DOCTORAL SCHOOL IN BIOLOGY
SECTION: Biodiversity and Ecosystem Analysis
XXVII CYCLE, A.A. 2014/2015

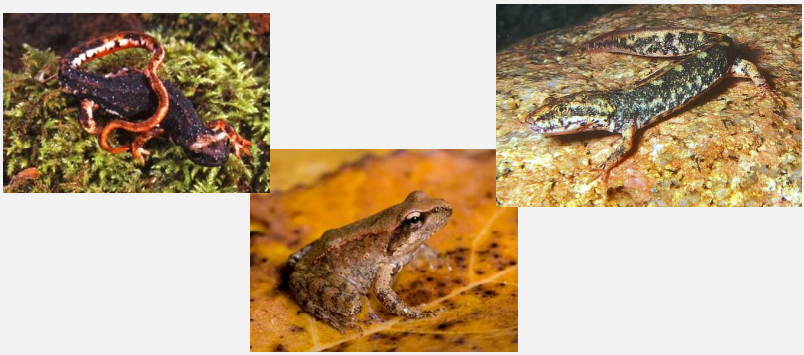
Ph.D. student: VALENTINA ROVELLI

Applying genetic and genomic methodologies for the conservation of *Salamandrina perspicillata*, *Euproctus platycephalus* and *Rana italica*

Tutor: Dr. Leonardo Vignoli

Supervisor: Prof. Marco A. Bologna

Co-tutor: Dr. Ettore Randi



DOCTORAL SCHOOL IN BIOLOGY
Section: Biodiversity and Ecosystem Analysis

XXVII Cycle - Academic Year 2014/2015

Applying genetic and genomic methodologies for the conservation of *Salamandrina perspicillata*, *Euproctus platycephalus* and *Rana italica*

CANDIDATE Ph.D. STUDENT: Valentina Rovelli

TUTOR: Dr. Leonardo Vignoli

CO-TUTOR: Dr. Ettore Randi

SUPERVISOR: Prof. Marco A.Bologna

Thesis defense on 27th February 2015 in front of the following jury:

Dr. José A. Godoy (European committee member)
Dr. Mattias Larsson (European committee member)
Dr. David R.Vieites (European committee member)

European evaluators:
Dr. Robert Ekblom
Dr. Cino Pertoldi

The extinction of species, each one a pilgrim of four billion years of evolution, is an irreversible loss. The ending of the lines of so many creatures with whom we have traveled this far is an occasion of profound sorrow and grief. Death can be accepted and to some degree transformed. But the loss of lineages and all their future young is not something to accept. It must be rigorously and intelligently resisted.

Gary Snyder (1990)

INDEX

PREFACE 3

ABSTRACT 5

CHAPTER 1 7
- General introduction
- Aims

CHAPTER 2 15
- She gets many and she chooses the best: polyandry in *Salamandrina perspicillata* (Amphibia, Salamandridae)

CHAPTER 3 31
- Genotyping by sequencing (GBS) of large amphibian genomes: looking for a needle in a haystack?

CHAPTER 4 43
- Using Genotyping By Sequencing (GBS) for delineating conservation units for the Sardinian Brook Salamander (*Euproctus platycepahlus*)

CHAPTER 5 57
- Individual heterozygosity and demographic estimates by GBS in a non-model amphibian species: preliminary results

CONCLUSIONS 65

REFERENCES 67

ACKNOWLEDGMENTS 81

PREFACE

This thesis encompasses a general introduction, four independent researches and a section of general conclusions.

In particular, the thesis is structured as follow:

CHAPTER 1. A general introduction to the global biodiversity crisis, the global amphibian decline and a focus on the Italian situation. Here I highlight the need of genetic conservation of endemic amphibian species and enunciate the aims of my PhD project.

CHAPTER 2. The chapter is structured around the first submitted manuscript (abstract not included): "Rovelli V, Randi E, Davoli F, Macale D, Bologna MA, Vignoli L. She gets many and she chooses the best: polyandry in *Salamandrina perspicillata* (Amphibia, Salamandridae). Biological Journal of the Linnean Society".

CHAPTER 3. The chapter is structured around the second manuscript under submission (abstract not included): "Rovelli V, Ruiz-González A, Vignoli L, Macale D, Buono V, Davoli F, Vieites RD, Randi E. Genotyping by sequencing (GBS) of large amphibian genomes: looking for the needle in a haystack? Conservation Genetic Resources".

CHAPTER 4. The chapter is based on the paper in preparation (abstract not included): "Rovelli V, Randi E, Macale D, Davoli F, Vignoli L. Using Genotyping By Sequencing (GBS) for delineating conservation units for the Sardinian Brook Salamander (*Euproctus platycephalus*). Conservation Genetics".

CHAPTER 5. The chapter is based on some preliminary results related to the paper in preparation: "Rovelli V, Vieites D, Vignoli L, Davoli F, Buono V, Randi E. Individual heterozygosity and demographic estimates by GBS in a non-model amphibian species: *Rana italica*".

The conclusion section highlights the main outcomes of the present study, in the light of the proposed aims, and address future research directions.

ABSTRACT

Because of a global crisis of biodiversity numerous plant and animal species are rapidly disappearing, and among vertebrates amphibians represent the most threatened group. Together with species diversity, also the ancient and extremely diversified amphibian genome is at risk of extinction. Since they are the most ancient landdwelling animals, their genome very likely holds important keys to understand crucial evolutionary events, among which vertebrate terrestrialisation. For this reason, amphibian conservation is a global priority. Being genetic variability one of the key requisite for species to adapt to environmental changes, the conservation of genetic diversity became one of the objectives of crucial importance in conservation biology. The general aim of this thesis is to provide new insights about the mechanisms that contribute in shaping the pattern of genetic diversity observed in three Italian endemic amphibian species. All of them are species of conservation concern, protected by European and National laws. Due to their different conservation status and specific research issues, their study was approached with different methodologies and perspectives.

As for *Salamandrina perspicillata*, an investigation at a very thin resolution scale (one population) was carried out by using microsatellite markers. The main aim of this study was to clarify the mating strategy adopted by females and understand how they can improve their fitness. This study provided the first evidence of polygynandry in a salamandrid species and the first report, in natural conditions, of salamander females that choose males genetically dissimilar from themselves to obtain indirect benefit (higher offspring heterozygosity).

As for *Euproctus platycephalus*, due to its condition of endangered species (IUCN 2014), the analyses focused on the investigation of the actual population genetic structure. In particular, the aim of this study was to identify the possible Evolutionarily Significant Units (ESUs), along the whole distribution area of the species, in order to plan adult individuals' collection for an ex-situ breeding project. Since there were no other genetic markers available for the species, apart from mitochondrial DNA (which would have not provided the necessary resolution power), the novel Genotyping By Sequencing technique have been used. Clusterization analyses revealed the presence of four ESUs, very likely comprising also adaptive groups.

The same genomic approach has been used for the third target species, *Rana italica*, with the aim of investigating its current genetic structure at a regional scale (Latium region). In this case the clusterization analyses suggest the presence of two different genetic groups. Further analyses focused on the estimation of individual heterozigosity, in preparation for Heterozygosity-Fitness-Correlation evaluations.

Thus, before of GBS application on the above mentioned species, another goal of this project was to produce the first genomic tools for these amphibian species.

CHAPTER 1

General introduction

The loss of biodiversity

Biodiversity is the variety of life on Earth, and comprises ecosystems, species, populations within species, and genetic diversity among and within these populations (Frankham 2004). Actually, the biological diversity of the planet is rapidly depleting as direct and indirect consequence of human activities. Although the exact number of disappearing species is still unknown, it has been estimated that the current extinction rate is the highest observed in the last 1,000,000 years (Eldredge 1998). The scale of the phenomenon is enormous and has been called the 'sixth extinction', as its magnitude compares with that of the other five mass extinctions revealed in the geological records (Frankham 2004). In nature, extinction is part of the evolutionary processes, but nowadays species are being lost at a rate that exceeds the speciation process and, unlike previous mass extinctions, is mainly due to human activities. Therefore, many species urgently require human intervention to ensure their survival.

Humans derive many direct and indirect benefits from the living world. According to Frankham (2004), human being has a stake and an obligation in conserving biodiversity for four main reasons: 1) for the bioresources used by humans (which include food, many pharmaceutical drugs, natural fibres, rubber, timber, etc.); 2) for the ecosystem services it provides (examples include oxygen production by plants, climate control by forests, nutrient cycling, water purification, natural pest control, and pollination of crop plants); 3) for the pleasure humans derive from nature and 4) for ethical reasons, since humans do not have the right to drive other species to extinction.

The only way humans have to preserve biodiversity is to protect all its forms that, as recognized by the IUCN, are represented by ecosystems, species and genetic diversity (McNeely *et al*. 1990). Being genetic diversity the basal level of biodiversity, the conservation of genetic diversity became one of the objectives of crucial importance in conservation biology (Allendorf & Luikart 2007).

Genes are responsible for the traits exhibited by organisms and, as populations of species decrease in size or go extinct, unique genetic variants are lost. Even if genes reside within species, genetic diversity is considered as a separate category from species because each population holds its own "genetic potential". This genetic variation allows populations, and consequently species, to evolve in response to the various selective pressures such as diseases, competitors, predators, parasites, environmental changes (i.e. pollution and climate change).

When dealing with the management of populations for a conservative purpose, demography represents an aspect of primary importance, as extinction is mainly a matter of demographic processes: the failure of one or more generations to replace itself with subsequent generations (Lacy 1988; Lande 1988). The more a population becomes smaller the more it is subjected to uncontrollable stochastic demographic

factors, and only when population's fluctuations and environmental modifications have been evaluated it is possible to estimate the expected time to extinction (Goodman 1987). According to Caughley (1994), there are two main types of threats to populations' survival: deterministic and stochastic threats. Among the deterministic threats, we can find habitat destruction, pollution, resource overexploitation, species translocation, and global climate change. On the other hand, stochastic threats are random changes in genetic, demographic, or environmental factors.

Genetic stochasticity is represented by genetic drift and increased inbreeding. The main consequences resulting from these phenomena are: 1) the increase of homozygosity and of the frequency of deleterious alleles, that often lead to inbreeding depression and thus might decrease the short-term viability of a population; 2) the loss of genetic variants, which will compromise the evolutionary adaptive potential of a population, and can thereby reduce its long-term viability, especially in changing environments; 3) the increase of genetic divergence among small and isolated populations, as a consequence of the genetic drift acting independently in each of them. As regards the last point, crossing individuals between populations, for instance in restoration programs that aim to enhance gene flow between previously isolated populations, might then lead to outbreeding depression (Ouborg *et al*. 2010).

When deterministic and stochastic threats arise together, their synergic effect can be lethal for small populations, as they can reduce populations' fecundity and viability. Thus, under some conditions, extinction is likely to be influenced almost exclusively by genetic factors. Indeed, a crucial question is when and under what conditions genetic concerns are likely to influence population persistence (Nunney & Campbell 1993).

Thus, conservation genetics arose as a discipline that uses genetic theory and techniques to reduce the risk of extinction in threatened species (Frankham 2004). Its longer-term goal is to preserve species as dynamic entities capable of overcome environmental changes. It benefits from the use of molecular genetics techniques to elucidate aspects of species' biology relevant to conservation management. Among the major conservation issues manageable through a molecular approach we can list: 1) the analysis of the deleterious effects of inbreeding on reproduction and survival (inbreeding depression); 2) the evaluation of the amount of loss of genetic diversity and investigating the ability of populations to evolve in response to environmental change (loss of evolutionary potential); 3) the study of the populations' fragmentation and the gene flow reduction; 4) the analysis of random processes (genetic drift) overriding natural selection as the main evolutionary process; 5) the accumulation and loss of deleterious mutations; 6) the genetic management of small captive populations and the adverse effect of adaptation to the captive environment on reintroduction success; 7) the resolution of taxonomic uncertainties; 8) the definition of management units within species (Frankham 2004).

The past decades has seen a large usage of neutral-behaving genetic markers such as microsatellites (or single tandem repeats, STRs) and mitochondrial DNA (mtDNA) control region, in order to assess the basic genetic variables in animal and plant

populations, with particular attention to the taxa presenting conservation concerns (Ouborg *et al*. 2010). The use of these molecular tools allowed to identify cases of reduced effective population size, restricted gene flow, limited heterozygosity, but also inbreeding, past bottlenecks and hybridization or gene introgression, all factors that could seriously affect the population viability and long term survival, especially in times of strong human-driven environment modifications and fast climate changes. The same genetic markers allowed the researchers to reconstruct the phylogenetic relationship, social structure, kin affiliations and individual fitness estimates in many social species, particularly among mammals and birds (Ellegren & Sheldon 2008).

The study of the relationships of individuals with the environment (considered in its widest sense to include the habitat, the social structure, the food networks - especially the prey-predator relations and coevolution, the climate and the pathogens), based on their genetic background, and the returning effects of the environment in driving and shaping the genetic features of the individuals through natural and sexual selection, has seen a never-dropping interest. However, the limited resources usually available to researchers did not allow for the investigation of a large number of genetic markers, therefore often limited to a few genes or non-coding regions of interest. Nowadays, on the contrary, revolutionary technologies such as Next Generation Sequencing (NGS) allow for the screening of thousands of genome-wide genetic markers, e.g. single nucleotide polymorphisms (SNPs), or whole genome sequences, in a very short time and with a relatively limited economic effort (Schuster 2008). This huge upgrade can make it easier to deepen existing disciplines (as for Ecological Genetics -or Molecular Ecology- and Genome-Wide Association Studies -GWAS), or even to open the way to the development of new branches, such as Conservation Genomics (Ouborg *et al*. 2010). This latter emerging discipline can be simply defined as the application "of new genetic techniques to solve problems in conservation biology" (Allendorf *et al*. 2010), such as genetic drift, hybridization, inbreeding or outbreeding depression, natural selection, loss of adaptive variation and fitness. The whole genomes of some endangered species have been recently completed, starting from the Great Apes: chimpanzee, gorilla and orang utan (Locke *et al*. 2011); however, these data will not automatically provide useful data for their conservation (Frankham 2010), especially given the limited information about population variation deducible from single individual sequencing. Nonetheless, this will provide a great aid in identifying genetic markers that can be applied to the study of entire populations (Frankham 2010). Genomic information will turn out to be useful also to try and recover populations from strong inbreeding depressions, by identifying the genes exposing deleterious alleles (Allendorf *et al*. 2010) and augmenting the population variability through crosses of the most appropriate individuals (Frankham 2010). On the other side, the same techniques will allow to identify the *loci* most responsible for speciation or cryptic local adaptation, or for exposing populations to severe diseases (Allendorf *et al*. 2010). Having a minor focus on conservation issues, other disciplines (whose boundaries are often difficult to define) raised, such as evolutionary and ecological functional genomics (EEFG; Feder & Mitchell-Olds 2003).

