou M, Giscombe R, Qiu Q, Huang D, Pirskanen R, Lefvert AK (2002) Abnormal expression of CTLA-4 by T cells from patients with myasthenia gravis: effect of an AT-rich gene sequence. J Neuroimmunol 130: 224-232.	1
73	Zimmerman AM, Evenhuis JP, Thorgaard GH, Ristow SS (2004) A single major chromosomal region controls natural killer cell-like activity in rainbow trout. Immunogenetics 55: 825-835.	39
74	Perry GM, Danzmann RG, Ferguson MM, Gibson JP (2001) Quantitative trait loci for upper thermal tolerance in outbred strains of rainbow trout (*Oncorhynchus mykiss*). Heredity (Edinb) 86: 333-341.	3
75	Perry GM, Ferguson MM, Sakamoto T, Danzmann RG (2005) Sex-linked quantitative trait loci for thermotolerance and length in the rainbow trout. J Hered 96: 97-107.	6
76	Shirak A, Palti Y, Cnaani A, Korol A, Hulata G, Ron M, Avtalion RR (2002) Association between loci with deleterious alleles and distorted sex ratios in an inbred line of tilapia (*Oreochromis aureus*). J Hered 93: 270-276.	3
77	Somorjai IM, Danzmann RG, Ferguson MM (2003) Distribution of temperature tolerance quantitative trait loci in Arctic charr (*Salvelinus alpinus*) and inferred homologies in rainbow trout (*Oncorhynchus mykiss*). Genetics 165: 1443-1456.	2
78	Naidoo R, Chetty R (1998) The application of microsatellites in molecular pathology. Pathol Oncol Res 4: 310-315.	
79*	French D, Cermele C, Lombardi AM, Vecchione A, Midulla C, Del NA, Vecchione A (1998) Microsatellite alterations in uterine leiomyomas. Anticancer Res 18: 349-351.	7
80*	Christensen M, Jensen MA, Wolf H, Orntoft TF (1998) Pronounced microsatellite instability in transitional cell carcinomas from young patients with bladder cancer. Int J Cancer 79: 396-401.	22
81*	Mark Z, Toren A, Amariglio N, Schiby G, Brok-Simoni F, Rechavi G (1998) Instability of dinucleotide repeats in Hodgkin's disease. Am J Hematol 57: 148-152.	7
82*	Ko TM, Hwa HL, Tseng LH, Lin YW, Cheung YP (1998) Fluorescence microsatellite analysis to study the parental origin of the supernumerary chromosome in Down's syndrome. Int J Gynaecol Obstet 61: 149-153.	9
83*	Aldaz CM, Chen T, Sahin A, Cunningham J, Bondy M (1995) Comparative allelotype of in situ and invasive human breast cancer: high frequency of microsatellite instability in lobular breast carcinomas. Cancer Res 55: 3976-3981.	20
84*	Duerr RH, Barmada MM, Zhang L, Davis S, Preston RA, Chensny LJ, Brown JL, Ehrlich GD, Weeks DE, Aston CE (1998) Linkage and association between inflammatory bowel disease and a locus on chromosome 12. Am J Hum Genet 63: 95-100.	12
85*	el-Naggar AK, Hurr K, Huff V, Clayman GL, Luna MA, Batsakis JG (1996) Microsatellite instability in preinvasive and invasive head and neck squamous carcinoma. Am J Pathol 148: 2067-2072.	25
86*	Huddart RA, Wooster R, Horwich A, Cooper CS (1995) Microsatellite instability in human testicular germ cell tumours. Br J Cancer 72: 642-645.	9

	87*	Quinn AG, Healy E, Rehman I, Sikkink S, Rees JL (1995) Microsatellite instability in human non-melanoma and melanoma skin cancer. J Invest Dermatol 104: 309-312.	10
	88*	Talwalkar VR, Scheiner M, Hedges LK, Butler MG, Schwartz HS (1998) Microsatellite instability in malignant melanoma. Cancer Genet Cytogenet 104: 111-114.	7
	89*	Watanabe M, Imai H, Shiraishi T, Shimazaki J, Kotake T, Yatani R (1995) Microsatellite instability in human prostate cancer. Br J Cancer 72: 562-564.	36
5. QUANTITATIVE TRAIT LOCI MAPPING	90	Behnke JM, Iraqi F, Menge D, Baker RL, Gibson J, Wakelin D (2003) Chasing the genes that control resistance to gastrointestinal nematodes. J Helminthol 77: 99-110.	15
	7	Agresti JJ, Seki S, Cnaani A, Poompuang S, Hallerman EM, Umiel N, Hulata G, Gall GAE, May B (2000) Breeding new strains of tilapia: development of an artificial center of origin and linkage map based on AFLP and microsatellite loci. Aquaculture 185: 43-56.	63
	66	Cnaani A, Zilberman N, Tinman S, Hulata G, Ron M (2004) Genome-scan analysis for quantitative trait loci in an F2 tilapia hybrid. Mol Genet Genomics 272: 162-172.	40
	91	Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, Schluter D, Kingsley DM (2004) The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biol 2(5): e109. doi: 10.1371/journal.pbio.0020109.	160
	92	Danzmann RG, Jackson TR, Ferguson M (1999) Epistasis in allelic expression at upper temperature tolerance QTL in rainbow trout. Aquaculture 173: 45-48.	61
	93	Jackson TR, Ferguson MM, Danzmann RG, Fishback AG, Ihssen PE, O'Connell M, Crease TJ (1998) Identification of two QTL influencing upper temperature tolerance in three rainbow trout (*Oncorhynchus mykiss*) half-sib families. Heredity 80: 143-151.	19
	94	Lee BY, Penman DJ, Kocher TD (2003) Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. Anim Genet 34: 379-383.	102
	95	Lee BY, Hulata G, Kocher TD (2004) Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). Heredity (Edinb) 92: 543-549.	119
	96	Nakamura K, Ozaki A, Akutsu T, Iwai K, Sakamoto T, Yoshizaki G, Okamoto N (2001) Genetic mapping of the dominant albino locus in rainbow trout (*Oncorhynchus mykiss*). Mol Genet Genomics 265: 687-693.	49
	97	O'Malley KG, Sakamoto T, Danzmann RG, Ferguson MM (2003) Quantitative trait loci for spawning date and body weight in rainbow trout: testing for conserved effects across ancestrally duplicated chromosomes. J Hered 94: 273-284.	201
	5	Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM (2001) The genetic architecture of divergence between threespine stickleback species. Nature 414: 901-905.	281
	98	Reid DP, Szanto A, Glebe B, Danzmann RG, Ferguson MM (2005) QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). Heredity (Edinb) 94: 166-172.	91
	99	Sakamoto T, Danzmann RG, Okamoto N, Ferguson MM, Ihssen PE (1999) Linkage analysis of quantitative loci trait loci associated with spawning time in rainbow trout (*Oncorhynchus mykiss*). Aquaculture 173: 33-43.	54

6. MARKER-ASSISTED SELECTION	100	Garcia de Leon FJ, Dallas JF, Chatain B, Canonne M, Versini JJ, Bonhomme F (1995) Development and use of microsatellite markers in sea bass, *Dicentrarchus labrax* (Linnaeus, 1758) (*Perciformes: Serrandidae*). Mol Mar Biol Biotechnol 4: 62-68.	7
	101	Blaszczyk L, Chelkowski J, Korzun V, Kraic J, Ordon F, Ovesna J, Purnhauser L, Tar M, Vida G (2004) Verification of STS markers for leaf rust resistance genes of wheat by seven European laboratories. Cell Mol Biol Lett 9: 805-817.	1
	102	Hara M, Shekino M (2003) Efficient detection of parentage in a cultured Japanese flounder *Paralichthys olivaceus* using microsatellite DNA markers. Aquaculture 217: 107-114.	4
	29	Herbinger CM, Doyle RW, Pitman ER, Paquet D, Mesa KA, Morris DB, Wright JM, Cook D (1995) DNA fingerprintbased analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. Aquaculture 137: 245-256.	4
	103	Jackson TR, Martin-Robichaud DJ, Reith ME (2003) Application of DNA markers to the management of Atlantic halibut (*Hippoglossus hippoglossus*) broodstock. Aquaculture 220: 245-259.	5
	104	Sekino M, Saitoh K, Yamada T, Kumagai A, Hara M, Yamashita Y (2003) Microsatellite-based pedigree tracing in a Japanese flounder *Paralichthys olivaceus* hatchery strain: implications for hatchery management related to stock enhancement program. Aquaculture 221: 255-263.	4

8.2

Supporting information from Chapter 2:

Identification of 246 microsatellites in the Asiatic clam (*Corbicula fluminea*)

Table S1. Repeat motif and GenBank Accession number for 246 SSRs isolated in C. *fluminea*.

*Numbers between two repeat motives correspond to the number of base pairs in composed microsatellites.

Locus	Repeat motif*	GenBank Accession Number
Cf1	$[GTT]_{19}$	KF730321
Cf2	$[AAC]_{12}$	KF730322
Cf3	$[CAA]_{12}$	KF744990
Cf4	$[ATG]_{16}$	KF744991
Cf5	$[TTA]_{15}$	KM589068
Cf6	$[TGT]_{15}$	KM589069
Cf7	$[GTT]_{5}$	KM589070
Cf8	$[GTT]_{15}$	KF730323
Cf9	$[AAC]_{8}$	KF730324
Cf10	$[TGT]_{9}$	KM589071
Cf11	$[TTATA]_{6}$	KM589072
Cf12	$[GTT]_{5}$	KF744992
Cf13	$[AAAG]_{9}$	KM58073
Cf14	$[TGA]_{7}$ $[GAC]_{5}$	KF730325
Cf15	$[ATG]_{6}$ $[TCA]_{19}$ $[TGA]_{12}$	KM589074
Cf16	$[TTG]_{9}$	KF744993
Cf17	$[GTT]_{5}$	KF744994
Cf18	$[TGT]_{8}$	KM589075
Cf19	$[CAA]_{8}$	KF744995
Cf20	$[ATC]_{8}$	KF744996
Cf21	$[ATGT]_{8}$	KF744997
Cf22	$[TTA]_{7}$	KF744998
Cf23	$[AAAT]_{6}$	KF744999
Cf24	$[TTGAT]_{9}$	KM589076
Cf25	$[ATTTG]_{5}$ 16 $[ATTG]_{5}$	KF745000
Cf26	$[ATC]_{7}$	KF745001
Cf27	$[CAT]_{30}$	KM589077
Cf28	$[TGA]_{21}$	KF730326
Cf29	$[TAG]_{8}$	KM589078
Cf30	$[ATTT]_{5}$	KF745002
Cf31	$[AAC]_{7}$	KF730327
Cf32	$[TAA]_{24}$	KM589079
Cf33	$[AGT]_{6}$ 16 $[TGC]_{10}$	KM589080
Cf34	$[TGT]_{5}$	KM589081
Cf35	$[ATT]_{5}$	KF745003
Cf36	$[ACCA]_{7}$	KF745004
Cf37	$[AAC]_{17}$ $[ACT]_{10}$ 50 $[CAA]_{6}$ $[ACT]_{22}$	KF745005
Cf38	$[GTT]_{7}$	KF745006
Cf39	$[ATG]_{12}$	KF745007
Cf40	$[CTGTT]_{5}$	KF745008
Cf41	$[ATTC]_{14}$	KM589082
Cf42	$[TTG]_{7}$	KM589083
Cf43	$[TCA]_{14}$	KF745009
Cf44	$[ATC]_{8}$	KF745010
Cf45	$[GTATA]_{16}$ $[GTGTG]_{3}$	KF745011
Cf46	$[TATT]_{5}$	KF745012
Cf47	$[AAC]_{8}$	KF745013
Cf48	$[AAC]_{5}$	KF745014
Cf49	$[TGT]_{5}$	KF745015

Cf50	$[GTCT]_6$	KF745016
Cf51	$[TCG]_5$	KF730328
Cf52	$[CAAA]_5$ $[AAC]_7$	KM589084
Cf53	$[ACAA]_7$	KF745017
Cf54	$[AAC]_{19}$	KM589085
Cf55	$[TTG]_5$	KF745018
Cf56	$[TTTTG]_5$	KF745019
Cf57	$[AGAC]_5$ $[GACG]_6$	KM589086
Cf58	$[GCA]_9$ 15 $[AAC]_8$	KM589087
Cf59	$[AGTGC]_7$ 5 $[GGTAG]_9$	KM589088
Cf60	$[GAT]_{11}$ 60 $[GAT]_8$	KM589089
Cf61	$[TTAGT]_6$	KM589090
Cf62	$[GTCG]_8$	KF745020
Cf63	$[ATTG]_{21}$	KM589091
Cf64	$[CCATG]_7$ 15 $[ACCAC]_{14}$	KF745021
Cf65	$[AAC]_8$	KF745022
Cf66	$[AAC]_{11}$	KM589092
Cf67	$[TTAC]_6$	KM589093
Cf68	$[TGTC]_5$	KM589094
Cf69	$[CGTC]_9$	KF745023
Cf70	$[ATG]_9$	KM589095
Cf71	$[GTT]_6$	KF745024
Cf72	$[AAC]_5$	KF745025
Cf73	$[ACG]_6$	KM589096
Cf74	$[CACC]_9$	KF745026
Cf75	$[AAAC]_6$	KF745027
Cf76	$[TC]_{12}$	KF730329
Cf77	$[TGT]_{12}$	KF745028
Cf78	$[GACA]_5$	KM589097
Cf79	$[AGC]_7$	KF745029
Cf80	$[CTGT]_{11}$	KM589098
Cf81	$[GTCC]_7$	KF745030
Cf82	$[AAC]_5$	KF745031
Cf83	$[CGGA]_{10}$	KF745032
Cf84	$[CGGA]_6$	KM589099
Cf85	$[GTCC]_9$	KM589100
Cf86	$[TGT]_{11}$	KF745033
Cf87	$[ACA]_7$	KF745034
Cf88	$[AAG]_6$	KM589101
Cf89	$[TGG]_7$	KF745035
Cf90	$[ACC]_7$	KM589102
Cf91	$[TTG]_{12}$	KF745036
Cf92	$[GGAC]_9$	KM589103
Cf93	$[AGG]_8$	KF745037
Cf94	$[GTT]_9$	KM589104
Cf95	$[CGTC]_8$	KM589105
Cf96	$[GTG]_5$	KM589106
Cf97	$[TTTTA]_5$	KF745038
Cf98	$[TTTTC]_4$	KM589107
Cf99	$[TATTG]_3$	KM589108
Cf100	$[AAC]_{13}$	KM589109
Cf101	$[ATCT]_5$	KM589110
Cf102	$[TTTTC]_5$	KM589111
Cf103	$[ATGC]_4$	KM589112
Cf104	$[TAAAT]_3$	KM589113

Cf105	$[CAAAA]_3$	KM589114
Cf106	$[AAC]_7$	KM589115
Cf107	$[TAAA]_4$	KM589116
Cf108	$[TCCG]_8$	KM589117
Cf109	$[CGTC]_4$	KM589118
Cf110	$[GTGTG]_{11}$	KM589119
Cf111	$[TTGA]_{27}$	KM589120
Cf112	$[AAGTA]_4$	KM589121
Cf113	$[GTGTG]_4$	KM589122
Cf114	$[AAAAT]_3$	KM589123
Cf115	$[CAA]_6$	KM589124
Cf116	$[CAAC]_4$	KM589125
Cf117	$[CACCA]_4$	KM589126
Cf118	$[ACGTC]_3$	KM589127
Cf119	$[TAAAA]_4$	KM589128
Cf120	$[CGGAT]_3$	KM589129
Cf121	$[ATCT]_4$	KM589130
Cf122	$[GTCC]_4$	KM589131
Cf123	$[AGAAC]_5$	KM589132
Cf124	$[AAAC]_4$	KM589133
Cf125	$[GGTGT]_{13}$ $[TGGTA]_{21}$	KM589134
Cf126	$[AAC]_5$	KM589135
Cf127	$[GTTCC]_3$	KM589136
Cf128	$[CGGAA]_3$	KM589137
Cf129	$[ATACT]_6$ $[ACACC]_{30}$ 18 $[ATACC]_{15}$	KM589138
Cf130	$[CGTC]_5$	KM589139
Cf131	$[TCTC]_{15}$ $[TA]_{14}$	KM589140
Cf132	$[AAC]_{10}$	KM589141
Cf133	$[ACAA]_8$	KM589142
Cf134	$[GTCG]_9$	KM589143
Cf135	$[TTGAT]_4$	KM589144
Cf136	$[CGCC]_4$	KM589145
Cf137	$[CGT]_9$	KM589146
Cf138	$[TTAC]_5$	KM589147
Cf139	$[CCATA]_{12}$ $[ACCAC]_{25}$ 16 $[ATACC]_5$ 87 $[TATAC]_{10}$ 10 $[ACCAC]_{15}$	KM589148
Cf140	$[CCGA]_6$	KM589149
Cf141	$[ATT]_6$	KM589150
Cf142	$[GTC]_{13}$ $[CAT]_{17}$ $[CAA]_9$	KM589151
Cf143	$[CCCCA]_3$	KM589152
Cf144	$[ATTG]_{35}$	KM589153
Cf145	$[GTCC]_6$	KM589154
Cf146	$[TATTC]_5$	KM589155
Cf147	$[GAACA]_3$	KM589156
Cf148	$[ACT]_{88}$	KM589157
Cf149	$[GTGTG]_{14}$	KM589158
Cf150	$[TCCG]_6$ 12 $[ATA]_{13}$ 3 $[AGA]_{11}$	KM589159
Cf151	$[TGA]_{43}$	KM589160
Cf152	$[GGGGT]_3$	KM589161
Cf153	$[CAA]_7$	KM589162
Cf154	$[GTTCC]_3$	KM589163
Cf155	$[GAACA]_3$	KM589164
Cf156	$[GGAC]_{25}$	KM589165
Cf157	$[GAT]_{12}$	KM589166
Cf158	$[CGTG]_4$	KM589167
Cf159	$[CGTG]_4$ $[CGTC]_4$	KM589168

Cf160	$[ATTG]_6$	KM589169
Cf161	$[GGTAC]_{11}$	KM589170
Cf162	$[ACCAC]_{36}$	KM589171
Cf163	$[ATTT]_4$	KM589172
Cf164	$[GTCG]_7$	KM589173
Cf165	$[GCGGG]_3$	KM589174
Cf166	$[TTA]_9$	KM589175
Cf167	$[TCAT]_4$	KM589176
Cf168	$[TATGG]_{13}$	KM589177
Cf169	$[AAC]_{27}$	KM589178
Cf170	$[TGAA]_6$	KM589179
Cf171	$[ACCAC]_{14}$ 24 $[CCATA]_5$	KM589180
Cf172	$[TCTG]_4$ 30 $[ACGC]_4$	KM589181
Cf173	$[ACCAC]_4$ 36 $[CCATA]_7$ 61 $[ATACT]_5$	KM589182
Cf174	$[ATT]_9$	KM589183
Cf175	$[TGTTT]_4$	KM589184
Cf176	$[CAT]_{38}$	KM589185
Cf177	$[CGTC]_4$	KM589186
Cf178	$[CGTC]_{15}$	KM589187
Cf179	$[ACCAC]_8$ $[ACTAT]_{21}$ 6 $[CATAC]_{15}$ $[ACCAC]_{14}$	KM589188
Cf180	$[GTGTG]_3$ 68 $[GTGTG]_3$ 50 $[GTGTG]_7$	KM589189
Cf181	$[GTCC]_8$	KM589190
Cf182	$[CGATC]_3$	KM589191
Cf183	$[ATACC]_{21}$	KM589192
Cf184	$[GGAC]_8$	KM589193
Cf185	$[CCGC]_4$	KM589194
Cf186	$[AAC]_6$	KM589195
Cf187	$[GGTGG]_3$	KM589196
Cf188	$[AAAC]_6$	KM589197
Cf189	$[TTTG]_5$	KM589198
Cf190	$[GTT]_6$	KM589199
Cf191	$[CGTC]_{12}$	KM589200
Cf192	$[GGAC]_6$	KM589201
Cf193	$[ACCAC]_4$ 24 $[CCATA]_5$ 9 $[ACTAT]_7$	KM589202
Cf194	$[TTC]_6$ $[TTA]_7$	KM589203
Cf195	$[ATA]_6$	KM589204
Cf196	$[AATAG]_{17}$	KM589205
Cf197	$[AAAG]_7$	KM589206
Cf198	$[ATCCA]_9$ $[ATCCG]_4$	KM589207
Cf199	$[CATAC]_{12}$	KM589208
Cf200	$[TGGTA]_3$	KM589209
Cf201	$[CATAC]_7$ $[TATAC]_12$	KM589210
Cf202	$[GTGTG]_3$	KM589211
Cf203	$[CAAT]_5$	KM589212
Cf204	$[TATAC]_{10}$ 10 $[ACCAC]_{22}$ $[ACCAT]_{11}$	KM589213
Cf205	$[TTTTG]_5$	KM589214
Cf206	$[ACGG]_{10}$	KM589215
Cf207	$[ATTG]_{14}$	KM589216
Cf208	$[TAA]_{13}$	KM589217
Cf209	$[TCAT]_{11}$ $[CAA]_5$	KM589218
Cf210	$[CCATA]_{33}$	KM589219
Cf211	$[TAA]_5$	KM589220
Cf212	$[TGAAA]_3$	KM589221
Cf213	$[CCGT]_7$	KM589222
Cf214	$[TTAC]_4$	KM589223