As a part of the general biodiversity crisis, amphibians are facing an imminent extinction emergency (Wake & Vredenburg 2008). Since 1970, scientist have observed fast and widespread population declines (Pounds & Crump 1994; Young *et al*. 2001), but the hypothesis of a global amphibian decline trend was suggested for the first time in 1989, during the First World Congress of Herpetology (Blaustein & Wake 1990; Collins & Storfer 2003). Since 1993, more than 500 populations of different amphibian species were considered to be declining or listed as requiring particular conservation actions (Blaustein & Kiesecker 2002). The great interest towards amphibians increased mainly for their sensibility to environmental changes, which makes them very good bioindicators. Even if, after 1989, there was a constant increase in literature about this topic, researchers did not find a consensus about the causes. However, during the Third World Congress of Herpetology in 1997, the need of establishing the global conservation status for Amphibian aroused even more powerfully. According to Collins & Storfer (2003), since 1990 we can observed three trends: 1) an increase in reports of amphibian populations decline and extinction worldwide; 2) the causes of decline seemed to be occurring simultaneously and over great distances; and 3) amphibian populations were declining also in protected natural areas. The latter was the most alarming issue because it meant that habitat protection, perhaps the best way to ensure a species' survival, was failing in many cases, and there were no comprehensive explanation for this phenomenon.

In 1994 Blaustein *et al*. (1994) highlighted the lack of long term studies about this topic, and the consequent impossibility in understanding the real extent of the phenomenon. Because amphibian populations are usually subjected to seasonal demographic fluctuations, further data were needed to clarify the global claimed trend. Despite the fact that it appeared clear that there is not a single cause for amphibian populations' declines and a number of culprits have been recognized, at the global scale few studies have provided convincing proof of causal relationships (Carey *et al*. 2001). Among the anthropic causes, the better studied is the habitat modification (Alford & Richards 1999). The vegetation removal or modification has a very strong impact on several populations: it exposes terrestrial species to altered microclimates, to soil drying, to habitat complexity reduction, so decreasing amphibian abundance and diversity. The same effects are generated by urbanization and infrastructures creation, which originate fragmentation and isolation of populations, increasing the risk of local extinctions. A second anthropic cause is represented by the introduction of alien species. Invasive alien species, in fact, can colonize a new habitat altering its equilibrium; in particular, they can predate on and/or compete with the amphibian autochthonous species, introduce new pathogens, and hybridize with the former species (Alford & Richards 1999). A third cause, as suggested by Collins & Storfer (2003), is represented by overexploitation, that is an excessive collection of individuals in the wild. At a wider geographic scale, we find the global climate change, which implies global warming, increase of ultraviolet radiations, exposition to

contaminants (for example derived from the use of DDT, PCB and CFC) and the increase in sensibility to pathogens.

The lack of a complete understanding about the decline extension and gravity led the IUCN to starting a global assessment on Amphibian (GAA: Global Amphibian Assessment) and a new data collection about distribution, abundance, populations' trends, relationships with the habitat and threats for all the 5743 described species (Stuart *et al*. 2004). The latest assessment of the Status of the World's Vertebrate reports that 41% of amphibians species are threatened (Baillie *et al*. 2010). Amphibians have the highest proportion of threatened species among vertebrates, but also the highest proportion of Data Deficient and the lowest proportion of Least Concern species (Baillie *et al*. 2010).

The Italian scenario

Moving to a European scenario, and in particular focusing on the Italian one, we find that as for amphibians Italy is the country owning the highest number of overall species, but also of endemic and threatened species in Europe (IUCN, Version 2014.3).

The European countries and the Member States of the European Union subscribed several conservation agreements, among which the Bern Convention (1979), the Washington Convention (1975), the Alpine Convention (1991) and the Convention on Biological Diversity (1992). For what concerns only the European Union, the conservation policy is based on two major directives: the Birds Directive (79/409/CEE) and the Habitat Directive (92/43/CEE). The most powerful conservation instrument that was born from these two directives is the Nature 2000 Network, which consists of sites containing habitat and species of unique conservation value. These sites are called Special Area of Conservation (SAC) and Special Protection Area (SPA), and most of them comprise very important habitat for amphibian species. However, as mentioned above, the presence of protected natural areas often is not enough to guarantee an adequate conservative standard, since the delimitation of those areas doesn't take into account potential changes in populations' distribution in response to external pressures. In Italy SAC and SPA cover about the 17% of the national territory, and the opportunity of creating ecological corridors and environmental restoration programs could guarantee the survival of a number of amphibians, otherwise convicted to a sudden decline.

The first step for a correct analysis in conservation biology is represented by the study of natural populations and the assessment of their conservation status. At the bottom of this investigation we find the monitoring activity, which allows the collection of data on population structure, ecology and genetics, on the presence of threats, and that can address long-term management actions. Thus, genetic monitoring of threatened species represents one of the most effective tools to investigate and prevent populations' decline.

As highlighted by Beebee (2005) and then by Calboli *et al*. (2011), there is a strong need for the application of conservation genetics to amphibian populations.

Indeed, the precipitous amphibians' decline involves also their genomic diversity, whose analysis could instead provide crucial insights for the understanding of several evolutionary questions, as for example vertebrates' evolution.

Aims

The overall goal of this thesis is to contribute in deepening the knowledge about three conservation concern amphibian species endemic to Italy: *Salamandrina perspicillata*, *Euproctus platycephalus* and *Rana italica*. All these species are protected by European Habitat Directive and National laws, but present different conservation status and priorities.

Salamandrina perspicillata is listed as Least Concern (LC) by IUCN (IUCN 2014), so actually there is not an immediate conservation emergency. For this reason I approached the population genetic issue on this species favoring an analysis at a very thin scale, by using microsatellite markers. I mainly investigated the mating system of the species, in order to identify the mechanisms through which individuals can increase their fitness and passing on their genetic pool. In particular, I tested one population of *S. perspicillata* for the presence of multiple paternity, already discovered in other salamanders. Since *S. perspicillata* is the sister taxon of all others Salamandridae, an assessment of the status of multiple paternity features along all salamandrids would be needed, in order to fully understand the evolution of mating systems within the family.

As for *Euproctus platycephalus*, this species is listed as Endangered (EN) by IUCN (IUCN 2014), and due to its condition of highly threatened species, it is also object of an ex-situ and in-situ conservation program, founded by EAZA and carried out by Fondazione Bioparco di Roma (Rome) in collaboration with Roma Tre University. Previous genetic analysis, focused on investigating the population genetic structure and phylogeography of the species, were based on mitochondrial DNA analyses. However further genetic studies were urgently needed to assess the current genetic status of the species and to plan individuals' collection for the ex-situ breeding program.

I carried out a detailed population genomics study throughout the whole distribution range of the species. Since microsatellite loci for this species are still unavailable, and I would have needed markers with a great resolution power for the analyses, I used Next Generation Sequencing techniques. In a conservation genetic context, for most species of interest (i.e. those threatened) there are no sequence resources available; however, this issue can be overcome by using NGS techniques, such as the novel genotyping by sequencing (GBS) technique. Therefore, a parallel goal of my project was to produce the first genomic tools for the species.

For what concerns *Rana italica*, it is listed as Least Concern (LC) by IUCN (IUCN 2014). Very few studies have been carried out on this species and in the

Latium region it is facing a slow decline since the beginning of the‘70s (Bologna et al. 2000). A previous study revealed that the overall geographical pattern of genetic variation found among the species’ populations closely matches the one of the so-called ‘southern-richness, northern-purity’ (Canestrelli *et al*. 2008). In particular in the Latium region the authors found a low genetic variability at allozyme loci. Due to the fact that there were no other genetic markers already available for the species, and since we needed a good resolution power for a fine-scale population genetic research, the same NGS technique used for *E.platycephalus* study has been adopted.

I analyzed samples belonging to seven populations from Latium region, for which aging (obtained from a skeletochronologic analysis) and demographic data have already been produced. The main aim of this preliminary work was to investigate the current genetic structure of the species in the study area, identifying possible genetic clusters. Once having genetically defined the groups, I estimated individual heterozigosity within the different populations. The next step will be to correlate these estimates with the skeletochronologic/demographic data, in order to identifying possible Heterozigosity-Fitness-Correlations (HFC).

CHAPTER 2

She gets many and she chooses the best: polyandry in *Salamandrina perspicillata* (Amphibia, Salamandridae)

Introduction

The evolution of animal mating systems has been traditionally approached from a male perspective. However, deepening animal reproductive strategies from the female perspective has provided new insights into this issue (Gowaty 1992), evidencing how females of many species significantly enhance their reproductive success by mating with multiple males, and re-interpreting polyandry as an active mating strategy (Bateman 1998; Zeh, Newcomer & Zeh 1998; Zeh & Zeh 1996, 1997). Indeed, since polyandry has been recognized as a pervasive feature of natural populations, the dogmatic role of females as the choosy, monogamous sex has been challenged (Trivers 1972).

Multiple paternity as consequence of females mating with multiple males has been described in a broad number of animal taxa, as for example mammals (Borkowska, Borowski & Krysiuk 2009; King, Banks & Brooks 2013), birds (Griffith, Owens & Thuman 2002), reptiles (Garner & Larsen 2005), amphibians (Adams, Jones & Arnold 2005; Tennessen & Zamudio 2003) and insects (Arnqvist & Nilsson 2000). Among amphibians, both Anura and Caudata showed polyandry. In particular, in the suborder Salamandroidea (Zhang *et al*. 2008), fertilization occurs internally by means of the transfer of a spermatophora from the male into the female spermatheca (Sever 1991). This enables long-term sperm storage (Sever 2002) and facilitates the occurrence of sperm competition and female cryptic choice phenomena (Parker 1998; Birkhead & Pizzari 2002; Wigby & Chapman 2004). Within this suborder, all species investigated so far have shown polyandrous mating strategies by females, with various and species-specific outcomes as for male's reproductive success (number of sired offspring), depending on several features characterizing the species reproductive behaviour (Table 1 and references therein).

Three main outcomes resulting from sperm competition and or female choice (phenotypic or cryptic) have been proposed for polyandric Caudata: 1) early male advantage: the first males sired a larger number of offspring than the second with topping-off mechanism (Jones, Adams & Arnold 2002; Tennessen & Zamudio 2003); 2) last male advantage: the sperm is stored in the spermatheca as stratified tangled masses favouring the ova fertilization by the spermatophorae belonging to the last male encountered by the female (Sever *et al*. 1999);3) mixed fertilization: no clear advantage for first or last male, with female choice (Darwinian or cryptic) as the underlying mechanism behind the observed male reproductive success (Garner & Schmidt 2003; Jehle *et al*. 2007; Chandler & Zamudio 2008). As for the latter category, relatedness between the females and her partners has been proposed as strong predictor for male success.

However, the observed patterns were not consistent within species, and the few taxa investigated on this issue were all tested in experimentally manipulated condition by mating each female to two (rarely three) males. Moreover, in nature, polyandric animals could behave differently than in simplified two-male mating experiments when the quality and the number of males are selected by the experimenter and not by the female (Zeh & Zeh 1994). Indeed, it should be mandatory that sperm precedence or male selection patterns discovered experimentally is corroborated by testing the same issue in full natural condition.

In the present work, we studied the mating system of the Northern Spectacled Salamander (*Salamandrina perspicillata,* Savi 1821), one of the two species belonging to the genus *Salamandrina*, the sister group of all other living Salamandridae taxa (Zhang *et al*. 2008). *Salamandrina* is a genus endemic to the Italian peninsula, and it is one of the most terrestrial within the family. The reproductive activity takes place on land from autumn to early spring with transfer of a spermatophora from male to female. The species exhibits terrestrial courtship behaviour with complete absence of physical contact between sexes (Bruni & Romano 2011). At the beginning of spring, when the oviposition period starts, only females enter the water to lay eggs on wood, leaves and stones (Della Rocca, Vignoli & Bologna 2005). Eggs are laid one by one and every female can produce from 20 up to a maximum of 60 eggs (Angelini, Vanni& Vignoli 2007).

Along with the description by Sever & Brizzi (1998), *S. perspicillata* holds the simple type of spermatheca, but so far, there is no evidence for the occurrence of polyandry or polygyny. Due to sperm degradation activity after oviposition (Brizzi *et al*. 1995), it is more likely that the expected occurrence of multiple paternity would result from polyandrous behaviour, instead of from sperm storage across different reproductive seasons. Most experimental studies on amphibian mating systems used aquatic explosive breeder species as a model due to the easy achievement of mating also in captive conditions (Tennessen & Zamudio 2003; Adams *et al*. 2005; Gopurenko *et al*. 2006; Liebgold *et al*. 2006). Since *S. perspicillata* is very difficult to mate in captivity, we focused on verifying the presence of multiple paternity under natural conditions.

According to the mating strategies shared by other Salamandridae species, we expected to find evidences of multiple paternity also in *Salamandrina*. Moreover, we expected to detect a male differential contribution in the fertilization of eggs. In detail, by collecting females in full natural condition before the oviposition phase, we aimed to answer the following key questions: (1) is polyandry, and the possible relative sperm competition, an important component in the mating system of the species? If the answer to this question is affirmative, (2) is there evidence for a male differential contribution in the fertilization of the eggs? Finally, (3) do the females choose the males (or their sperms) on the basis of their degree of genetic relatedness (Garner & Schmidt 2003; Jehle *et al*. 2007), and (4) does this choice produce indirect benefits (genetic) to the offspring?

Table 1. Synopsis of reproductive features related to polyandry of studied Salamandroidea species with data on the outcome of sperm competition/female choice on fertilization of eggs. Abbreviations: H: habitat where courtship occurs; A: aquatic; T: terrestrial; Br: breeding strategy; E: explosive; P: prolonged; Ref: References (see below).