Cf215	$[ACT]_{32}$	KM589224
Cf216	$[TTTG]_{5}$	KM589225
Cf217	$[AAT]_{11}$	KM589226
Cf218	$[TCAT]_{27}$	KM589227
Cf219	$[TGTGG]_{3}$ 4 $[TATAG]_{6}$ 29 $[TGGTG]_{15}$	KM589228
Cf220	$[GTCC]_{25}$	KM589229
Cf221	$[ATACCA]_{10}$ 5 $[ACCAC]_{19}$	KM589230
Cf222	$[CAT]_{7}$	KM589231
Cf223	$[TGTGG]_{3}$	KM589232
Cf224	$[ATG]_{12}$	KM589233
Cf225	$[AGAC]_{7}$	KM589234
Cf226	$[GTGTG]_{5}$	KM589235
Cf227	$[ATG]_{5}$	KM589236
Cf228	$[GTGTG]_{3}$	KM589237
Cf229	$[CCATA]_{20}$	KM589238
Cf230	$[TGTA]_{4}$	KM589239
Cf231	$[GGTAT]_{6}$ 57 $TAGTA]_{6}$	KM589240
Cf232	$[CGTA]_{4}$	KM589241
Cf233	$[AG]_{32}$	KM589242
Cf234	$[GTCC]_{22}$	KM589243
Cf235	$[CCGT]_{12}$	KM589244
Cf236	$[ATT]_{8}$	KM589245
Cf237	$[TGT]_{5}$	KM589246
Cf238	$[CCATA]_{6}$ $[ATACT]_{9}$ 25 $[ACCAC]_{5}$	KM589247
Cf239	$[ATT]_{7}$	KM589248
Cf240	$[GTGTG]_{3}$	KM589249
Cf241	$[GTTTT]_{3}$	KM589250
Cf242	$[ATTG]_{9}$	KM589251
Cf243	$[ATT]_{9}$	KM589252
Cf244	$[GAT]_{31}$ 10 $[TGA]_{5}$ 83 $[GAT]_{23}$ 47 $[GAT]_{10}$ 44 $[GAT]_{25}$	KM589253
CF245	$[GTATA]_{31}$	KM589254
Cf246	$[AAAAT]_{6}$	KM589255

Table S2. Size in base pairs, forward (F) and reverse (R) primers and PCR amplification results for 97 SSRs in C. *fluminea*.

Locus	Size	Primer Sequences (5' → 3')	PCR amplification
Cf1	432	F: GTTCGGTGGACTGGTTTGAT R: CATGGAGAATGACGCCAAAT	Positive
Cf2	29	F: GCAGTCGTTGTGTTGTTGCT R: GGGCGATTAGGACCTTTGA	Positive
Cf3	269	F: TCGCCCTTTCTGGTCATATC R: GTGATGCAGGGGAAAACAAT	Positive
Cf4	136	F: AAATCAGCAATCCGTTAATTTGA R: TTTACCCGTCGCAGTTCAAT	Positive
Cf5	386	F: GATGCTCGGAGTGTGTTCAA R: GCGGTGACAATCAACAGATG	Negative
Cf6	149	F: GCGTTCACGTCTTACACACC R: TGTTCGTTTGTCACGTTCTG	Negative
Cf7	372	F: CCGCGGTATAACTGATGGTT R: CCGCGATATTTCGACACTTT	Negative
Cf8	298	F: TGGACTGGTTTGATGTCGAT R: TGACGCCCAATAACAATACG	Positive
Cf9	217	F: GTTCGGTGGACTGGTTTGAC R: GCGCGGATATGTTATTTTGC	Positive
Cf10	139	F: CATATCCACTAGCCTGACGAA R: AAGGGAACAAGATATGTTTGGA	Negative
Cf11	279	F: AGGGAATACAGCCGAACTGA R: CCAGCTTTGTTTTCAGTTTTTCA	Negative
Cf12	135	F: AGCTTGTGTAATAGTGGCAGTG R: TGATGGATGTGTGTAATGGTTAGA	Positive
Cf13	394	F: AGATAACACGTGCGCTACCC R: GCTGAAGACGGCCCTTATTA	Negative
Cf14	129	F: ATTGTTTTCCCCTGCATCAC R: GAGGGAAATATGGCCCATTG	Positive
Cf15	419	F: TCGGGAAATATGATGCGAATA R: GTCGTCGACGTCATCATCA	Negative
Cf16	100	F: TCTTATGTAGATGCGGGAGCTT R: ACAGCACGTACTTGTTTAGCC	Positive
Cf17	260	F: CAGTTAAACGCATCGGAACC R: TAGCATCACCACCATCATCG	Positive
Cf18	257	F: AAAATTGTGTCCGTCCCAAA R: TTTTCGTTTTTGAGGCTTGA	Negative
Cf19	150	F: ATGCGTGAAACGTTGAATGA R: TTACGCGTTATCCGGAGTTT	Positive
Cf20	136	F: TGCCTTTTGATTTACCGTAACC R: TGTGAAACACGGCTTACTGC	Positive
Cf21	300	F: AGCATGCAAAACAGTTACATCA R: TGGTCCTCTTAAAAAGTACAGATGG	Positive
Cf22	200	F: CTTTGCCAAATATGCCATCC R: CATAATGTGGCCTCCTTTGA	Positive
Cf23	263	F: AGATGCCGATTATTTCATACAAC R: GCGTTACAGTAAAAGATGGTTGA	Positive
Cf24	298	F: TTTAATTTTGTTGTTTGTACCAGTTG R: TGCCCAAGACCAAATCAAAT	Negative
Cf25	421	F: CCCGACCCGTATACATTGTT R: GCATAATATGGCCTGGCAGT	Positive
Cf26	146	F: TGTCACTTATCAAAATGTGTCAAAAA R: CATCGTGGTGTCGAGATGTT	Positive
Cf27	388	F: ACATCCATATCCCGTGGAGT R: ATCAGCGACCCGTTAATTTG	Negative
Cf28	118	F: TTGATGCTGCTTCTGCTGTT R: AAACAAAACATTCGCGGAAC	Positive
Cf29	119	F: CCTCATGTTCCTCGAAAAGC R: TTCTATTCGTTTGGAATGGATCT	Negative

Cf30	221	F: CCGAGAATGTCGGTAGCATT R: GCATGGTCGCGAAAACTAAT	Positive
Cf31	450	F: GTCAACAATGGACCACACCA R: GAAACCCGTGCGTATGTTTT	Positive
Cf32	149	F: TTTATTCTTGGGAAGACAAGATTC R: TGTTGCTGTTCATTCTGCAA	Negative
Cf33	260	F: AAAAGCAGAAACAACAGCATCA R: TGCTTCGCTACTTCAACTGC	Negative
Cf34	132	F: CATGGCCAAAGTAATCCAACT R: CCAGTTTTTCGGCATCAAAT	Negative
Cf35	311	F: GCGATGGGTTCAGAGACAGT R: GGCATCCAACTCAAGCAAAT	Positive
Cf36	330	F: AATTTCATGTCCCCTGCAAA R: GACCAAAATTGGCCACGTC	Positive
Cf37	420	F: ATCGGAACAAGCATCAACAT R: GGGGGTAAATGTGAAAGTCG	Positive
Cf38	269	F: GCAACAGCAGAAACGAACCT R: TATTCAGCAGCAGCATCGTC	Positive
Cf39	115	F: GACGATGCTGCTGCTGAATA R: TGTCCTTAGGCTGTTTGCTTG	Positive
Cf40	124	F: TTCTCAGTGCATGCAATTGTT R: GGAAAGCCTTCGTCCAATTT	Positive
Cf41	380	F: CCTTCACAGGTAACCCCATTT R: AGGCCGAATGCTACACTGAG	Negative
Cf42	225	F: TGCAGTAACCACAGCGTCAT R: CAAAGAAAACAAGAAAACAAAAAGG	Negative
Cf43	141	F: TGATGATCATAGCAATCATACTCCT R: GGCTATGTTGATGCTGCTGA	Positive
Cf44	127	F: TGAAATAGTCTGCCGATGTG R: TGATAACTGCGATGTTTTTCG	Positive
Cf45	266	F: TACCCCAGCCAGACACCTTA R: CCCCACACATATATTCTATTCCA	Positive
Cf46	448	F: AACCACAGATTCCATCCACA R: CGGGCTGATTCATTATGTGA	Positive
Cf47	137	F: TGGATTAATTGGTTCCCGTTT R: CGATCGTGACATTCAACACC	Positive
Cf48	235	F: ATTGTCGGCGGTTAAAGATG R: ACCCAGCGTTCACGTCTTAC	Positive
Cf49	149	F: TAAAGGCCCCGTAATCTCTG R: AAGCCAAAGGGCTAGAGGTC	Positive
Cf50	202	F: AGTGCGATTTCATGGTCTGA R: TGCATGCCTTTTAATTTTTGG	Positive
Cf51	363	F: CGTTTCCTGGTCGTTGTTTT R: GAGGGAGATGTGGAAGACGA	Positive
Cf52	251	F: ATAACGTTCGGGCATTGTTC R: TGTGTATGCTGCATTGGAAAA	Negative
Cf53	132	F: CGTAATGCTGGTTTCCCATC R: GATGACAAATGGCGGAAATC	Positive
Cf54	214	F: AAAACGACGCCATCAATCAT R: TCCCGTTTCCTTTTTCCTTT	Negative
Cf55	313	F: TCTCACGAGGATGGGAGTCT R: AATACGCCGGTTATGTTTGG	Positive
Cf56	425	F: TCAGCAGTAAGCGCAGAAGA R: CAGTTCTTCGTCGGTGTCAA	Positive
Cf57	268	F: CCCTGAGGGGTTTGACATTA R: GCTGCTGCTGCGTCTACATA	Negative
Cf58	267	F: TGTTGTCTTTGAAGTTGAAGCA R: GACGCCCACCTAGCCTATTA	Negative
Cf59	406	F: TGAGGTGGACATTGCCACTA R: GAAGGCGACACAAGTGTTCA	Negative
Cf60	251	F: GCCTGCCAAAACTTCTTCAT R: TTGTAACAAAACGGCAAAATACA	Negative
Cf61	143	F: ATTCCTCGCCATTTTTGATG	Negative

		R: GCGTGCGTTTTCATTGTCT	
Cf62	139	F: GGCACTGCCTCTTATCAAGC R: CCGGAAACGATGTGGTCTAC	Positive
Cf63	364	F: AAGCGAAACTTTTGCGGTTA R: TCCATAGTCAAGGGCACAAA	Negative
cF64	224	F: CTTGGGTTGCTTGTGTTGAG R: CCCCTACACTCCTCCAAGTT	Positive
Cf65	329	F: CGAGTGTGTATGCCGAGAAC R: CAGTGTGGTTTGTTGTGTGG	Positive
Cf66	302	F: TTTGGACGTACAGGCATTCA R: AAAACTGCCCTGTGACCCTA	Negative
Cf67	349	F: TCAAGTAGTGCGTCCAAACG R: AGCCCTACCCTTCTCCACAC	Negative
Cf68	444	F: CTGTTTGAGCTCGCTTTTGA R: GGGGCAATATACGACCAATG	Negative
Cf69	320	F: ATCGGTGGGTCTGTCAGTCT R: GGTAAAAATGGTCGCAGGAA	Positive
Cf70	200	F: GATGACATTTCACCCAGCAA R: CCACAGCTTCTTCTTCTACCA	Negative
Cf71	226	F: TCCAAGTCACAGTCATGTTTCC R: GGTCGTTGATCACGAGCATT	Positive
Cf72	298	F: GGCGCAACAAAGCAAATATC R: TGTTTCTTCGGAAAGCTGAA	Positive
Cf73	301	F: AATGTGGTCATGGTCGGAGT R: TTCATTTCTCATCATGTCTCACG	Negative
Cf74	117	F: TGTTGAAAATCGGGAGACAA R: GGATGGAGAGATATCGGGTTG	Positive
Cf75	277	F: TTGAAACCAGGCGTTAAACC R: CGACAATTCGACTTCCCAAT	Positive
Cf76	205	F: AGACAAGGCGTTCGCTTTTA R: CACCATCGTGATGATGAACG	Positive
Cf77	221	F: TCGCTGGTTTTTAATGAAAGG R: CAAAACAATTCTCCACCTGACA	Positive
Cf78	308	F: ACAGATGGACGGACAAATCC R: TCTGACAAAACTTCCCACTGAA	Negative
Cf79	229	F: TTTTTGTGGTGATGGTGGTG R: TGCTGTTGCTAAAATGCTTCTC	Positive
Cf80	207	F: TGGTGTGGGAGACATTTTGA R: GGGCCATAATTCTGACATGC	Negative
Cf81	277	F: AGCGACGCGAGTCATAAAAT R: CAAGGTAGCGACCCAAATGT	Positive
Cf82	210	F: AGACGAGGGATGATCACAAAA R: CACACTGAATCGATGGCAATA	Positive
Cf83	150	F: GGAGTTGCGTTCACGAGAATA R: TGTGTTGGTGGCATACAATG	Positive
Cf84	135	F: CATAAACCCAAAATAAGCCACA R: CAAATGCCCATCTGTGAGAA	Negative
Cf85	103	F: CCTTTCAAAGAAAAGGCGATA R: TGGCACAGTTATAGCCAACG	Negative
Cf86	125	F: AGGCAGGATGCGTAGATTGA R: CGAGGTGTAGCCAACACAAG	Positive
Cf87	125	F: GGACGGGTTTTAGCCATACA R: TGAAATATCGGTCCCATACAAA	Positive
Cf88	212	F: GCCATTAGAAGTTTACAAGTCAGAA R: TGGGACAATTGCACAAGTAAA	Negative
Cf89	147	F: CAGTAGCAGCAGCAGTGGAG R: CCTCCTCCACATAAAACTTTGC	Positive
Cf90	131	F: TGCATGCAAGCTACTCATAACT R: TGTGGTGTGTGGTGTGGTCT	Negative
Cf91	132	F: AAACTGCCTCTGGTCGAAAA R: TCGCAGTTTGCTTGTACGTC	Positive
Cf92	106	F: GGAGGAGTTGATGCCACAAG R: TTTTTATGCCCCCACAAAGT	Negative

Cf93	130	F: GTCGTGATGCCCTCATCTTT R: ACCCAGATGTTCCGTACCAC	Positive
Cf94	106	F: TACGCCTGTCTTCGACCTTT R: TGGAACAGAGTTGATTCGTCA	Negative
Cf95	100	F: AGCGAGGAGGGGGATATAGG R: AAAGTTATGGACCGGACACG	Negative
Cf96	100	F: CCAAGTTTGCCCTGTGAGAT R: CCATATGGTATTGCTGGGAAA	Negative
Cf97	217	F: TAGGGATTATCGGGACGATG R: ATGCCAGATTTAGGGCATTG	Positive

8.3

Supporting information from Chapter 3:

SNP identification in two invasive species: zebra mussel (*Dreissena polymorpha*) and Asian clam (*Corbicula fluminea*)

8.4

Supporting information from Chapter 4:

Genetic characterization of the invasive zebra mussel (*Dreissena polymorpha*) in the Iberian Peninsula

Table S1. Microsatellite primer sequences, concentrations of reagents and thermal profiles used for PCR amplifications. *F* = Forward PCR primer; *R* = Reverse PCR primer. *Standard final concentrations for PCR mix reagents in a final volume of 30 µl per sample.

PCR reaction	Microsatellite	*F* / *R* final concentration (µM)*	Thermal profile
PCR – Multiplex 1	DpolC5	0.2	Initial denaturation step 94°C 3’ Cycling step 10 TOUCH-UP cycles → 94°C 30”; 50-60°C 2’ (increasing 1°C every cycle); 72°C 2’ 25 cycles → 94°C 30”; 56°C 2’; 72°C 2’ Final extension step 72°C for 10’
	DpolB9	0.2	
	DpolB6	0.2	
	DpolA6	0.2	
	DpolB8	0.2	
PCR – Multiplex 2	Dp31	0.267	Initial denaturation step 94°C 3’ Cycling step 12 TOUCH-DOWN cycles → 94°C 30”; 62-50°C 2’ (decreasing 1°C every cycle); 72°C 2’ 25 cycles → 94°C 30”; 50°C 2’; 72°C 2’ Final extension step 72°C for 10’
	Dp43	0.267	
	Dp68	0.4	
	Dp86	0.2	

Table S2. **Bayesian analyses STRUCTURE clusters composition among locations**

		Using all 21 locations							Using only 3 Spanish locations						
		k = 2 (%)		*k* = 5 (%)					*k* = 2 (%)		*k* = 5 (%)				
Country	Location	1	2	1	2	3	4	5	1	2	1	2	3	4	5
Romania	1	90.6	09.4	79.5	03.4	03.5	09.3	04.3	41.5	58.5	72.1	05.1	07.4	12.3	03.2
Spain	2	02.8	97.2	02.3	50.1	43.7	02.8	01.1	-	-	-	-	-	-	-
	3	01.6	98.4	01.8	61.6	34.4	01.4	00.8	-	-	-	-	-	-	-
	4	03.9	96.1	02.0	59.1	33.6	03.0	02.3	-	-	-	-	-	-	-
	5	02.4	97.6	01.6	61.7	33.4	02.1	01.1	-	-	-	-	-	-	-
	6	01.4	98.6	02.2	58.4	37.3	01.4	00.8	-	-	-	-	-	-	-
	7	03.4	96.6	03.2	55.1	38.4	02.5	00.8	-	-	-	-	-	-	-
	8	02.9	97.1	02.8	46.4	47.9	01.8	01.1	-	-	-	-	-	-	-
	9	02.6	97.4	01.4	41.9	53.3	01.8	01.6	-	-	-	-	-	-	-
	10	03.5	96.5	03.1	45.4	48.4	02.1	01.0	95.4	04.6	03.2	88.9	03.1	03.5	01.3
	11	09.9	90.1	05.5	28.6	58.3	04.2	03.3	93.3	06.7	02.5	87.5	02.7	04.3	03.1
	12	08.3	91.7	02.8	24.1	65.9	04.2	03.1	92.0	08.0	02.4	86.2	03.7	04.8	02.9
Italy	13	92.3	07.7	17.7	05.0	01.9	73.5	01.8	19.2	80.8	07.5	04.2	61.1	25.5	01.7
	14	94.2	05.8	01.2	02.3	01.4	93.8	01.3	13.0	87.0	00.9	02.1	92.9	02.8	01.2
France	15	91.6	08.4	50.7	04.4	05.2	30.8	08.9	28.2	71.8	18.1	06.7	20.6	48.4	06.2
	16	91.8	08.2	51.3	06.5	04.6	26.2	11.4	28.2	71.8	15.0	06.9	17.8	51.0	09.3
UK	17	89.2	10.8	46.9	06.2	04.5	33.1	09.3	29.1	70.9	12.7	06.4	16.7	58.7	05.5
	18	85.3	14.7	53.6	09.1	05.9	23.3	08.1	27.0	73.0	15.3	09.3	13.1	57.5	04.8
	19	90.8	09.2	36.5	05.1	07.0	38.5	12.9	25.1	74.9	04.8	07.2	23.2	56.7	08.1
North America	20	96.4	03.6	09.4	02.8	02.8	05.9	79.1	10.4	89.6	10.2	03.3	05.3	09.7	71.5
	21	98.8	01.2	08.0	01.3	01.3	04.7	84.8	05.3	94.7	07.0	01.5	05.2	10.1	76.2

Table S3. Results of GeneClass assignment test. In bold: the most probable source for each population. Refer to Table 1 for sample numbers.