Family	Species	H	Br	Sper	Outcome	Ref
Ambystomatidae	*Ambystoma maculatum*	A	E	S	Early male	1, 2
	Ambystoma texanum	A	P?	S	-	3
	Ambystoma tigrinum	A	P	S	-	4
Plethodontidae	*Desmognathus ochrophaeus*	T	P	C	Mixed	5
	Desmognathus ocoee	T	P	C	Mixed/Early male?	5, 6
	Plethodon cinereus	T	P	C		7
Salamandridae	*Salamandra salamandra*	T	P	S	Mixed	8, 9
	Notophtalmus viridescens	T	P	S	Mixed-Femalechoice	10
	Taricha granulosa	T	E	S	Early male	11
	Lissotriton vulgaris	A	P	S	Last male/Femalechoice	12, 13
	Ichthyosaura alpestris	A	P	S	Mixed-Female choice	14

[1]Tennessen & Zamudio 2003; [2]Chandler & Zamudio 2008; [3]Gopurenko *et al*. 2007; [4]Williams & Dewoody 2009; [5]Houck 1985; [6]Adams *et al*. 2005; [7]Sever & Siegel 2006; [8]Steinfartz *et al*. 2006; [9]Caspers *et al*. 2013; [10]Gabor *et al*. 2000; [11]Jones *et al*. 2002; [12]Gabor & Halliday 1997; [13]Jehle *et al*. 2007; [14]Garner & Schmidt 2003.

Materials and methods

Origin of females salamanders and study design

The study site is located within the Vejo Regional Park (latitude 42.105, longitude12.405; Latium region, Central Italy), quite close to the city of Rome. Here the studied population inhabited an oak (*Quercus cerris*) wood surrounding a small tributary (about 300 m)of the Crémera river. According to (Vignoli *et al*. 2010), individuals of this population follow four different behavioural and ecological phases: 1) courtship (October–November) when animals are in terrestrial activity (feeding and mating); 2) post courtship (December–January) when salamanders have probably ended mating activity but can still be found continuing feeding activity on land; 3) laying phase (February–early May) when ovipositing females are found in water and males on land; 4) pre-aestivation (late May–June) when a few salamanders are found on land, but showing reduced activity.

We collected a total of 43 females and 11 males of *S. Perspicillata* between October 2011 and April 2012. Individuals collected in fall-winter period, well before the ovipostion phase and likely before mating activity, were intended to be only used to estimate population genetic parameters. The collection of already fertilized and pregnant females in spring, during the oviposition phase and before they entered into the water, was aimed at analysing the polyandrous mating system in the study species under full natural condition. These females were found very close to the water, along the stream banks, just before entering the water for oviposting (Della Rocca *et al.* 2005). Sex determination in the field was accomplished by the observation of the external morphology of the cloaca, as according to the method developed by Vignoli *et al*. (2010). Each individual was marked by means of a picture of the ventral pattern (Della Rocca *et al*. 2005). Males and females were temporarily housed in two separate terraria (75x50x45h cm) at the Department of Sciences of Roma Tre University. The terraria were maintained in an environmentally controlled room, at a photoperiod of 12:12 h, temperature ranging from 12–19 °C and air humidity of about 60%. Salamanders were fed with live food (*Drosophila heidi*) and provided with *ad libitum* every two days.

At the beginning of April 2012 we moved the female salamanders at the facilities of the Bioparco in Rome into an outdoor enclosure (100x150x100h cm) filled with 5 cm of soil from the place of origin, covered with leaves, moss, small pieces of wood and some stones as hiding places for the animals, and providing animals with natural conditions as for the climate and photoperiod. Moreover, we set up 14 small terraria (34x20x22h cm) aimed at housing pregnant females for the oviposition activity. In each box, filled with reverse osmosis processed/purified water, we put true wood sticks (previously sterilized with an autoclave to prevent mould growth) as substrate where females can oviposit, and a floating piece of polystyrene covered with leafs as a terrestrial hiding/resting place. By observing female's vent swelling, we selected 10 individuals assumed to be pregnant and ready for laying eggs (Angelini *et al*. 2007), and we kept them individually in single fauna boxes for at least one week. At the end of the oviposition period eight females out 10 had laid eggs, as summarized in Table 3. Eggs were kept in small plastic tanks, 20x30x15h cm. To avoid cannibalism among the larvae, we maintained a maximum density of lower than 10 larvae. Every day we fed the larvae with *Artemia salina* nauplii (Crustacea) and renewed the water inside the fauna boxes.

As the larvae metamorphosed, they were moved into new terraria furnished with moss and fed with Collembola. When all the individuals had metamorphosed, we collected tissue samples by clipping a small tip of the tail (2/3 mm), which was then immediately stored in ethanol 95% at -20 °C. Observations of clipped individuals revealed no effects of invasive sampling on their short term survival. After the sample collection, we released all the individuals at their place of origin in the Vejo Park. This study, together with Steinfartz *et al*. (2006), is the only study on mating systems where all the sampled individuals - adults and larvae - were released into the wild.

Adults genotyping and quality control

Adults' tail tips were digested using a proteinase K solution (56 °C - overnight). Total genomic DNA was extracted using a ZR-96 Quick-gDNA^{TM} kit (Zymo Research), and all the individuals were genotyped at 10 microsatellite loci (Hauswaldt *et al*. 2012). The forward primers were labelled with fluorescent dye (Applied Biosystems): SALA-10H, SALA-H2, SALA-A11 and SALA-B8 with 6-FAM, SALA-3F and SALA-NG4 with NED, SALA-NC7 and SALA-A3 with PET, and SALA-D4 and SALA-D9 with VIC. Negative controls were always included in the analysis to check for contaminations. PCRs were performed in a 6 µl mix composed of a 0.8 µl reaction buffer 10X (5 PRIME), 0.80 µl BSA (Bovine Serum Albumine) 0.2%, 0.36 µl DNTPs 2.5 mM and 0.04 µl Taq polymerase (5 U/µl – 5 PRIME), brought to volume with H2O. In order to optimize the costs and time spent on the analyses, we ran two multiplex PCR (M1 included SALA-D4, SALA-3F, SALA-H2, SALA-NC7, SALA-B8; M2 included SALA-D9, SALA-NG4, SALA-10H, SALA-A3, SALA-A11). When further amplifications were needed to verify the data, only simplex PCR were used. Amplifications were performed using the following thermal profile: a first denaturating step at 94 °C for 2 min was performed, 35 cycles at 94 °C for 30 s, 50 °C for 30 s for M1 and 60 °C for 30 s for M2 as annealing temperatures, then 72 °C for 45 s for the synthesis, and a final extension step at 72 °C for 10 min. PCR products were electrophoresed with GeneScan^{TM} – 500 LIZ (Applied Biosystems) as a marker ladder in an ABI 3130XL sequencer and allele sizes were scored using the software GENEMAPPER v.4.0® (Applied Biosystems).

To set up the PCR reactions we used seven out of the eight mothers, performing a number of replicates ranging from six to 12. Since replicates were concordant and genotypes were always confirmed, for the remaining 47 adult individuals we performed only four replicates. As a first step for the quality control, allelic drop-out (ADO), false alleles (FA) and a percentage of positive PCR were calculated with the software GIMLET v.1.3.3 (Valiere 2002). As the adults represent a random sample of the breeding population, their consensus genotypes were used to estimate allele frequencies, expected and observed heterozygosities, deviation from the Hardy-Weinberg equilibrium, the number of alleles, the probability of identity and the probability of identity between sibs using the program GenAlEx 6.5® (Peakall & Smouse 2012). The software MICRO-CHECKER (Van Oosterhout *et al*. 2004) was used to test the adult genotypes for the presence of null alleles, which can be problematic in parentage analysis as they create false homozygotes and increase the genetic differentiation (Dakin & Avise 2004; Carlsson 2008), as well as for the presence of large allele dropouts and scoring errors due to stuttering. In order to assess the reliability of the obtained individual genotypes we used the software package RELIOTYPE (Miller, Joyce & Waits 2002), accepting a confidence level higher or equal to 90%.

Larvae genotyping and quality control

Larval genotyping was accomplished following the same protocol used for the adults, performing two individual replicates. During this step, the genotypes of 16 larvae were found to be incompatible with their mother's genotype, a likely consequence of allelic dropout. In order to solve this problem, we performed four additional replicates for each sample: in this way we were able to correct all the PCR errors. All the larval genotypes were verified with the software RELIOTYPE, accepting a confidence level of 90%, as for the adults.

Parentage analysis

Parentage analyses were performed with the software COLONY 2.0.5.0 (Jones & Wang 2010) and GERUD 2.0 (Jones 2005). We used COLONY to reconstruct the most likely number of paternal genotypes and assess the probability to correctly obtain the maternal genotypes. COLONY uses full-pedigree likelihood methods by taking into account the information of all the individuals at the same time (Wang 2004); this approach allows us to infer the parentage and sibships jointly (Jones & Wang 2010). COLONY divides the samples into three groups: the offspring sample (OFS), a candidate father sample (CFS) and a candidate mother (CMS). Individuals in the OFS are further categorized as full-sibs (sharing both parents), half-sibs (sharing a single parent) or unrelated (sharing no parents) (Jones & Wang 2010). Samples in CFS and CMS represent all the individuals that have some probability of being father and mother. If one of these groups is empty (in our case CFS) the software will estimate the reliability of the given genotypes (in our case maternal) and simultaneously will reconstruct the genotypes in the other group, without inferring any relationship. COLONY allows us to define beforehand some known relationships between the individuals; since we knew which offspring each female belongs to, we were able to constrain the relationship configuration. We assumed that if the origin population is large, individuals mate randomly, and genetic markers are in Hardy-Weinberg and linkage equilibria. Allelic drop-out and other errors, such as mutation, are taken into account, as are estimates. Furthermore, with a small sample size, and given some prior information, the number of possible configurations can be very large (Jones & Wang 2010). To look for the best configuration, the software uses a simulated annealing algorithm (Kirkpatrick, Gelatt & Vecchi 1983), which works by calculating the likelihood of all the possible configurations. After a comparison of two configurations, COLONY retains the one with the maximum likelihood, while the other is rejected, and then goes on until a solution is reached (Jones & Wang 2010). We set up the software by selecting "Full-Likelihood" as the method of analysis and "Complexity Prior" as the Sibship Prior. As regards the known paternal sibship, we selected no relationship, while for the maternal sibship we loaded the relative file. The analyses were performed using only the genotypes that overcame the threshold confidence of 90%, as assessed with the software Reliotype. We made an exception for only one genotype, which had a reliability of 89%. Since it belonged to one of the eight mothers

(Spe006), and those individuals were used to set up the multiplex PCR and were verified several times, we are quite confident that this genotype reflected the real one. To validate the results obtained with COLONY, we tested our data for multiple paternity also with the computer software GERUD version 2.0 (Jones 2005) and the related software GERUDSIM 2.0. Due to the very high computation time required by the program to run with more than five loci, we were not able to use all the nine loci. We selected the five loci on the basis of their PISibs values, previously calculated with GenAlEx. In contrast to COLONY, GERUD returns an estimate of the minimum number of fathers, together with an assessment of the maternal genotypes.

Pairwise relatedness coefficients between the female and each of the reconstructed paternal genotype were estimated by means of SM estimator (ranging from -1 (least similar) to 1 (identical) using the software MER (Wang 2004). If the females chose the males in relation to their genetic relatedness (the more diverse or the more similar) we expected a significant relationship between female-male genetic distance and the proportion of sired offspring by a given male (%OFF). We tested this hypothesis versus the case of random fertilization (null hypothesis), expecting a random distribution of %OFF among males with different SMs. Although %OFFs among males siring the same clutch are not independent, we assumed that %OFF is the direct outcome of the paternal selection by the females that acts upon the characteristics (phenotype or genotype) of the males. Hence, we used %OFF as dependent variables as is, without managing the apparent lack of independence among males' mating success. We performed a linear regression by using %OFF as dependent variable and the SM coefficients estimated for the given male as independent variable. In order to take into account the clutch size (CS) in the regression analysis, we performed a Factor Analysis (extraction method: PCA) using CS and %OFF by each male as variables. Then, in a new regression analysis, we used the scores of the first factor (eigenvalue: 1.161; explained variance: 58%) significantly associated to CS (loading: 0.767) and %OFF (loading: -0.767) as a new dependent variable, and the SM coefficients as independent variable (we consider only factors with eigenvalue > 1 and loadings in the excess of 0.71 as significantly associated to a given factor; Tabachnick & Fidell 2001).Finally, the average heterozygosity within clutches was estimated with software GenAlEx (Peakall & Smouse 2012).

Results

Overall, eight females laid eggs, for a total of 323 eggs (40.4 eggs per female on average). Seven of these females were collected pregnant between 20 March and 17 April; the remaining one was collected on 24 January. The egg mortality was quite low, with 278 eggs hatched and a survival rate of 86%; mortality was due to unfertilized eggs or mould growth. The survival rate of the larvae was 90%, with 251 individuals reaching the metamorph stage. All 278 hatchlings were used in the molecular analysis and we were able to genotype 273 of them. The other five larvae were cannibalized by sibs and we could not collect good quality DNA from the individual's remains.

As concerns the quality control, we found that six loci presented allelic dropout (ADO), but the average value of 0.017 is quite low (ranging from 0.003 to 0.109; Table 2). Only one locus presented false alleles, returning a general mean value of 0.003, while the percent of positive PCR was 0.91. Only one locus (Sala-10H) presents null alleles (Table 2). All the microsatellite loci were polymorphic and met the expectations of the Hardy-Weinberg equilibrium, except one (Sala-NC7), which was excluded from the analyses. The remaining nine loci showed an average number of alleles of 6.1, ranging from three to 11. The number of alleles per locus that was observed and the expected heterozygosity are summarized in Table 2. The probability of identity (PI) at nine loci was 8.3E-08, while the probability of identity between sibs (PISibs) was 1.2E-3, suggesting the absence of a shadow effect between sibs. We excluded from the analysis 26 larval individuals, which had a reliability value lower than 90%, while 53 adults out of 54 passed the reliability control. According to COLONY results, no one clutch had been sired by one single male, but females mated with at least two males up to a maximum of four. In particular, our results showed that two ovipositions had been sired by two males, five clutches by three males, and one clutch by four males (Table 4). GERUD outcomes confirmed these results, as shown in Table 3. The number of fathers did not correlate with clutch size (Spearman rank correlation: $r = 0.302$; $n = 8$; $p = 0.467$). The reconstruction of the more likely paternal genotypes, performed with COLONY, revealed that in all the analysed ovipositions a single male sired 50% or more of the offspring (range 50%-78%). The remaining males sired on average 16.06% of the offspring (range 2%-33%) (Figure 1).