Country		Romania						Spain						Italy		France		Great Britain			North America	
Pop		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	Assigned to	**93.75**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.08	0.00	0.00	2.08	2.08	0.00	0.00	0.00	0.00
	Rejected from	0.00	100.00	100.00	97.92	100.00	87.50	93.75	100.00	97.92	97.92	91.67	95.83	85.42	100.00	37.50	41.67	39.58	66.67	87.50	31.25	68.75
2	Assigned to	0.00	**45.00**	8.33	8.33	8.33	8.33	6.67	3.33	1.67	6.67	0.00	1.67	0.00	0.00	0.00	0.00	0.00	1.67	0.00	0.00	0.00
	Rejected from	53.33	0.00	5.00	18.33	16.67	1.67	1.67	1.67	0.00	5.00	20.00	3.33	58.33	100.00	35.00	28.33	10.00	58.33	58.33	20.00	73.33
3	Assigned to	0.00	3.92	**50.98**	15.69	7.84	0.00	3.92	3.92	3.92	1.96	3.92	3.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Rejected from	56.86	9.80	0.00	5.88	5.88	1.96	0.00	1.96	0.00	3.92	13.73	1.96	72.55	100.00	39.22	76.47	21.57	62.75	66.67	27.45	78.43
4	Assigned to	0.00	3.85	5.77	**59.62**	3.85	3.85	5.77	9.62	1.92	3.85	0.00	1.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Rejected from	55.77	15.38	3.85	1.92	15.38	3.85	3.85	7.69	3.85	15.38	13.46	7.69	71.15	98.08	34.62	17.31	11.54	65.38	63.46	19.23	71.15
5	Assigned to	0.00	11.76	7.84	7.84	**54.90**	3.92	5.88	1.96	1.96	1.96	1.96	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Rejected from	47.06	7.84	1.96	5.88	0.00	0.00	0.00	0.00	3.92	7.84	21.57	3.92	68.63	100.00	33.33	25.49	15.69	58.82	62.75	7.84	60.78
6	Assigned to	0.00	15.38	7.69	0.00	15.38	**38.46**	7.69	0.00	7.69	0.00	0.00	7.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Rejected from	38.46	0.00	0.00	23.08	7.69	0.00	0.00	0.00	0.00	0.00	30.77	0.00	69.23	100.00	38.46	23.08	15.38	61.54	76.92	15.38	76.92
7	Assigned to	0.00	3.80	11.39	6.33	7.59	5.06	**43.04**	2.53	1.27	7.59	3.80	5.06	0.00	0.00	1.27	0.00	0.00	0.00	1.27	0.00	0.00
	Rejected from	41.77	11.39	7.59	11.39	15.19	5.06	0.00	3.80	2.53	11.39	11.39	5.06	63.29	98.73	32.91	24.05	11.39	48.10	54.43	22.78	70.89
8	Assigned to	0.00	6.45	4.84	11.29	8.06	1.61	6.45	**40.32**	6.45	6.45	0.00	8.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Rejected from	61.29	16.13	17.74	19.35	12.90	14.52	8.06	4.84	8.06	11.29	20.97	3.23	72.58	100.00	40.32	27.42	25.81	61.29	70.97	20.97	85.48
9	Assigned to	0.00	7.41	6.17	3.70	4.94	7.41	6.17	4.94	**45.68**	7.41	0.00	6.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Rejected from	67.90	23.46	20.99	34.57	19.75	6.17	9.88	9.88	0.00	16.05	24.69	11.11	75.31	100.00	48.15	46.91	32.10	70.37	75.31	20.99	88.89
10	Assigned to	0.00	3.33	8.33	1.67	6.67	1.67	5.00	6.67	8.33	**55.00**	1.67	1.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Rejected from	53.33	16.67	15.00	23.33	23.33	15.00	5.00	6.67	3.33	1.67	15.00	3.33	66.67	100.00	31.67	28.33	23.33	60.00	66.67	21.67	83.33
11	Assigned to	0.00	0.00	0.00	0.00	0.00	0.00	1.96	0.00	0.00	0.00	**94.12**	1.96	0.00	0.00	0.00	0.00	0.00	0.00	1.96	0.00	0.00
	Rejected from	45.10	35.29	29.41	37.25	47.06	19.61	9.80	29.41	13.73	37.25	0.00	13.73	64.71	100.00	27.45	21.57	15.69	49.02	56.86	7.84	66.67
12	Assigned to	0.00	1.92	1.92	3.85	1.92	5.77	1.92	3.85	7.69	5.77	3.85	**61.54**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Rejected from	57.69	32.69	30.77	34.62	48.08	15.38	7.69	21.15	7.69	26.92	17.31	1.92	69.23	98.08	28.85	38.46	26.92	63.46	59.62	11.54	65.38
13	Assigned to	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.23	0.00	0.00	**93.55**	0.00	0.00	3.23	0.00	0.00	0.00	0.00	0.00
	Rejected from	29.03	93.55	93.55	96.77	93.55	90.32	77.42	87.10	74.19	93.55	70.97	67.74	3.23	70.97	0.00	0.00	0.00	29.03	22.58	9.68	29.03
14	Assigned to	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**96.67**	1.67	0.00	0.00	0.00	1.67	0.00	0.00
	Rejected from	33.33	86.67	86.67	80.00	90.00	90.00	68.33	83.33	68.33	86.67	68.33	60.00	1.67	0.00	0.00	0.00	0.00	23.33	10.00	0.00	5.00
15	Assigned to	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.13	0.00	0.00	2.13	0.00	**74.47**	8.51	6.38	2.13	4.26	0.00	0.00
	Rejected from	27.66	89.36	85.11	85.11	93.62	76.60	78.72	80.85	80.85	87.23	78.72	74.47	38.30	93.62	0.00	4.26	0.00	34.04	38.30	10.64	23.40
16	Assigned to	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00	0.00	4.00	0.00	18.00	**60.00**	12.00	0.00	0.00	2.00	0.00
	Rejected from	44.00	96.00	96.00	94.00	96.00	90.00	90.00	94.00	94.00	96.00	84.00	82.00	54.00	98.00	14.00	0.00	10.00	48.00	52.00	12.00	42.00
17	Assigned to	0.00	0.00	0.00	1.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.67	1.67	0.00	8.33	3.33	**60.00**	11.67	11.67	0.00	0.00
	Rejected from	35.00	85.00	88.33	81.67	91.67	78.33	73.33	85.00	80.00	86.67	71.67	73.33	50.00	95.00	11.67	5.00	0.00	33.33	28.33	16.67	25.00
18	Assigned to	0.00	0.00	0.00	0.00	0.00	0.00	3.45	0.00	0.00	0.00	0.00	0.00	3.45	0.00	0.00	3.45	3.45	**72.41**	10.34	3.45	0.00
	Rejected from	27.59	93.10	82.76	86.21	86.21	79.31	62.07	75.86	68.97	82.76	72.41	65.52	24.14	96.55	6.90	0.00	3.45	0.00	27.59	6.90	24.14
19	Assigned to	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.57	3.57	0.00	0.00	0.00	7.14	10.71	**75.00**	0.00	0.00
	Rejected from	21.43	92.86	85.71	78.57	89.29	64.29	64.29	85.71	71.43	92.86	57.14	82.14	42.86	92.86	3.57	3.57	0.00	7.14	0.00	3.57	7.14
20	Assigned to	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.64	1.64	1.64	0.00	4.92	**77.05**	13.11
	Rejected from	86.89	98.36	100.00	100.00	100.00	96.72	100.00	98.36	96.72	98.36	98.36	98.36	86.89	100.00	63.93	63.93	54.10	85.25	86.89	0.00	6.56
21	Assigned to	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.79	0.00	8.93	**89.29**
	Rejected from	91.07	100.00	100.00	100.00	100.00	100.00	98.21	100.00	100.00	100.00	96.43	98.21	94.64	100.00	64.29	53.57	50.00	83.93	89.29	3.57	0.00

8.5

Supporting information from Chapter 5:

Genetic characterization of the Asian clam species complex (*Corbicula*) invasion in the Iberian Peninsula

Table S1. Haplotype and *Corbicula* lineage identified in the 175 individual analyzed in this study based on phylogenetic COI and 28S markers.

Location	Individual	COI gene Haplotype	Lineage	28S gene Haplotype	Lineage
CA	CA01	COI1	RA	S4/S5	RA/SC
	CA02	COI1	RA	S4/S5	RA/SC
	CA03	COI1	RA	S4/S5	RA/SC
	CA04	COI1	RA	S4/S5	RA/SC
	CA05	COI1	RA	S4/S5	RA/SC
	CA06	COI1	RA	S4/S5	RA/SC
	CA07	COI1	RA	S4/S5	RA/SC
	CA08	COI1	RA	S4/S5	RA/SC
	CA09	COI1	RA	S4/S5	RA/SC
	CA10	COI1	RA	S4/S5	RA/SC
	CA11	COI3	RA	S4/S5	RA/SC
	CA12	COI1	RA	S4/S5	RA/SC
	CA13	COI1	RA	S4/S5	RA/SC
GA	GA01	COI1	RA	S4/S5	RA/SC
	GA02	COI1	RA	S4/S5	RA/SC
	GA03	COI1	RA	S4/S5	RA/SC
	GA04	COI1	RA	S4/S5	RA/SC
	GA05	COI1	RA	S4/S5	RA/SC
	GA06	COI1	RA	S4/S5	RA/SC
	GA07	COI1	RA	S4/S5	RA/SC
	GA08	COI1	RA	S4/S5	RA/SC
	GA09	COI1	RA	S4/S5	RA/SC
	GA10	COI1	RA	S4/S5	RA/SC
	GA11	COI1	RA	S4/S5	RA/SC
	GA12	COI1	RA	S4/S5	RA/SC
	GA13	COI1	RA	S4/S5	RA/SC
TE	TE01	COI1	RA	S3/S3	RA/RA
	TE02	COI1	RA	S4/S5	RA/SC
	TE03	COI1	RA	S4/S5	RA/SC
	TE04	COI1	RA	S4/S5	RA/SC
	TE05	COI1	RA	S4/S5	RA/SC
	TE06	COI1	RA	S4/S5	RA/SC
	TE07	COI5	RA	S4/S5	RA/SC
	TE08	COI1	RA	S4/S5	RA/SC
	TE09	COI1	RA	S3/S3	RA/RA
	TE10	COI1	RA	S4/S5	RA/SC
	TE11	COI6	RlcB	S3/S3	RA/RA
	TE12	COI1	RA	S4/S5	RA/SC
AL	AL01	COI1	RA	S1/S2	RA/RlcB
	AL02	COI1	RA	S4/S5	RA/SC
	AL03	COI1	RA	S1/S3	RA/RA
	AL04	COI1	RA	S1/S1	RA/RA
	AL05	COI1	RA	S4/ S5	RA/SC
	AL06	COI1	RA	S4/ S5	RA/SC
	AL07	COI1	RA	S4/ S5	RA/SC
	AL08	COI1	RA	S4/ S5	RA/SC
	AL09	COI1	RA	S4/ S5	RA/SC
	AL10	COI1	RA	S4/ S5	RA/SC
	AL11	COI1	RA	S4/ S5	RA/SC
	AL12	COI1	RA	S4/ S5	RA/SC
ZG	ZG01	COI1	RA	S4/S5	RA/SC
	ZG02	COI1	RA	S4/S5	RA/SC
	ZG03	COI1	RA	S4/S5	RA/SC
	ZG04	COI1	RA	S4/S5	RA/SC
	ZG05	COI1	RA	S4/S5	RA/SC
	ZG06	COI1	RA	S4/S5	RA/SC
	ZG07	COI1	RA	S4/S5	RA/SC

	ZG08	COI1	RA	S4/S5	RA/SC
	ZG09	COI1	RA	S4/S5	RA/SC
	ZG10	COI1	RA	S4/S5	RA/SC
	ZG11	COI1	RA	S4/S5	RA/SC
	ZG12	COI1	RA	S4/S5	RA/SC
RB	RB01	COI1	RA	S1/S5	RA/SC
	RB02	COI1	RA	S1/S5	RA/SC
	RB03	COI1	RA	S1/S5	RA/SC
	RB04	COI1	RA	S1/S5	RA/SC
	RB05	COI1	RA	S1/S3	RA/RA
	RB06	COI1	RA	S1/S5	RA/SC
	RB07	COI1	RA	S1/S5	RA/SC
	RB08	COI1	RA	S1/S5	RA/SC
	RB09	COI1	RA	S1/S5	RA/SC
	RB10	COI1	RA	S1/S5	RA/SC
	RB11	COI1	RA	S1/S5	RA/SC
	RB12	COI1	RA	S1/S5	RA/SC
	RB13	COI1	RA	S1/S5	RA/SC
XE	XE01	COI6	RlcB	S4/S9	RA/RlcB
	XE02	COI1	RA	S4/S5	RA/SC
	XE03	COI1	RA	S4/S5	RA/SC
	XE04	COI7	SC	S8/S10	SC/SC
	XE05	COI6	RlcB	S4/S9	RA/RlcB
	XE06	COI1	RA	S4/S5	RA/SC
	XE07	COI1	RA	S4/S5	RA/SC
	XE08	COI1	RA	S4/S5	RA/SC
	XE09	COI1	RA	S4/S5	RA/SC
	XE10	COI1	RA	S4/S5	RA/SC
	XE11	COI1	RA	S4/S5	RA/SC
	XE12	COI1	RA	S4/S5	RA/SC
DE	DE01	COI2	RlcB	S4/S7	RA/SC
	DE02	COI2	RlcB	S4/S7	RA/SC
	DE03	COI2	RlcB	S4/S7	RA/SC
	DE04	COI2	RlcB	S4/S7	RA/SC
	DE05	COI4	RlcB	S4/S7	RA/SC
	DE06	COI4	RlcB	S4/S7	RA/SC
	DE07	COI2	RlcB	S4/S7	RA/SC
	DE08	COI2	RlcB	S4/S7	RA/SC
	DE09	COI4	RlcB	S4/S7	RA/SC
	DE10	COI4	RlcB	S4/S7	RA/SC
	DE11	COI2	RlcB	S4/S7	RA/SC
	DE12	COI2	RlcB	S4/S7	RA/SC
	DE13	COI2	RlcB	S4/S7	RA/SC
	DE14	COI2	RlcB	S4/S7	RA/SC
	DE15	COI2	RlcB	S4/S7	RA/SC
	DE16	COI2	RlcB	S4/S7	RA/SC
RH	RH01	COI1	RA	S4/S5	RA/SC
	RH02	COI1	RA	S4/S5	RA/SC
	RH03	COI1	RA	S4/S5	RA/SC
	RH04	COI1	RA	S4/S5	RA/SC
	RH05	COI1	RA	S4/S5	RA/SC
	RH06	COI1	RA	S4/S5	RA/SC
	RH07	COI1	RA	S4/S5	RA/SC
	RH08	COI1	RA	S4/S5	RA/SC
	RH09	COI1	RA	S4/S5	RA/SC
	RH10	COI1	RA	S4/S5	RA/SC
	RH11	COI1	RA	S4/S5	RA/SC
	RH12	COI1	RA	S4/S5	RA/SC
RO	RO01	COI1	RA	S4/S5	RA/SC
	RO02	COI1	RA	S4/S5	RA/SC
	RO03	COI1	RA	S4/S5	RA/SC
	RO04	COI1	RA	S4/S5	RA/SC
	RO05	COI1	RA	S4/S5	RA/SC

	RO06	COI1	RA	S4/S5	RA/SC
	RO07	COI1	RA	S4/S5	RA/SC
	RO08	COI1	RA	S4/S5	RA/SC
	RO09	COI1	RA	S4/S5	RA/SC
	RO10	COI1	RA	S4/S5	RA/SC
	RO11	COI1	RA	S1/S2	RA/RlcB
BR	BR01	COI2	RlcB	S6/S6	RlcB/RlcB
	BR02	COI2	RlcB	S6/S6	RlcB/RlcB
	BR03	COI1	RA	S6/S6	RlcB/RlcB
	BR04	COI2	RlcB	S6/S6	RlcB/RlcB
	BR05	COI1	RA	S6/S6	RlcB/RlcB
	BR06	COI1	RA	S2/S3	RlcB/RA
	BR07	COI1	RA	S6/S6	RlcB/RlcB
	BR08	COI1	RA	S2/S3	RlcB/RA
	BR09	COI2	RlcB	S6/S6	RlcB/RlcB
	BR10	COI1	RA	S2/S3	RlcB/RA
	BR11	COI2	RlcB	S6/S6	RlcB/RlcB
	BR12	COI1	RA	S6/S6	RlcB/RlcB
GR	GR01	COI1	RA	S1/S8	RA/SC
	GR02	COI1	RA	S4/S5	RA/SC
	GR03	COI1	RA	S4/S5	RA/SC
	GR04	COI1	RA	S4/S5	RA/SC
	GR05	COI1	RA	S4/S5	RA/SC
	GR06	COI1	RA	S1/S8	RA/SC
	GR07	COI1	RA	S1/S8	RA/SC
	GR08	COI1	RA	S1/S8	RA/SC
	GR09	COI1	RA	S1/S9	RA/RlcB
	GR10	COI1	RA	S4/S5	RA/SC
	GR11	COI1	RA	S1/S3	RA/RA
	GR12	COI1	RA	S1/S2	RA/RlcB
SB	SB01	COI1	RA	S4/S5	RA/SC
	SB02	COI1	RA	S4/S5	RA/SC
	SB03	COI1	RA	S4/S5	RA/SC
	SB04	COI1	RA	S4/S5	RA/SC
	SB05	COI1	RA	S4/S5	RA/SC
	SB06	COI1	RA	S4/S5	RA/SC
	SB07	COI1	RA	S4/S5	RA/SC
	SB08	COI1	RA	S4/S5	RA/SC
	SB09	COI1	RA	S4/S5	RA/SC
	SB10	COI1	RA	S4/S5	RA/SC
	SB11	COI1	RA	S4/S5	RA/SC
	SB12	COI1	RA	S4/S5	RA/SC
	SB13	COI1	RA	S4/S5	RA/SC
TY	TY01	COI1	RA	S1/S2	RA/RlcB
	TY02	COI1	RA	S1/S2	RA/RlcB
	TY03	COI1	RA	S1/S3	RA/RA
	TY04	COI1	RA	S1/S2	RA/RlcB
	TY05	COI1	RA	S1/S3	RA/RA
	TY06	COI1	RA	S1/S3	RA/RA
	TY07	COI1	RA	S1/S2	RA/RlcB
	TY08	COI1	RA	S1/S2	RA/RlcB
	TY09	COI1	RA	S1/S2	RA/RlcB
	TY10	COI1	RA	S1/S3	RA/RA
	TY11	COI1	RA	S1/S3	RA/RA
	TY12	COI1	RA	S1/S2	RA/RlcB

Table S2. Pairwise genetic differentiation between sampling locations. Below Diagonal: COI gene; above diagonal: 28S gene. In Bold significant p-values with a 0.05 threshold after Bonferroni correction with 10000 replicates.