The analysed offspring revealed polygyny: two out of eight clutches belonging to different females shared a same father (#6) (Figure 1). Intriguingly, the sired percentage of this male differed significantly between the two clutches (68% and 3% respectively). When the males' fitness (log-transformed %OFF) was tested against the genetic relatedness with the females (SM coefficient), we found an inverse and significant relationship ($R = -0.626$; $p = 0.001$) that was even stronger when clutch size (Factor1) was taken into account ($R = 0.705$; $p = 0.0002$): that is the more genetically dissimilar male has a higher paternity share (Figure 2). Moreover, the male presenting polygyny (#6) showed a negative value of SM when sired most of the clutch (68%), but a positive SM value when siring a very low portion of the offspring (3%). This result perfectly corroborated the observed relationship between %OFF and male SM for all the studied clutches.

To investigate indirect benefits of polyandry, we analysed the effects of number of fathers and of genetic relatedness between female and males (both the male with higher mating success and the average of all males siring a given clutch) on offspring average heterozygosity (HET). In this respect, we found that the number of fathers siring the clutch (NF) and the genetic similarity of the male with higher mating success correlated significantly with an inverse pattern with the offspring average heterozygosity (NF*HET: $r = -0.783$; $p = 0.021$; SM*HET: $r = 0.881$; $p = 0.004$), whereas the average SM of all males siring a given clutch showed no relationship (average SM*HET: $r = -0.262$; $p = 0.531$) (Figure 3).

Table 2. Description of the nine microsatellite loci used in this study that were not out of Hardy-Weinberg equilibrium. The table shows for each locus the values for allelic dropout (ADO), false alleles (FA), percent of positive PCR (% PCR), presence or absence of null alleles (NA, calculated on adult genotypes only), the allelic size range (Size, in bp), name of fluorescent dye (Dye), number of alleles (No. All), observed (H_0) and expected (H_E) heterozygosity.

Locus	ADO	FA	% PCR	NA	Size	Dye	No. All.	H_0	H_E
SALA-10H	0.000	0.000	0.94	yes	225-261	6 FAM	5	0.488	0.648
SALA-3F	0.004	0.000	0.98	no	120-170	NED	8	0.814	0.657
SALA-H2	0.005	0.000	0.99	no	225-270	6 FAM	7	0.721	0.747
SALA-NG4	0.109	0.000	0.76	no	125-148	NED	4	0.395	0.475
SALA-A3	0.000	0.000	0.80	no	240-274	PET	3	0.310	0.312
SALA-A11	0.000	0.032	0.91	no	307-350	6 FAM	8	0.791	0.779
SALA-B8	0.020	0.000	0.94	no	300-360	6 FAM	11	0.806	0.857
SALA-D4	0.018	0.000	0.96	no	110-140	VIC	5	0.643	0.629
SALA-D9	0.003	0.000	0.96	no	128-148	VIC	4	0.550	0.557
Mean	0.017	0.003	0.91	-	-	-	6.11	0.613	0.629

Table3. Results of the oviposition activity and paternity analysis. The table shows the mother's ID, the number of eggs laid by each female, the number of offspring effectively genotyped and the number of fathers found respectively with software COLONY and software GERUD.

Mother's ID	N° of eggs	N° of offspring genotyped	N° of fathers COLONY/GERUD*
Spe006	35	24	2/3
Spe007	38	36	3/3
Spe008	30	20	3/2
Spe009	42	30	3/3
Spe010	56	52	3/2
Spe011	45	37	3/2
Spe012	37	36	4/3
Spe013	40	38	2/2
TOT	323	273	--
Mean	40.4	34.1	--

*while COLONY returns an estimate of the most likely number of fathers, GERUD returns the minimum number.

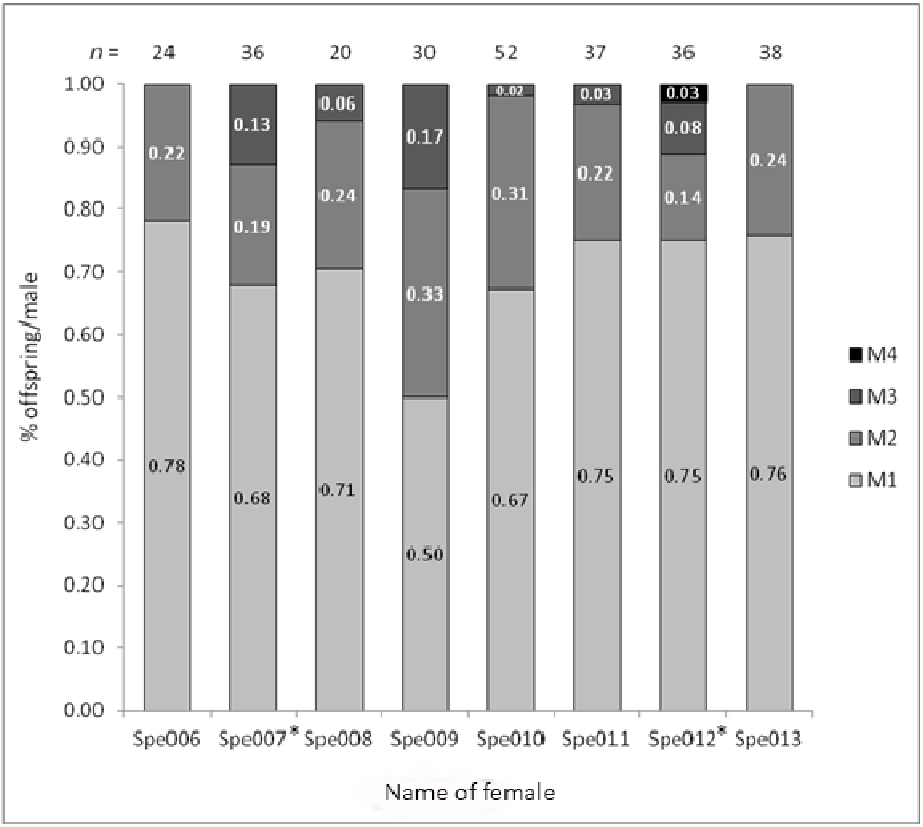


Figure 1. Number of fathers per clutch and percentage of the offspring sired by each father. Each clutch is represented by the code of the female. For each clutch, the fathers are called M1, M2, M3 and M4 based on the decreased percentage sired by each male (i.e. M1 is the most successful male). Asterisks indicate the two clutches that shared the same father (#6): in the clutch named Spe007 it resulted as M1 (with a percentage of sired eggs of 68%), while in the clutch named Spe012 it resulted as M4 (percentage of sired offspring of 3%).

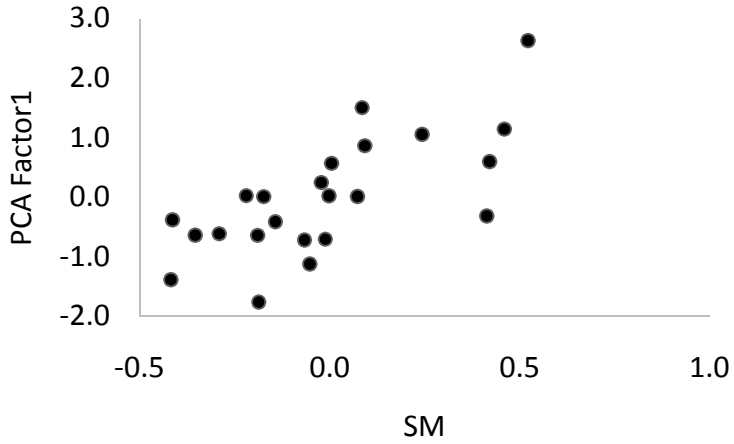


Figure 2. Scatterplot showing the negative relationship between the male-female genetic dissimilarity (SM coefficient) and the proportion of sired offspring (PCA Factor1): R = 0.705. Note that the observed positive relationship is because PCA Factor1was inversely associated to the percentage of sired offspring (loading: -0.767).

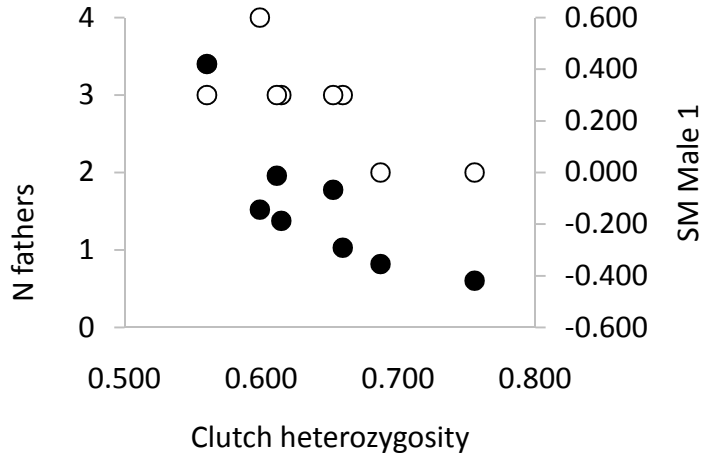


Figure 3. Scatterplot showing the relationships between the average offspring heterozygosity with (a) the number of fathers (white circles) and (b) the genetic dissimilarity (SM coefficient) between the most successful male (black circles) and the female.

Discussion

Main findings

Overall, we found that in the studied population of*Salamandrina perspicillata* multiple paternity occurs as a pervasive reproductive strategy under full natural conditions. The set of microsatellite markers used in this study was shown to be very informative, allowing us to document patterns of polyandry as well as to document polygyny in *S.perspicillata*. This is the first direct evidence for polygynandry in a wild population of a salamandrid species. Moreover, we demonstrated a role for the female-males genetic relatedness in female choice (Darwinian or cryptic). This choice provides an indirect benefit for the offspring in terms of heterozygosity, being higher in clutches mostly sired by a male genetically dissimilar to the female.

Polygynandry mating strategy

According to our results, every clutch has a minimum of two fathers, suggesting that the reproductive system of this species includes multiple mating by females. Even though we analysed only eight clutches, we found multiple mates as a pervasive and consistent strategy adopted by all females. Interestingly, other studies revealing multiple paternity in other Caudata species with variable sample size (N ranges 13-41) reported from a small to a significant portion of the analysed clutches sired by a single male (Garner & Schmidt 2003; Tennessen & Zamudio 2003; Gopurenko *et al*. 2006, 2007; Caspers *et al*. 2014). Our comparison of multiply sired clutches revealed no differences in fecundity, indicating that polyandry per se affords no fitness benefits to the studied salamanders. Moreover, although our sample is rather small we found evidence of polygyny with a male siring two different females. However, as the lower the number of sired eggs the more the reliability of the genotypes' reconstruction decreases, the evidence of polygyny has to be taken with caution. Indeed, since the male who sired the two different females in one case fertilized only one egg, it is possible that the reconstructed genotype belongs to a different but genetically similar male (possibly a close relative). Although polygynandry has been demonstrated in other tailed amphibian taxa (see Gopurenko *et al*. 2006) and likely occurs in most Caudata taxa, this would be the first evidence in a salamandrid species.

Since in our study mortality at egg and larval stages was very low, we identified the number of fathers which was very likely close to the real one, making our conclusions very sound. On the contrary, in other works on amphibian multiple paternity (i.e. Tennessen & Zamudio 2003) the percentage of analysed eggs was on average 7% of the total female fecundity due to embryo mortality at very early stages. When a very low proportion of eggs belonging to a given clutchis considered with the purpose of discovering and interpreting multipaternity, there is a high risk of misinterpreting the real number of males sired and the actual percentage of the oviposition fertilized by each single male. The high portion of the clutch not available

for analysis due to embryo mortality can hide one or more males involved in the clutch fecundation.

In the light of the observed pervasive pattern of polyandry, our analyses on the degree of genetic similarity between mates yielded three interesting results: (i) males mated to a given female and siring most of the offspring tend to be genetically dissimilar from their sexual partner; (ii) a same male, when mated with two females, sired a proportion of the offspring inversely correlated with his genetic similarity to the female; (iii) genetic dissimilarity between mating partners is positively correlated with offspring degree of heterozygosity.

Female choice and male-female genetic similarity

The first result, that is, high genetic dissimilarity between males and the female with which they mated is positively related to the proportion of siring offspring, is in agree with the common idea of inbreeding avoidance where females increase genetic variability among their offspring by being polyandryc and by choosing mates that are genetically different from themselves (Stapleton *et al*. 2007; Sluter *et al*. 2007; Hoffman *et al*. 2007). As far as we know for tailed amphibians, only three experimental manipulative studies revealed such a pattern with a female choice based on male relatedness (Garner & Schmidt 2003; Jehle *et al*. 2007; Chandler & Zamudio 2008). Our study is the first evidence of such a mating system in natural condition in a salamander species. Two studies on newts (*Ichtyosauria alpestris* and *Lissotriton vulgaris*) revealed that in two-male mating experiments less-related males are preferred, or, at least, more successful fathers, with neither phenotypic traits selected by the females nor male mating order having an effect on male reproductive success (Garner & Schmidt 2003; Jehle *et al*. 2007). Interestingly, Chandler & Zamudio (2008) found in controlled mating experiments in the field on *Ambystoma maculatum* the opposite pattern (outbreeding avoidance or inbreeding preference), with the more successful males not too distantly related to their mates. However, the interpretation of this finding by Chandler & Zamudio (2008) is complicated by the presence of stored sperm from the previous breeding season, the effect of body size on the proportion of sired offspring, and the interaction between body size and relatedness. In any case, deviations from inbreeding avoidance have already been described in other taxa (e.g. Cohen & Dearborn 2004; Jennions *et al*. 2004), and the pattern observed by Chandler & Zamudio (2008)could be a further exception.

Despite *S. perspicillata* lacks of evident secondary sexual dimorphic traits present in other salamander species, such as a larger tail crest, nuptial pads, and greater body size (Angelini *et al*. 2007), the females of this species have shown to be very selective as for mating (Darwin 1871; Bateson 1983; Andersson 1994). This might suggest a strong effect of sexual selection within the female (attributes of males perceived during the copulatory courtship and/or attributes of sperms selected within the spermatheca) that potentially could override the effects of selection acting at earlier stages (i.e. mating order, timing of mating). Indeed, the females of this species should be able to recognize the less genetically related male by selecting it at the

precopulatory (mate choice) and/or the postcopulatory/prefertilization phase (when the sperms are within the spermatheca).