		CA	GA	TE	AL	ZG	RB	XE	DE	RH	RO	BR	GR	SB	TY
Iberian Peninsula	CA		-0.0400	-0.0092	-0.0118	-0.0417	**0.1349**	-0.0269	**0.3797**	-0.0417	-0.0382	**0.6136**	0.0068	-0.0400	**0.3914**
	GA	0		-0.0092	-0.0118	-0.0417	**0.1349**	-0.0269	**0.3797**	-0.0417	-0.0382	**0.6136**	0.0068	-0.0400	**0.3914**
	TE	0.0062	0.0069		-0.0269	-0.0109	0.0813	0.0074	**0.4161**	-0.0109	-0.0232	**0.5964**	0.0065	-0.0092	**0.3169**
	AL	-0.0065	0	0		-0.0135	0.0758	-0.0110	**0.3937**	-0.0135	-0.0344	**0.5583**	-0.0323	-0.0118	**0.2654**
	ZG	-0.0065	0	0	0		**0.1346**	-0.0286	**0.3764**	-0.0435	-0.0400	**0.6176**	0.0051	-0.0417	**0.3948**
	RB	0	0	0.0069	0	0		**0.1535**	**0.5204**	**0.1346**	**0.1051**	**0.6380**	**0.0926**	**0.1349**	**0.3332**
	XE	0.1456	0.1488	-0.0099	0.1383	0.1383	0.1488		**0.3474**	-0.0286	-0.0306	**0.5830**	-0.0121	-0.0269	**0.3554**
	DE	**0.9804**	**0.9850**	**0.8945**	**0.9844**	**0.9844**	**0.9850**	**0.7417**		**0.3764**	**0.3728**	**0.6546**	**0.3811**	**0.3797**	**0.5688**
France	RH	-0.0065	0	0	0	0	0	0.1383	**0.9844**		-0.0400	**0.6176**	0.0051	-0.0417	**0.3948**
Romania	RO	-0.0136	0	-0.0076	0	0	0	0.1272	**0.9838**	0		**0.5830**	-0.0199	-0.0382	**0.3366**
TX, USA	BR	**0.3718**	**0.3767**	0.1858	**0.3636**	**0.3636**	**0.3767**	0.0082	**0.5619**	**0.3636**	0.3498		**0.5442**	**0.6136**	**0.5130**
	GR	-0.0065	0	0	0	0	0	0.1383	**0.9844**	0	0	**0.3636**		0.0068	**0.2497**
	SB	0	0	0.0069	0	0	0	0.1488	**0.9850**	0	0	**0.3767**	0		**0.3914**
	TY	-0.0065	0	0	0	0	0	0.1383	**0.9844**	0	0	**0.3636**	0	0	

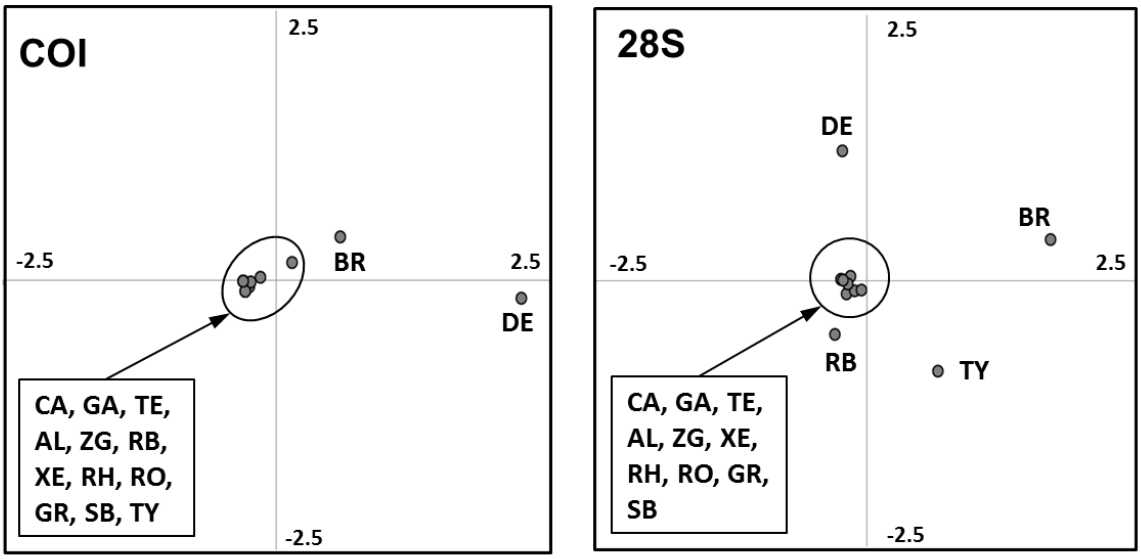


Figure S1. Multidimensional Scaling (MDS) analysis based on pairwise genetic differentiation (ΦST) between sampling locations for mt COI (left) and nuclear 28S (right) genes. Encircled locations corresponded to no significant pairwise ϕ_{ST} comparisons (P> 0.05) among sites after Bonferroni correction.

8.6

Supporting information from Chapter 6:

Validated methodology for quantifying infestation levels of dreissenid in environmental DNA (eDNA) samples