Female incongruence in mate preference

The second result referred to the single case we found of a male that had sired two different females with contrasting mating success. Indeed, when this male fertilized a female genetically dissimilar, the proportion of the sired offspring was high (in this case the male was the most successful out of three males with 68% of sired eggs); on the contrary, when the genetic dissimilarity between male and female was low, the male sired a very low portion of the offspring (in this latter case the male was the least successful male out of four). Even if the result is based only on a single male mating to two females, this evidence perfectly corroborates the findings showed for the first result (see above), and suggests that the siring success likely could be a matter of male-female individual relative affinities rather than based on intrinsic features of the male. Although this result needs to be confirmed with further observation, the general pattern evidenced for the study species would indicate that *S. perspicillata* females are incongruent in their mate preference for a particular male (Neff & Pitcher 2005), with each preferring a different male, i.e. the one genetically more dissimilar.

Offspring genomic divergence: indirect benefits for the choosy females

The third result showed that the higher heterozygosity among offspring appeared to be the consequence of the uneven sired offspring proportion among mating males biased towards the less genetically similar male. Intriguingly, the number of fathers seemed to have a detrimental effect on clutch heterozygosity degree, likely because the more are the mates the less are the chances that the offspring are fertilized by the best male only. On the contrary, higher levels of heterozygosity were correlated to the genetic dissimilarity of the most successful male. Thus, polyandry per se did not provide a genetic indirect benefit to the offspring. This could indicate that multiple mating enables female to distribute fertilizations among several males, thus reducing the impact of low efficiency in the evaluation of male quality (Gabor, Krenz & Jaeger 2000). For example, if the females are unable to discriminate among males based on phenotypic traits, then they may gather indirect benefits from a polyandric mating strategy as a sort of genetic bet-hedging. Thus, since in *Salamandrina* there are no evident phenotypic traits on which females can rely on for mate choice at the precopulatory phase, mating multiple times may enhance the chances for a female to fertilize her offspring with sperm from genetically dissimilar males through cryptic choice. Indeed, we can speculate that the observed pattern might be a matter of cost/benefit mechanism in which the gained benefit of multiple mating counterbalances the negative effect of the number of mates on the offspring heterozygosity.

In the species in which females are highly selective when it comes to mating, the question why they are so choosy assumes a crucial evolutionary significance (Neff &

Pitcher 2005). In mating systems in which males provide resources (i.e. food or shelter) as direct benefits to the females or their offspring, females should recognize and select those males that are able to provide more resources (Moller & Jennions 2001). However, in the nonresource-based mating systems, in which females still show a preference among males despite they provide no resources but genes (i.e. sperm), as the study system is, the answer is less straightforward (Neff & Pitcher 2005). Models have shown that nonresource-based mating systems require that choosy females gather an indirect benefit (genetic) through increased offspring fitness (survival and/or reproductive advantage at the adult stage). Since in *S. perspicillata* we observed females each preferring a different male, the observed pattern would suggest that the genetic quality should reflect interactions between paternal and maternal genomes rather than the inheritance of the called "good genes" (Neff & Pitcher 2005). Thus, although we cannot exclude a complementary effect of sperm competition to the observed pattern, we hypothesize that behind the biased mating success among males likely there was a females' cryptic choice by means of which they directly manipulate sperm usage and bias fertilization to the male that will produce offspring of higher genetic quality (higher offspring heterozygosity) (Neff & Pitcher 2005). According to the genetic compatibility hypothesis, a male with compatible genes will produce offspring with higher fitness only when matched with a specific maternal haplotype thanks to favourable gene–gene interactions through, for example, heterozygote advantage (Trivers 1972; Zeh & Zeh 1996, 1997; Neff & Pitcher 2005). As in amphibians larval survival can be related to genetic diversity/heterozygosity (Beebee 2005), we hypothesize that the study species mating system is compatible with the genetic compatibility model: the females obtain indirect benefit (higher offspring heterozygosity) by choosing males genetically dissimilar from themselves.

Ecological and behavioural implications

All the females that oviposited during our study were collected already pregnant in nature. Only one female collected in January, well before the oviposition phase, laid eggs. This confirmed that *S. Perspicillata* can potentially store sperm for a long period (> 60 days) (Sever 2002; Adams *et al*. 2005). This evidence makes an in-depth analysis on the possible occurrence of sperm competition even more necessary. The fact that all the other females collected in January did not lay eggs could be likely due to an early sampling (i.e. most females had probably not mated at that point).

As for the male contribution to the clutch fertilization, our results revealed that one single male sired more than the half of the whole oviposition, while other males on average fecundate less than 20% of the eggs (2%-33%). The studied population showed a male-biased sex-ratio of 6.67 (Vignoli *et al*. 2012). If we oversimplify the studied reproductive system by assuming that each male fertilized only a single female, and taking into account that on average one female is fecundated by three males, we can roughly estimate that the percentage of males in the population that reproduce would be at most 45% with an operational sex ratio less biased (1:3). Moreover, our calculations are likely overestimates of the real ratio, since we found

multiple mates by males and it has been observed that over the prolonged reproductive period one male can court several females by monopolizing them thus limiting other males the access to the female (Bruni & Romano 2011). Those salamander species showing an aquatic reproductive activity (i.e., *Ichthyosaura alpestris* (Rafinski & Osikowski 2002), *Lissotriton vulgaris* (Sever *et al*. 1999), *Notophtalmus viridiscens* (Gabor *et al*. 2000), *Taricha granulosa* (Jones *et al*. 2002)), show a mechanism of first male or last male advantage (but see Jehle *et al*. 2007 for *L. vulgaris*). In these species sexually dimorphic morphological traits are often conspicuous and individuals arrive at the same breeding site simultaneously, so females have at their disposal several males, gaining the opportunity to make a careful precopulatory selection on males signalling fertility benefits. The species characterized by terrestrial mating, as for example *S. salamandra* (Caspers *et al*. 2014), usually do not show evident secondary sexual dimorphic traits and present mainly a pattern of sperm mixing, suggesting sperm storage associated to female cryptic choice and sperm competition have a more relevant role in paternity outcomes. Since *S. perspicillata* is amongst the salamander species with the most terrestrial habits, we can hypothesize that the observed pattern of unbalanced fitness among fecundating males reflects the presence of a postcopulatory mechanism.

Conclusions

In summary, we answered our starting key question, revealing that females of *S. perspicillata* show multiple paternity. Moreover, we found that there is strong evidence for an uneven male contribution in egg fertilization, with one male siring most of the eggs in a clutch. Our findings suggest a female cryptic choice at the base of the observed pattern with a genetic indirect benefit (higher offspring heterozygosity) provided by the selection of the male more genetically dissimilar from the female. Although these data should be interpreted with caution due to the small sample size and the lack of experimental manipulative testing, the observed pattern was apparently confirmed by different and independent findings (results 1-3), thus providing robustness to our conclusions. Indeed, given our limited sample size and accounting for the fact that under field condition several factors are potentially confounding any genetic effect, we found a strongly supported pattern. However, since the studied females were collected already mated, we have no data with which to infer which kind of mechanism is actually at work in this species. In the light of the obtained results, a more in-depth analysis with an experimental approach would be required in order to achieve a full understanding of *S. perspicillata*'s reproductive strategy.

CHAPTER 3

Genotyping by sequencing (GBS) of large amphibian genomes: looking for a needle in a haystack?

Introduction

About 360 million years ago, amphibians became the first vertebrates to live on land (Calboli *et al*. 2011). Their ancient and extremely diversified genome very likely holds important keys to understanding crucial evolutionary events, among which vertebrate terrestrialisation (Calboli *et al*. 2011). Currently, amphibians are facing a global extinction crisis of unprecedented magnitude that is heavily threatening the survival of many species and their genomic diversity (Stuart *et al*. 2004; Calboli *et al*. 2011). Although 41% amphibian species are listed as threatened in the IUCN Red List, they have benefited least from conservation efforts (Hoffmann *et al*. 2010). Without prompt and focused conservation actions in consequence of habitat loss and fragmentation, diseases, invasive species, global climate change, chemical contaminants and overexploitation, a large number of amphibian species will be lost in the near future (Alford & Richards 1999; Pimm & Raven 2000; Collins & Storfer 2003; Stuart *et al*. 2004).

Until now, conservation genetic studies, based on small numbers of variable loci, have revealed important insights about the structure of endangered amphibian populations (Beebee 2005). However, novel next-generation sequencing (NGS) approaches, based on large-scale sequence information, have the potential to enormously improve our research ability. The opportunity of using thousand of genome-wide genetic markers would greatly increase our understanding of populations' genetic structure, and, especially, would help in detecting loci under selection. This is of particular importance because can give us a measure of the adaptive potential of populations, and provide critical estimates of their ability in responding to rapid environmental changes, such as those caused by climate change or emergin infection deseases (Calboli *et al*. 2011). As suggested by (Allendorf *et al*. 2010), the chance to examine thousand of genetic markers at the same time made possible to solve many conservation issues that have been unattainable until now. For example, it will be possible to estimate neutral population parameters, such as effective population size, or to study the genetic basis of local adaptation or inbreeding depression (Allendorf *et al*. 2010).

The study of amphibian genome, however, presents an intrinsic difficulty, due to its unusual size for vertebrates. It ranges from an average size of 9.36 Gb for anurans up to 35.90 Gb for salamanders (Gregory, 2011), this latter representing the second largest accepted animal genome (Dufresne & Jeffery 2011).

The analysis of such a huge genome would have been impossible with standard Sanger sequencing methods. The rapid progress NGS platforms, has bypassed the steps for marker assay development and library construction, allowing a direct comparison of large numbers of sequences for identifying DNA polymorphisms in a

wide range of species (Narum *et al*. 2013). Genotyping-by-sequencing (GBS) is one of the new NGS-based methods that increase the power of SNP genotyping, by both simultaneously collecting millions of short-read sequences for every individual and improving sequence coverage per locus through genome complexity reduction (Elshire *et al*. 2011). However, the performances of GBS in species with very large genomes and high genetic diversity are largely unknown (but see Chen *et al*. 2013). GBS in amphibians posses important challenges due to large genome sizes but also due to a lack of detailed genomic information [i.e., with the exception of the recently published western clawed frog (*Xenopus tropicalis*) genome sequence (Hellsten *et al*. 2010), there are no reference genome sequences available]. Therefore, amphibians are challenging and interesting species for assessing the effectiveness of GBS.

In this work we reported the comparison of GBS results for two amphibian species of conservation concern endemic to Italy, the Sardinian Brook Salamander (*Euproctus platycephalus*) and the Italian stream frog (*Rana italica*). Both species are protected by the Habitat Directive (Annex IV) and the Bern Convention (Appendix II), and are targets of local conservation efforts. Moreover, *E. platycephalus* is listed as Endangered (EN) by IUCN. While *R. italica* has a genome size of 5.66 pg (Gregory 2011), the exact dimension of *Euproctus platycephalus*' genome is unknown. However, since its two most phylogenetically closely related species, *E. montanus* and *E. asper*, have respectively a genome size of 23.10 and 27.62 pg, we can hypothesize that *E. platycephalus* should have a genome size > 20 pg.

Overall our results demonstrated the suitability of GBS for SNP discovery and genotyping in two conservation concern amphibians, by providing informative reference on restriction enzyme selection as well as plexing level issue in NGS analyses when dealing with large genome sizes.

Methods

DNA samples and GBS protocol optimization

Tissues samples, collected by clipping a small tip of the tail (2/3 mm), were digested using a proteinase K solution (56 °C - overnight). The total genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen – Cat. no. 69506), according to the manufacturer's protocol, from 95 *Rana italica* (RI) and 95 *Euproctus platycephalus* (EP) samples. DNA quality was assessed by running 100 ng of each DNA sample on 1,5% agarose gel. DNA concentration was determined using the QuantiT ™ PicoGreen® dsDNA Assay Kit (Invitrogen). For optimization of the GBS protocol, a single DNA sample (500 ng) from each species was digested for 2 h with the restriction enzymes *Ape*KI, *Eco*T22I and *Pst*I, using a tenfold excess of enzyme and reaction conditions as specified by the enzyme manufacturer (New England Biolabs). After ligation of appropriate adapters (adapter amounts were determined by titration as described in (Elshire et al. 2011) and PCR (see below), fragment size distributions of each test library were visualized using an Agilent BioAnalyzer 2100.

Preparation of Illumina libraries for next-generation sequencing

Two 96-plex *Eco*T22I GBS libraries, comprising 95 DNA samples of each species, henceforward abbreviated as *EP* (*Euproctus platycephalus*) and *RI* (*Rana italica*), and a negative (no DNA) control, were prepared according to (Elshire *et al*. 2011). Briefly, individual DNA samples were digested with the restriction enzyme and adapters were legated as described previously. The adapters comprised a set of 96 different barcodes containing adapters and a "common" adapter. The oligonucleotide sequences of the barcode adapters were as follows:

(a) 5´-ACACTCTTTCCCTACACGACGCTCTTCCGATCTxxxx and

(b) 5´-CWGyyyyAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

where "xxxx" and "yyyy" denote the barcode (AACGCACATT, AACGTGCCT, AACTGG, AAGACGCT, AATAACCAA, AATGAACGA, AATTAG, ACAACCAACT, ACAACT, ACAGT, ACCAGGA, ACGCGCG, ACGGTACT, ACTGCGAT, ACTGCT, ATAGAT, ATATAA, ATATCGCCA, ATCCG, ATCTGT, ATGAGCAA, ATGGCAA, ATTAT, CAACCACACA, CAAGT, CACCA, CAGAGGT, CAGATA, CAGTGCCATT, CATAT, CATCTGCCG, CCACTCA, CCGAACA, CCTCG, CCTTGCCATT, CGCAACCAGT, CGCACCAATT, CGTCGCCACT, CGTGGACAGT, CGTGTCA, CGTTCA, CTAAGCA, CTCAT, CTCGCGG, CTCGTCG, CTCTA, CTCTCGCAT, CTTAG, CTTGA, GAAGCA, GAAGTG, GAATGCAATA, GAGCGACAT, GCAAGCCAT, GCACGAT, GCCAACAAGA, GCCTACCT, GCGCCG, GCGCTCA, GCGTACAAT, GCGTCCT, GCTCCGA, GGAACGA, GGAAGACAT, GGACAG, GGACAG, GGAGTCAAG, GGATA, GGCTTA, GGTATA, GGTGCACATT, GGTGT, GTCGCCT, GTGACACAT, TAGATGA, TAGCAG, TAGCCAA, TAGCGGAT, TATCA, TATGT, TATTCGCAT, TCACGGAAG, TCACTG, TCAGAGAT, TCCGAG, TCTTGG, TGAAT, TGACGCCA, TGCAGA, TGCCGCAT, TGCTT, TGGCAACAGA, TGGCACAGA, TGGCCAG, TTATTACAT, TTCGTT, TTGCTG), and barcode complement, respectively. The EcoT22I common adapter was as follows:

(a) 5′-CWGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG and

(b) 5′-CTCGGCATTCCTGCTGAACCGCTCTTCCGATCT.