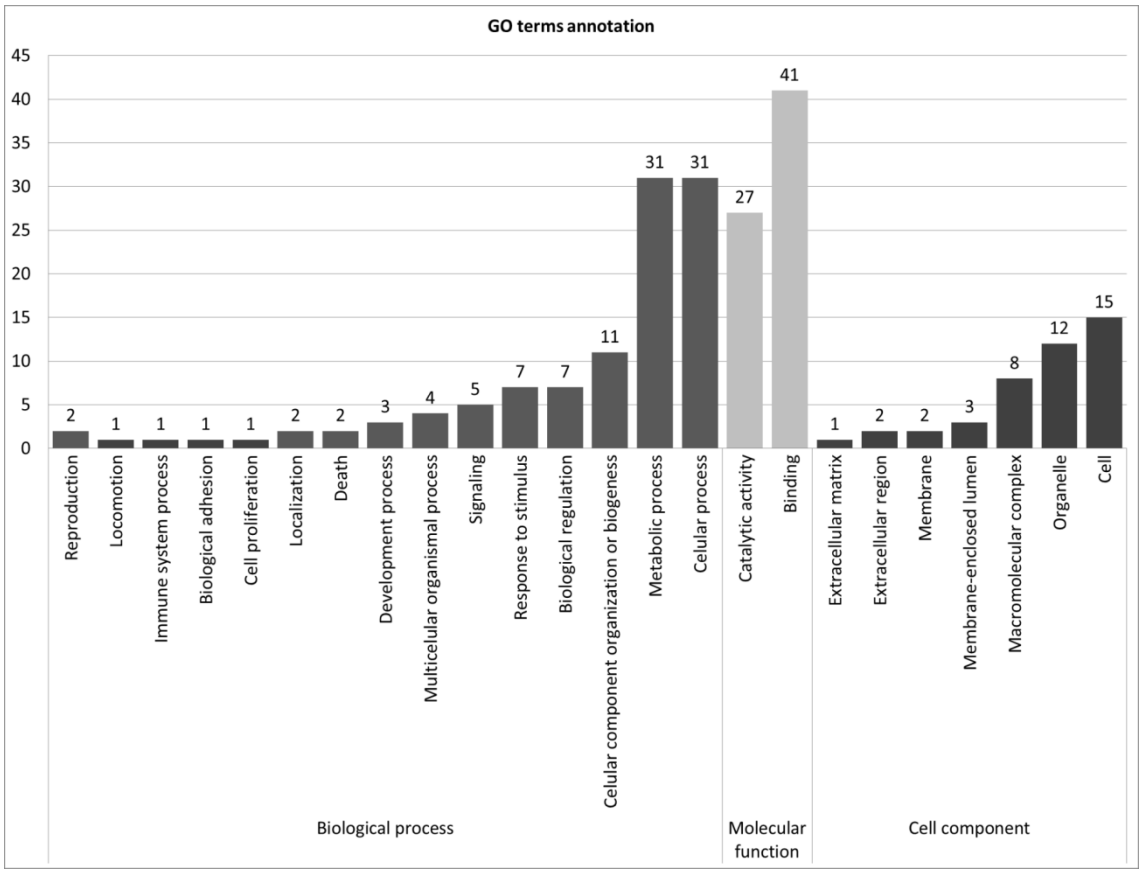


Figure S1. **Gene Ontology (GO) gene annotation for MPS sequences.**

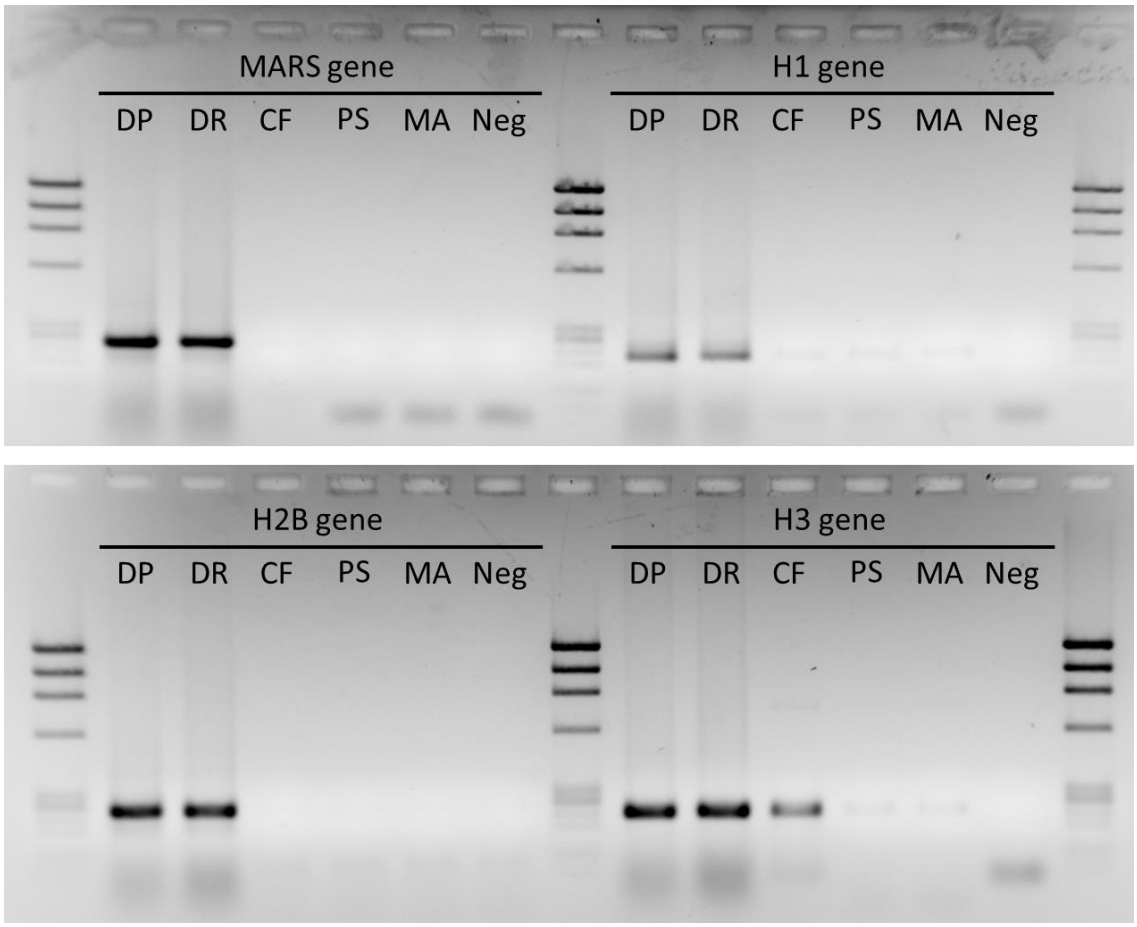


Figure S2. Agarose gel (1.5%) of representative adult individuals in the species specificity PCR. DP: *D. polymorpha*, DR: *D. rostriformis*, CF: *C. fluminea*, PS: *Pomacea sp*., MA*: M. auricularia*, Neg: negative PCR control.

```
                1   5          15         25         35         45         55
                |   |          |          |          |          |          |
Contig000076    CGCGCGCTCC ACTGACAAGA AGAAGCGCAG GAGGAGGAGG GAATCCTACG CCATCTACAT
DP adult tissue  .......... .......... .......... .......... .......... ..........
DR adult tissue  .......... .......... .......... .......... .....A.... ..........
Water samples .......... .......... .......... .......... .......... ..........

                    65         75         85         95        105        115
                    |          |          |          |          |          |
Contig000076    CTACAAAGTC TTGAAGCAGG TGCACCCCGA CACCGGAGTG TCCAGCAAGG CCATGTCGAT
DP adult tissue  .......... .......... .......... .......... .......... ..........
DR adult tissue  .......... C.C....... .......... .......... .......... .......C..
Water samples .......... .......... .......... .......... .......... ..........

                    125        135        145        155        165        175
                    |          |          |          |          |          |
Contig000076    CATGAACAGC TTTGTCAACG ACATCTTCGA GCGCATTGCT GCCGAGGCTT CCCGCCTTGC
DP adult tissue  .......... .......... .......... .......... .......... ..........
DR adult tissue  .......... .......... .......... ......C... .......... ..........
Water samples .......... .......... .......... .......... .......... ..........

                    185        195        205        215        225        235
                    |          |          |          |          |          |
Contig000076    CCACTACAAC AAGCGATCCA CCATCACAAG CAGAGAGATC CAGACCGCTG TGCGTCTCCT
DP adult tissue  .......... .......... .......... .......... .......... ..........
DB adult tissue  .......... .......... .......... .......... .......... ..........
Water samples .......... .......... .......... .......... .......... ..........

                    245    251
                    |      |
Contig000076    GCTGCCTGGT G
DP adult tissue  .......... .
DR adult tissue  .......... .
Water samples .......... .
```

Figure S3. Alignment of H2B sequences with the MPS sequence of reference. DP adult tissue represents all H2B sequences for the three *D. polymorpha* adult individuals. DR adult tissue represents H2B sequences for the two *D. rostriformis* adult individuals analyzed. Water samples sequence represents all environmental samples, all of them matching completely with adult *D. polymorpha* H2B sequences.

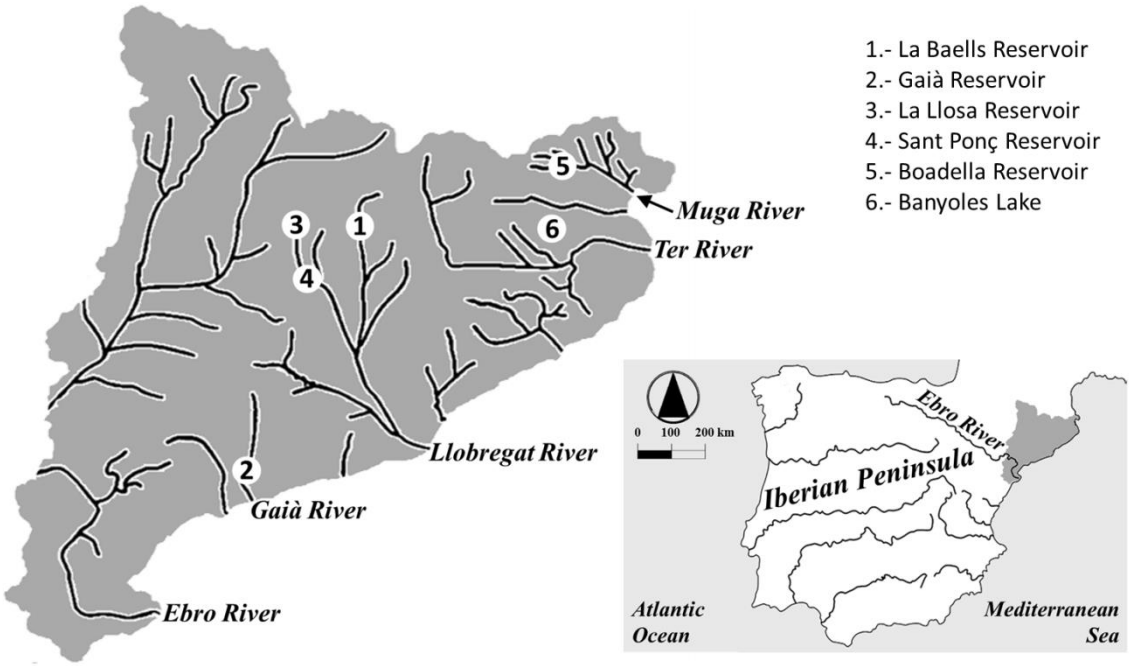


Figure S4. Map of the situation of the water samples analyzed in this study. All locations are situated in the Northwest Iberian Peninsula (grey area).

Table S1. Morphological characterization of adult dreissenid individuals collected from the Iberian Peninsula. Numbers in Sample correspond to different sampling recipients. N_m: number of mussels checked and measured; n_c: number of mussels checked only; N_{tot}: total number of mussels checked.

Location	Ebro River, Ribarroja reservoir					Llobregat River, La Baells reservoir								
Sampling date	May 13, 2013					May 13, 2013								
Sample	I/IV	II/IV	III/IV	IV/IV	Total	I/VIII	II/VIII	III/VIII	IV/VIII	V/VIII	VI/VIII	VII/VIII	VIII/VIII	Total
Coordinates	41° 15' N; 0° 26'' W					42° 71' N; 1° 53'' W								
Shell length (mm)	n_m	n_m	n_m	n_m	n_m	n_m	n_m	n_m	n_m	n_m	n_m	n_m	n_m	n_m
1	3	2		2	7									
2	5	5	4	8	22									
3	13	17	10	22	62									
4	16	18	19	33	86									
5	17	19	15	31	82			1						1
6	17	18	13	15	63	1								1
7	21	22	18	30	91				1		1			2
8	22	24	28	31	105									0
9	25	23	17	23	88	1	1						2	4
10	19	20	21	20	80	1								1
11	22	15	24	19	80	1		2				1		4
12	24	20	16	24	84	5			2		1	3	3	14
13	29	16	18	17	80	7	2	1	1	1	2	1		15
14	25	16	17	15	73	7	1		1	3	3	2	3	20
15	26	23	23	16	88	8	3	3	3	1	4	3	4	29
16	21	15	19	18	73	6	5	3		1	2	2	3	22
17	20	18	22	22	82	6	1	1	5	2		6	3	24
18	15	20	20	16	71	8	2	2	5	2	4	7	2	32
19	9	12	17	13	51	7	7	3	6	3	5	6	8	45
20	10	8	10	15	43	18	7	4	5	4	11	8	19	76
21	4	5	7	5	21	12	11	7	10	9	13	10	15	87
22	4	3	4	4	15	17	15	11	17	16	10	13	17	116
23	5	1	1		7	16	20	14	21	18	21	17	18	145
24	1	2			3	18	24	21	18	16	21	14	12	144
25	1				1	17	31	19	19	15	20	15	14	150
26						5	18	27	17	17	13	19	14	130
27						3	15	10	10	10	11	10	8	77
28						2	9	11	6	12	5	3	6	54
29						3	3	5	3	3	3	2	2	24
30									3	1		2	1	7
31							1	3	1	1				6
32								1						1
n_m total	374	342	343	399	1,458	169	176	148	154	135	150	144	154	1,230
n_c	331	469	310	445	2,909									
N_{tot}	705	811	653	844	3,013									

The zebra mussel (*Dreissena polymorpha* Pallas, 1771) and the Asian clam (*Corbicula fluminea* Müller, 1774) are considered two of the worst invasive aquatic species worldwide, causing major ecological and economic impacts on aquatic ecosystems.

Both species are present in the Iberian Peninsula since several years. The zebra mussel was first detected in the middle reaches of the Ebro River in 2001, and the Asian clam was first recorded in 1980 in the mouth of the Tajo River. While the current range distribution of the zebra mussel is found mainly in the Northeast of the Iberian Peninsula, the Asian clam is much more widespread and can be found in all major Iberian basins. However, introduction history and colonization routes of these two invasive species in the Iberian Peninsula remain mainly unknown.

In this respect, we have focused on understanding the possible sources of invasion of the two species, and how they have colonized and have expanded across the Iberian basins, using molecular and population genetics techniques. We have optimized new sets of microsatellites in each species using bioinformatic analysis of the results obtained by massive genome sequencing, and we also identified new SNPs. After selecting the best molecular markers, we characterized the genetic structure of the Iberian populations to infer their possible invasion routes. Finally, we have developed and optimized a genetic method based on environmental DNA and Real Time PCR to detect larvae of dreissenid mussels in water samples.