Individual ligations were pooled, and purified using QIAquick PCR purification kit (Qiagen). Genomic fragments were then amplified in a 50-µL volume containing 2-µL pooled DNA fragments, 1× Taq Master Mix (New England Biolabs), and 25 pmol, each, of the following primers:

(a) 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT and

(b) 5′-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT. Temperature cycling consisted of 72 °C for 5 min, 98 °C for 30 s, followed by 18 cycles of 98 °C for 30 s, 65 °C for 30 s, and 72 °C for 30 s, with a final extension step at 72 °C for 5 min. The EcoT22I GBS library was purified again, as above, and an aliquot was run on the Agilent BioAnalyzer 2100 for evaluation of fragment sizes and the presence of adapter dimers. After quantification on the Nanodrop 2000 (Thermo Scientific), each library was diluted and sequenced (single-end reads only) in a single flow cell channel on the Illumina HiSeq 2000 at the Cornell University Genomics Core Laboratory.

In order to overcome the large genome size of *EP*, and explore the multiplexing level issue in GBS analysis, we re-runned the same 96-plex *EP* library on an additional sequencing Illumina lane. The combined results of both 96-plex sequencing data gave rise to the same output that would be obtained by a 48-plex level.

DNA sequence analysis: SNP discovery and genotyping

Raw DNA sequences were analyzed with the Universal Network Enabled Analysis Kit (UNEAK) pipeline, tailored for species that lack a reference genome sequence (Lu et al. 2013). This pipeline is implemented in TASSEL v3.0 (http://www.maizegenetics.net/images/stories/bioinformatics/TASSEL/uneak_pipeline_documentation.pdf). Briefly, the raw Illumina DNA sequence data (100-bp qseq files) were first trimmed to remove barcodes. The sequence remnants were then either trimmed further or padded with 3' A's to 64-bp lengths. Sequences were then aligned to each other, both to identify unique sequences, or "sequence tags", and to generate clusters of related sequences. For each cluster, a network was generated, in which sequence tags were organized according to mutation steps (i.e., mutational relationship). A single base-pair mismatch was allowed among cluster members. Networks were then filtered such that only SNPs originating from reciprocal tag pairs were retained (see Lu *et al.* 2013). SNPs from more complicated networks that often result from alignment of paralogs and repeats, or sequencing errors were discarded. To further reduce the impact of sequencing errors, we also set the error tolerance rate (ETR) parameter to 0.03, slightly above the expected Illumina sequencing error rate (0.04 %). Pipeline default parameters were used for filtering the resulting table of genotypes. First of all, individual with high failure rates (i.e. individual call rates <0.15 or 0.30 = missigness >70% or 85 %) were filtered from the database. The genotypes' table was then filtered using default parameters, except that the minimum value of the minimum minor allele frequency (mnMAF) was set to 0.05. Further filtering of the data set was done to eliminate SNPs present in < 80 % of sample DNAs.

Results

Library fragment size distribution

The fragment size distributions of GBS libraries from *EP* and *RI* genomic DNA digested with different restriction enzymes are shown in Figure 1. Discrete peaks (i.e., repetitive DNA fragments) were present in most libraries. The size distribution curve for both species was smoothest for ApeKI, but it was pretty large, and fragments over ~ 500bp are too long to sequence on the Illumina platform. Also, the *EP* genome is so large that this frequent cutter (ApeKI has a 5-base recognition sequence but it cuts more like a 4-base cutter because of a degeneracy in the recognition site) would have probably produced too many fragments to get deep enough sequence coverage per SNP locus to be able to call heterozygotes with any confidence. On the other hand, the fragment size distribution is tighter with EcoT22I, so this enzyme should probably produce deeper sequence coverage than PstI. Based on these results, we chose to sequence GBS libraries derived from EcoT22I genomic digests containing little repetitive DNA.

Number of sequences and SNPs

Sequencing results showed that all 95 samples for each species were represented. As concerns *EP* a total of 279,893,656 and of 501,580,539 reads was found for the 96-plex and for the 48-plex levels, respectively, while for the 96-plex plate of *RI* the number of total reads was 269,695,656. From these raw sequences, 247,872,958 and 451,654,647 good barcoded reads were obtained respectively for the 96-plex and 48-plex *EP* plates; as for the *RI* plate 181,237,253 good barcoded reads were found. The distribution of reads number in individual samples from the EcoT22I libraries for the two species and for the different plex levels is shown in Figure 2.Resulting number of unique sequence tags were 24,892,397 for the 96-plex plate and 45,790,964 for the 48-plex plate of *EP*, while we get 23,863,809 for the 96-plex plate of *RI*. After analyzing the raw sequences with the UNEAK pipeline using default parameters, we obtained 134,803 (96-Plex) and 156,753 (48-plex) SNP loci for *EP*; resulting SNPs for the 96-plex plate of the *RI* were 20,399. The average call rate per locus for the unfiltered SNPs was 0.337 and 0.390 for 96-plex and 48-plex of *EP*, respectively; for *RI* this value was 0.495 (see Table 1).

As summarized in Table 1, after the removal of individuals presenting high failure rates (i.e call rate lower than 85% and 70% for *EP* and *RI*, respectively) the average call rate per locus for the 96-plex and 48-plex of *EP* was 0.353 and 0.409 respectively, and 0.559 for the 96-plex of *RI*. The seven individuals for *EP* and the 14 individuals for *RI* that had a call rate lower than the selected thresholds were excluded from further analyses. When loci were filtered for call rate < 0.8, the resulting numbers of SNPs on *EP* were 909 and 1480 for the 96-plex and 48-plex, respectively; while for *RI* a total number of 2531 SNPs were obtained. As expected, an increase in the call rates was observed, with a value of 0.887 (96-plex) and 0.885 (48-plex) for *EP*.

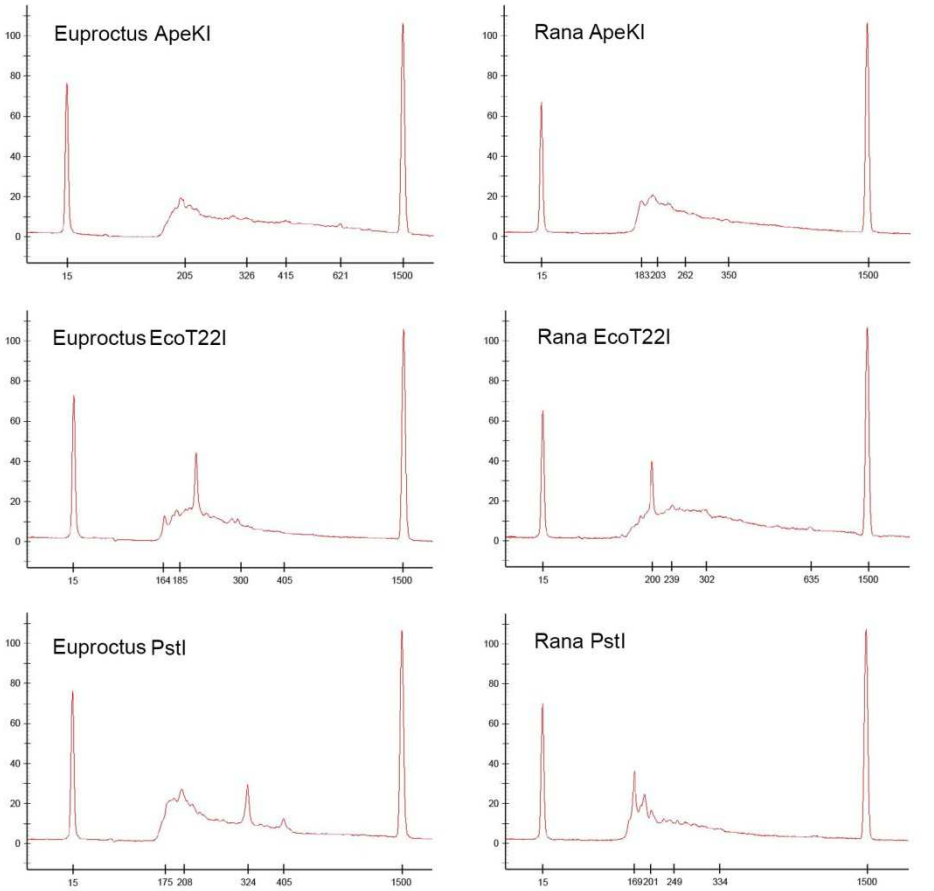


Figure 1. Fragment size distribution of GBS libraries made with a single DNA sample (*left*: newt; *right*: frog) using three restriction enzymes (*top*: ApeKI; *middle*: EcoT22I; *bottom*: PstI). The x-axis represent elution time and the y-axis shows fluorescence units. Numbers below hatch marks on the x-axis indicate fragment size (bp). Tall peaks at 15 and 1500 bp are size standards.

The resulting value for *RI* was 0.850. After further filter for MAF < 0.05, 595 and 961 SNPs were obtained for *EP*, for the 96-plex and 48-plex, respectively; as for *RI*, the number of resulting SNPs was 854. Graphical comparisons among the number of SNPs obtained after applying different filtering criteria, depicted according to the study species and plex level, are shown in Figure 3. The overall call rate for the two plex levels of *EP* after MAF filtering didn't change, while for *RI* it slightly increased to 0.860.

Table 1 Statistics summary results for the two species, divided in statistics per locus and statistics per individual according to the different plex levels. Filter1 for individuals with call rate < 0.15; Filter2 for individuals with call rate < 0.30; Filter3 for markers with call rate < 0.80; Filter 4for markers with MAF < 0.05.

	Euproctus platycephalus								*Rana italica*			
	96-plex				48-plex				96-plex			
	No filter	**Filter1**	**Filter3**	**Filter4**	**No filter**	**Filte1**	**Filter3**	**Filter4**	**No filter**	**Filter1**	**Filter2**	**Filter4**
N. individuals	95	88	88	**88**	95	88	88	**88**	95	82	82	**82**
SNPs	134803	13480	909	**595**	156753	156753	1480	**961**	20399	20399	2531	**854**
Call Rate	0.337	0.353	0.887	**0.887**	0.390	0.4087	0.885	**0.885**	0.495	0.559	0.850	**0.860**
Statistic per locus												
Min average coverage	1.004	1.004	1.233	**1.157**	10.032	1019	1.568	**1.66**	1.040	1.060	1.160	**1.360**
Max average coverage	7.526	7.522	37.081	**37.104**	7.490	11.603	49	**48.377**	27.910	27.910	50.110	**57.360**
Av. coverage	2.370	2.393	8.989	**8.814**	2.150	3.389	12.112	**11.833**	7.120	7.240	12.920	**15.140**
Median	1.850	1.872	7.112	**6.902**	1.660	2.731	9.638	**9.406**	5.390	5.520	10	**11.720**
Missingmess	0.662	0.646	0.112	**0.112**	0.694	0.591	0.114	**0.114**	0.504	0.440	0.140	**0.130**
St. deviation	1.630	1.644	7.164	**7.189**	1.440	2.515	9.490	**9.269**	5.890	5.910	10.420	**12.140**
Statistic per individual												
Min average coverage	1.000	1.000	1.000	**1.000**	1.000	1.000	1.011	**1.011**	1.000	1.000	1.000	**1.030**
Max average coverage	127.663	132.886	125.715	**119.59**	253.320	254	254	**254**	146.620	161.93	161.93	**161.93**
Av. coverage	2.410	2.452	8.586	**8.826**	2.250	3.548	12.120	**11.853**	6.880	7.490	12.350	**14.720**
Median	2.010	2.079	9.010	**5.909**	1.740	2.465	8.050	**7.880**	5.280	5.670	9.870	**10.700**
Missingmess	0.662	0.646	0.151	**0.112**	0.694	0.591	0.114	**0.114**	0.506	0.440	0.140	**0.130**
St. deviation	2.442	2.509	11.189	**10.724**	2.540	3.863	15.85	**16.868**	6.530	7.110	11.340	**16.220**

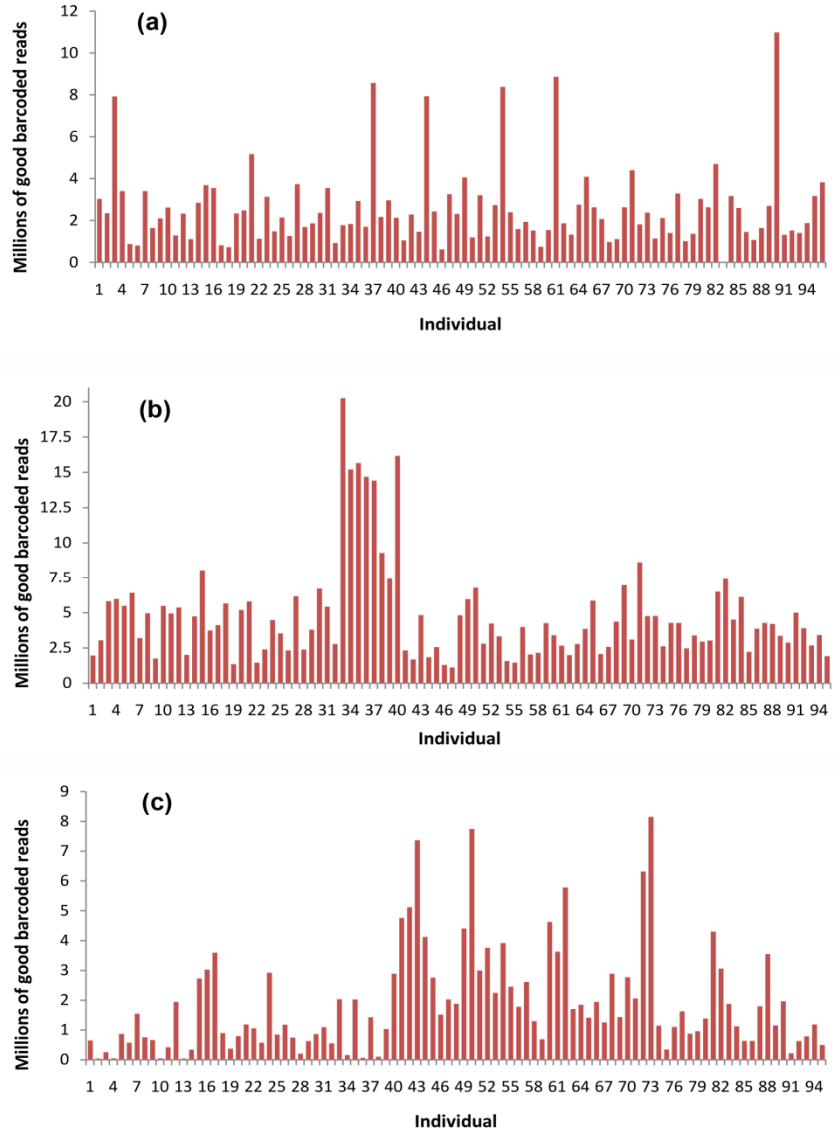


Figure 2. Distribution of the number of sequence reads. (a) represents the number of good barcoded reads in 95 DNA samples of a 96-plex depth plate of *Euproctus platycephalus*, (b) shows the number of good barcoded reads in 95 DNA samples of a 48-plex depth plate of *Euproctus platycephalus*, and (c) shows the number of good barcoded reads in 95 DNA samples of a 96-plex depth plate of *Rana italica*.

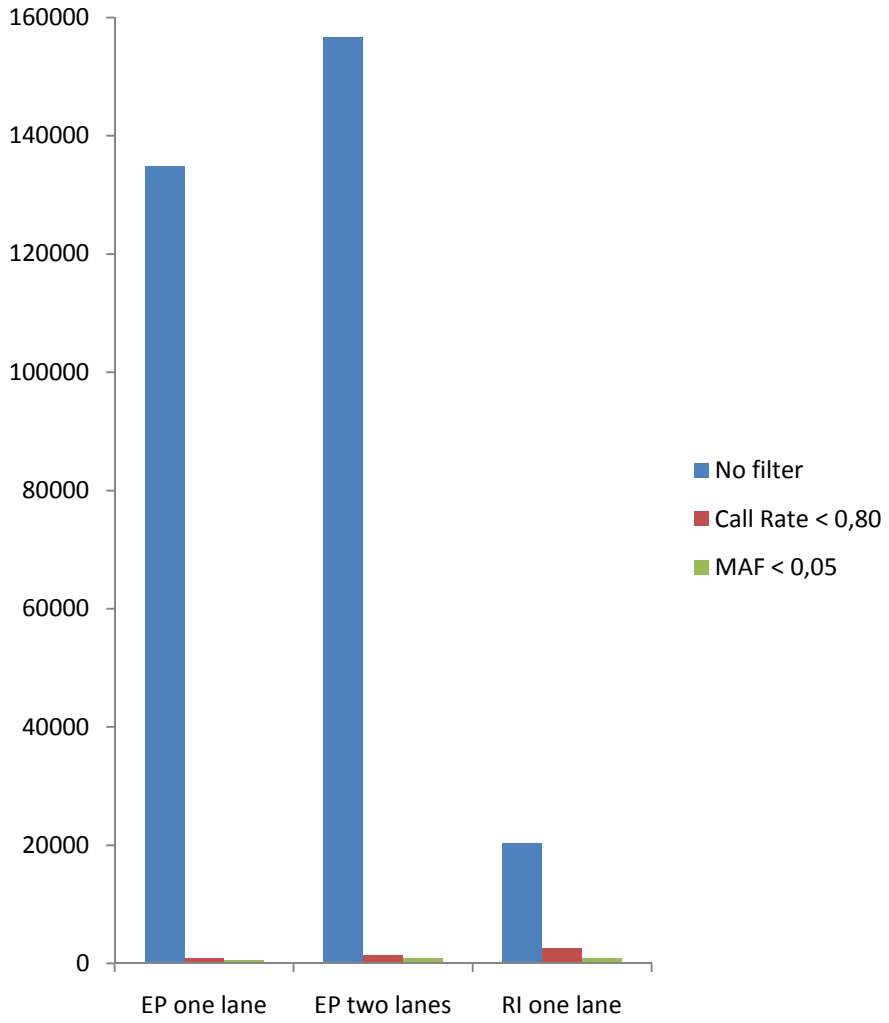


Figure 3. Number of SNPs obtained for the two species for the different plex levels. For *Euproctus platycephalus*, results for the 96-plex (EP one lane) and 48-plex (EP two lanes) depth plates are shown. For *Rana italica* we have only 96-plex depth (RI one lane). Blue bars represent the number of SNPs obtained from the application of the UNEAK pipeline, without filters application. Red bars represent the number of SNPs retained after having filtered for Call Rate < 0.80, while Green bars represent the final number of SNPs obtained after having filtered for MAF < 0.05, that are 595, 961 and 854, respectively.

Sequence coverage depth

Overall, the unfiltered SNPs identified by the UNEAK pipeline had a low average coverage depth (Table 1). The average coverage value per locus, for the unfiltered dataset, was 2.370 and 2.150 for the 96-plex and 48-plex experiments, respectively; while the average coverage value per individual was 2.410 and 2.250. In *RI* we found that the average coverage per locus was 7.12, and the average coverage per individual 6.880. After excluding individuals with low call rates, we got similar average coverage values per locus of 2.393 (96-Plex) and 3.389 (48-plex) for *EP*. Similarly, the average coverage results per individual were 2.452 (96-Plex) and 3.548 (48-plex). With regard *RI*, the average coverage per locus was 7.240, while the average coverage per individual was 7.490. After filtering markers with more than 20% of missing data, we obtained a significant increase in the average coverage. The average coverage per locus for *EP* changed to 8.989 and 9.638, for the 96-plex and 48-plex, respectively; while for *RI* we get a value of 12.920. The average coverage per individual, increase to 8.586 and 12.120, according to the lower and higher plex level of *EP*; for *RI* we get a value of 12. The average missing data ratio is ca. 11% for both species (Table 1).

Discussion

To the best of our knowledge, this study is the first attempt of GBS application on amphibian species and the first set of novel SNP markers developed for *EP* and *RI*. Overall, we demonstrated the suitability of GBS for SNP discovery and genotyping in two conservation concern amphibians, by providing, after stringent filtering, 961 (595 for the 96-plex library) and 854 novel and reliable SNPs for the Sardinian Brook Salamander and the Italian Stream Frog, respectively. In addition, we provided informative reference on restriction enzyme selection as well as plexing level issue in GBS analysis when dealing with large genome sizes. Our results demonstrated that GBS is a robust and suitable method for genotyping large amphibian genomes and for further development of SNP-based conservation genomics studies.

It is remarkable that we obtained a very low number of SNPs in comparison to those obtained in other species on which GBS technique has been applied. For example, (White *et al*. 2013) found about 6000 SNPs in the bank vole, *Myodes glareolus* (Order Rodentia, Family Arvicolinae), while (Chen *et al*. 2013) found about 18,000 SNPs in two conifer species, *Pinus contorta* and *Picea glauca* (Order Pinales, Family Pinaceae). The bank vole has a genome size comparable to *R. italica* [the exact dimension is not reported, but taking as references some related species according to Gregory (2011), we can suppose it is about 3 pg], while the two conifer species have a genome size comparable to *E. platycephalus* (20 pg on average).

Being the differences in resulting SNPs really considerable, it is unlikely that they are due only to a greater genome complexity in amphibians. On the contrary, it is plausible that they reflect the genetic diversity within the species. The vole and conifers have a wide distribution area: *P. contorta* is a common tree in western North America, *P. glauca* is native to the northern temperate and boreal forest in North

America, and *Myodes glareolus* ranges from the British Isles through continental Europe and Russia to Lake Baikal, reaching in the south northern Turkey and northern Kazakhstan. On the other hand, both *E. platycephalus* and *R. italica* have a small distribution area, being the first one endemic to Sardinia Island and the second one to peninsular Italy (Sindaco *et al*. 2006). Moreover, their dispersal abilities and population sizes are reduced, and, as suggested by Allendorf & Luikart (2007) and Frankham (2004), all these factors can decrease genetic diversity.

According to our results, future GBS analyses on amphibian genomes should address the restriction enzyme selection towards EcoT22I, which proved to be able in minimizing the amount of repetitive DNA during GBS library preparation process. When dealing with complex and large genomes like in amphibians, it is recommended to increase sequence coverage depth per locus, and one way to achieve this result is the selection of a restriction enzyme that does not cut frequently in the genome (Chen *et al*. 2013). However, there is the need to find a balance between coverage depth and the total number of called SNPs, since less frequent cutters produce higher coverage but also less SNPs. There are lot of cases, for example genome-wide association studies (GWAS) and genomic selection (GS), where a large number of SNPs are needed. In such a case, the better strategy would be to use a frequent cutter and try to increase coverage depth by running additional lanes and/or using a lower multiplexing level.

One of the advantage of the GBS methodology is that at a low cost it is possible to increase the number of SNPs and/or sequence coverage per SNPs locus by running the same library in further sequencing lanes or by running the libraries at a lower multiplexing levels (Elshire *et al*. 2011; De Donato *et al*. 2013). In order to overcome the large genome size of *EP* and explore the effect of plexing level on SNP numbers and depth of coverage, we re-ran the same 96-plex *EP* library on an additional sequencing Illumina lane (i.e. giving rise to the same output that would be obtained by a 48-plex level). In this study, by running a 96-plex GBS library 2 times (i.e. 48-plex) for *EP* we overall increased the number of SNPs identified from 596 to 961 (i.e. ca. 61% increase) but also, and importantly, increased the average sequence depth per locus from 8.81 to 11.83, increasing statistical support both for identifying variants and calling heterozygotes.

Our SNPs calling results follow the observation reported by (De Donato *et al*. 2013), regarding the increase in the number of SNPs with each sequencing lane. The authors found that this effect is especially pronounced during the first few replicated runs, gaining more than 1.5 fold increase in the number of SNPS with the second run. For *EP* we run two independent lanes at 96-plex that we later on combined to produce a 48-plex, and looking at Figure 3 we can see that in the second lane we obtained an increase of about half of the first one, reflecting the same cited trend. However, this result is true only for the filtered loci, while as regards the unfiltered loci we get almost the same number for the two different lanes.

Comparing the number of unfiltered SNPs obtained for *EP* and *RI*, this value turns out to be higher for *EP*, for both plex levels used *(EP: 96-plex=134.803; 48-*

plex=156.753 vs RI: 96-ple= 20.339). However, taking into consideration the number of SNPs obtained after having filtered for both call rate (<0.8) and MAF (<0.05), we got a similar number of SNPs for the 48-plex of *EP* and the 96-plex depth plate of *RI*. These outcomes reflect the differences in size between the two analyzed genomes. Since *EP*' genome is approximately more than five times bigger than the *RI* one, we could expect a higher total number of SNPs for the first species, that is what we found for the unfiltered loci (Fig. 3). In the case of *EP*, in fact, the restriction enzyme can cut in many more sites on respect to *RI*, producing a higher number of tags and correspondently of SNPs. However, right due to the high number of produced fragments for *EP*, during the amplification step it is possible that not every fragment is amplified, resulting in a very low average read number per tag. Thus, due to the low quality and consequent low reliability of many SNPs, the number of SNPs obtained from the two genomes after filtering resulted comparable. Even if the initial number of SNPs for *RI* was lower on respect to *EP*, among them there were many more reliable markers.

The previous outcomes are confirmed also by the results obtained for coverage. Indeed, looking at Table 1, it is evident the differences between the average coverage values for *EP* and for *RI*, as regards the Unfiltered and Filter1 columns. For what concern the newt, this value doesn't overcome the 2.5x, while for the frog it reaches at least 6.8x. In contrast with what we could expect, after the second run the average coverage value for *EP* remained similar, becoming on the contrary even lower. Only after having filtered for call rate < 0.80, the average coverage for both locus and individual level, at both plex levels, increased. Very interestingly, the 48-plex level for *EP* returned a value comparable to that obtained for the 96-plex depth plate of *RI*.

Here we reported that for a species with a huge genome (i.e. *EP. ca. > 20 pg)*, it is possible, by running one additional sequencing lane or using a lower multiplexing level (i.e. 48-plex) obtaining comparable results with those returned from species with approximately five times smaller genome (i.e. *RI* 5.66 pg).

In this pilot study, the average missing data ratio is ca. 11% for both species, and thus similar to those reported for large conifer genomes using a 48-plex level (i.e ca. 7-9%; (Chen *et al*. 2013).

In conclusion, GBS confirmed to be a cost-effective genotyping method, able to identify a high number of good quality SNPs also in non-model amphibian species with a very large genome that will be of great interest for further developing conservation genomics studies. The ability to examine hundreds of loci will increase the power and accuracy in estimating a variety of important parameters in conservation. For example, genomic techniques would be especially useful to study the genetic basis of local adaptation or inbreeding depression. Moreover, in the long term, they could be related to fitness and other demographic parameters in order to predict population viability or the capacity to adapt to climate change.

CHAPTER 4

Using Genotyping By Sequencing (GBS) for delineating conservation units for the Sardinian Brook Salamander (*Euproctus platycephalus*)

Introduction

Genetic diversity is one of the three fundamental levels of biodiversity, and it is directly linked to its conservation (McNeely *et al.* 1990; Reed & Frankham 2003). Conservation of genetic variability is important to the evolvability of populations because decreased genetic variability leads to reduced fitness and adaptation (Freeland *et al*., 2011). Small and isolated populations are particularly affected by loss of genetic variability, since the dynamics of genetic variation is strongly dominated by random genetic drift and inbreeding (Ouborg *et al.* 2010), leading them to a possible local extinction in a short time, as compared to larger populations. Loss of genetic diversity is enhanced in species with low dispersal capabilities, which reduces the rates of gene flow. The consequences of isolation can be even worse when the species is an insular endemism. Insular taxa indeed are prone to extinction because they are vulnerable to demographic stochasticity, strong climatic events, and anthropogenic disturbance (Cook & MacDonald 2001).

Resources for conservation have always been rather limited and, as highlighted by Allendorf *et al.* (2010), describing conservation units (CUs) is one of the most important contributions of genetics to conservation. Conservation units are population units identified within the species level, that can be useful to address management and conservation efforts (Funk *et al.* 2012). The two more widely used conservation units are evolutionarily significant units (ESUs) and management units (MUs). Summarizing the several definitions that have been proposed in the course of time (Dizon *et al.* 1992; Fraser & Bernatchez 2001; Moritz 1994; Ryder 1986; Vogler & Desalle 1994; Waples 1991), Funk *et al.* (2012) referred to ESUs as populations that have substantial reproductive isolation, which has led to adaptive differences so that the population represents a significant evolutionary component of the species. On the other hand, MUs represent demographically isolated populations whose population dynamics depend mainly on local birth and death rates rather than on immigration (Funk *et al.* 2012; Moritz 1994; Palsboll *et al.* 2007). According to these differences, Funk *et al.* (2012) suggest to use both neutral and adaptive loci to delineate ESUs, while MUs should be determined using only neutral loci. Nowadays, next generation sequencing allows for the screening of thousands of genome-wide genetic markers, e.g. single nucleotide polymorphisms (SNPs) or whole genome sequences, in a very short time and with a relatively limited economic effort, which provide markers that allow performing population genomic analyses (Frankham 2010). Therefore, these novel methods facilitate to deepen existing disciplines (e.g. Molecular Ecology and Genome-Wide Association Studies - GWAS), or even to open the way to the development of new branches, such as Conservation Genomics (Ouborg *et al.* 2010). This emerging discipline can be simply defined as the application “of new genetic

techniques to solve problems in conservation biology" (Allendorf *et al.* 2010), such as genetic drift, hybridization, inbreeding or outbreeding depression, natural selection, loss of adaptive variation and fitness.

Amphibian species are particularly vulnerable to loss of genetic diversity and local extinction, due to their low vagility and high sensibility to environmental changes (Kiesecker *et al.* 2001). According to the IUCN Red List of Threatened Species (Version 2014.3 - www.iucnredlist.org)**,** they are the most threatened vertebrate group on earth. The target species of the present work, the Sardinian Brook Salamander, *Euproctus platycephalus* (Caudata, Salamandridae), is an amphibian species endemic to Sardinia Island (Italy). It is actually found only on the eastern part of the island, from the North to the South, while records from the western part go back to 1995-2003 (see Lecis & Genètica de la Conservaciò 2004; Vignoli L. pers. comm.). It is a mountain species that prefers cooler waters and is usually found in streams, small lakes, pools and even artificial canals. Its terrestrial habitats are generally restricted to riverine scrubs or woodlands, and the species may also be found in cave systems. Population declines have been reported since the early 1980s (Puddu *et al*., 1988; Vanrooy & Stumpel 1995), probably caused by reduction and fragmentation of habitats (usually due to water redirection for agricultural purposes) (Vanrooy & Stumpel 1995), agricultural water pollution, illegal fishing methods and the introduction of allochthonous fishes. *Euproctus platycephalus* is classified as Endangered (EN) by the IUCN (2009), and it is listed on Appendix II of the Bern Convention and on Annex IV of the EU Habitats Directive; it is also protected by regional legislation (Regional Law n. 23/1998 (art. 5, c. 3)). Actually, the species is object of an ex-situ and in-situ conservation project, founded by EAZA and carried out by Fondazione Bioparco (Rome, Italy) and Roma Tre University (Rome, Italy). Since one of the aims of the project is creating and maintaining an ex-situ breeding stock, an in-depth knowledge of the population's genetic structure is of crucial importance. Moreover, because of its intrinsic value as declining endemic species, an update of the actual conservation status is required, in order to get essential information for in-situ management actions.

Given these assumptions, with this work we aimed at clarifying the genetic structure of *Euproctus'* populations, and identifying the main evolutionary significant units (ESUs). In parallel, we aimed at developing the first genomic resources for the species. Until now, indeed, *E. platycephalus* has belonged to the category of non-model species, which are those without genomic resources available. Together with the intrinsic difficulty of studying a non-model species, the analysis of *E. platycephalus'* genome represented an additional challenge. According to the Animal Genome Size Database (www.genomesize.com), in fact, the C-values of the two most closely related species, *E. montanus* and *E. asper*, are respectively 23.10 and 27.62. So we can assume, with a certain degree of confidence, that also the Sardinian Brook Salamander has a huge genome, with a C-value very likely greater than 20. However, this issue can be overcome by using NGS techniques, such as the novel genotyping by sequencing (GBS) technique (Elshire *et al.* 2011), which is a method for the reduction

of genome complexity and SNPs discovery (see Methods section for an in depth description).

Moreover, we performed a series of tests aimed at identifying the best method to reduce the number of SNPs up to 96, with the perspective of create a SNP array for the species.

Methods

Sampling and DNA extraction

During three expeditions in June, July and November 2012, 95 Sardinian Brook Salamander individuals were sampled from 9 sample sites in Sardinia, spread along the whole distribution area of the species. We collected samples as follow: 20 individuals from Monte Limbara, 8 individuals from Monte Albo, 11 individual from Supramonte di Oliena, 2 individuals from Roa Paolinu, 12 individuals from Villagrande, 9 individuals from Foresta Ortuabis, 12 individuals from Monte Ferru, 3 individuals from Perdasdefogu and 18 individuals from Settefratelli (Figure 1). Differences in number of individuals collected from distinct localities have to be attributed to the relative density of the populations and to the difficulty of sampling in certain sites.

Samples were collected by clipping a small tip of the tail (2/3 mm), which was immediately stored in ethanol 95% at -20 °C. Tissues were then digested using a proteinase K solution (56 °C - overnight). The total genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen – Cat. no. 69506), according to the manufacturer's protocol. DNA quality was assessed by running 100 ng of each DNA sample on 1,5% agarose gels. DNA concentration was determined using the QuantiT ™ PicoGreen® dsDNA Assay Kit (Invitrogen). Extracted DNA was sent to the Cornell Institute for Genomic Diversity to conduct GBS.

Genotyping by sequencing protocol and SNPs calling

GBS (Elshire *et al.* 2011) is a simple technique for constructing reduced representation libraries for the Illumina sequencing platform and is conceptually similar to RAD sequencing (Hohenlohe *et al.* 2010). Briefly, DNA from each individual was separately digested using the restriction enzyme EcoT22I (New England Biolabs, Ipswitch, MA), a 6bp cutting enzyme previously shown to work well for populations of unknown structure and highly heterozygous materials (Chen *et al.* 2013). The fragmented DNA was then ligated to a barcoded adaptor and a common adaptor, with appropriate sticky ends, by adding T4 ligase (New England Biolabs). The digestion and ligation were carried out in a 96-well plate. The wells each contained DNA from a different individual and a barcoded adaptor unique to that well. One control well did not contain any DNA. After ligation, samples were combined (5 µL each) and purified using a commercial kit (QIAquick PCR Purification Kit; Qiagen, Valencia, CA) to form a library. The library was then subjected to a PCR, using long primers that

matched the barcoded and common adaptors. The PCR has two functions. One is to perform a size-selection step, as the PCR preferentially amplifies fragments of an ideal length for Illumina sequencing. The second is that the long primers add a length of sequence to the fragments in the library. These sequences bind to the Illumina flow cell and are also used to prime subsequent DNA sequencing reactions. After PCR, the library was cleaned again using a Qiagen QIAquick PCR Purification kit. Libraries were purified as above and fragment sizes evaluated on an ExperionH automated electrophoresis station (BioRad, Hercules, CA). Single-end sequencing of one 96-plex library per lane was performed on an Illumina HiSeq instrument with 100 bp read chemistry.

Illumina data files were filtered to individual genotypes using the Universal Network Enabled Analysis Kit (UNEAK) pipeline (Lu *et al.* 2013), which is available as part of TASSEL 3.0 (Bradbury *et al.* 2007), by using standard parameters.

96 SNPs panel selection

Filtering of the VCF file resulting from the UNEAK pipeline was performed with the SNP & Variation Suite Resources (SVS 8.0.1, 2013-11-14; Golden Helix, Inc., Bozeman, MT) software. To remove those markers with more than 20% of missing data, SNPs were filtered out by call rate < 0.80. Because some populations were represented by only one or few individuals, the Filter Samples by Call Rate function was not applied. Minimum minor allele frequency (MAF) was set to 0.05. To filter out potential paralogs, following White *et al.* (2013) we discarded loci with a mean observed heterozygosity > 0.75, calculated with GenAlEx (Peakall & Smouse 2012).

Loci were tested for linkage disequilibrium by using software Genepop version 4.3 (Rousset 2008). After removal of linked markers, in order to get a first estimation of the number of populations (K), we performed a Principal Component Analysis (PCA) and a Discriminant Analysis of Principal Components (DAPC) with R package Adegenet (Jombart 2008; Jombart & Ahmed 2011), and a parallel run with Admixture (Alexander *et al.* 2009). ADMIXTURE program uses a cross-validation approach to help estimating which value of K has the best predictive value (Liu *et al.* 2013). ADMIXTURE focuses on maximum likelihood estimation (MLE) and calculates the estimates via a block relaxation approach which results in improvements in speed. As preliminary analysis, a number of K corresponding to the number of geographic sampling localities was tested (see Figure 1). The number of genetic populations was selected among those resulting from Admixture, choosing the K with the lower cross validation error.

In order to select the best panel of 96 SNPs, able to return the same results obtained with the whole SNPs panel, we scored the 637 loci separately according to three different indexes: pairwise Fst, locus Informativeness for ancestry (In) (Rosenberg *et al.* 2003), and PCA loadings. Pairwise Fst values among the resulting number of populations (obtained from PCA analysis, DAPC and Admixture) and by locus were assessed by using the software SVS. Locus Informativeness was estimated by the R package diveRsity version 1.7.6 (Keenan *et al.* 2013).

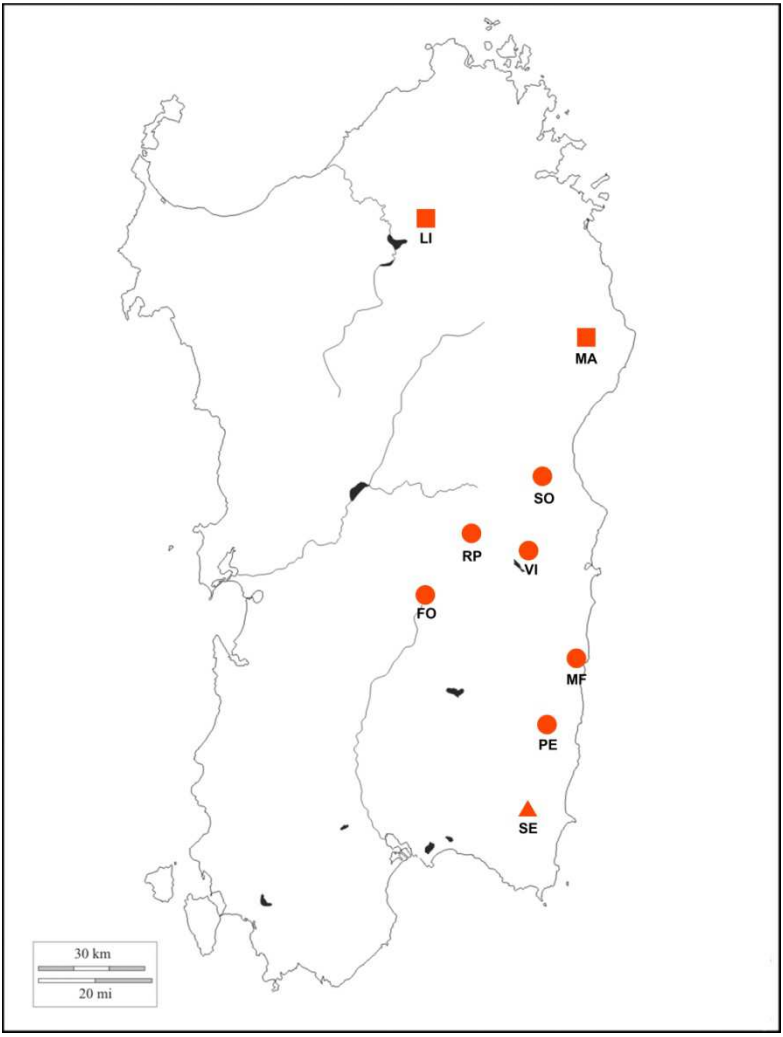


Figure 1. Location of sampling sites in Sardinia. LI = Limbara, MA = Monte Albo, SO = Supramonte di Oliena, RP = RoaPaolinu, VI = Villagrande, FO = Foresta Ortuabis, MF = Monte Ferru, PE = Perdasdefogu, SE = Settefratelli. Sites in the putative North are marked with squares, those in the putative Centre are marked with circles and the one in the South with a triangle.

Estimates of the PCA loadings were performed by R package Adegenet (Jombart 2008; Jombart & Ahmed 2011).

The first 96 SNPs for each index were then selected to create three different panels, named respectively 96Loci_Fst (based on the first 96 SNPs with maximum Fst value), 96Loci_Info (based on the first 96 SNPs with maximum Informativeness value) and 96Loci_Load (based on the first 96 SNPs with maximum Loadings values from PCA analysis). A fourth dataset, named 96Loci_Shared, was produced selecting the first 96 (within 637) shared loci among the three indexes. Association among Fst, Informativeness and Loadings values was assessed with a Spearman Correlation Test, performed in R. According to the results of the correlation test, a fifth panel named 96Loci_Mixed was created. Since Fst significantly correlated with Informativeness (r^2 = 0,4), we chose the first 32 markers from each index in the following order: Fst, Loadings and Informativeness. All the five panels were then used to perform further analysis with software Admixture. In order to test the resulting populations for Hardy-Weinberg equilibrium, we used software GenAlEx, applying the Bonferroni correction.

Results

Data quality and coverage

Illumina sequencing of 95 individuals on two lanes resulted in 501 580 539 reads. Of these, 451 654 647 were good barcoded reads (Glaubitz *et al.* 2014), that were used in the UNEAK pipeline. The UNEAK pipeline identified 156 753 biallelic SNP loci, with an average coverage per locus of 2.15×, and an average coverage per individual of 2.25x. When loci with more than 20% of missing data, minor allele frequency (MAF) < 0.05 and observed heterozygosity > 0.75 were excluded, 749 loci were retained, with an average coverage per locus of 11.83x and an average coverage per individual of 11.85x.

96 SNPs panel selection

The analysis of linkage disequilibrium found that 112 pairs were highly significant, returning 637 unlinked SNPs.

Results of PCA analysis show that samples are divided in three main clusters, as represented in Figure 2. Looking at the graph from left to right, the first cluster is represented by all the populations of the centre. A second cluster, in the right corner at the top, comprises the two populations in the north, Monte Limbara and Monte Albo. At the bottom of the figure there is the southern population of Settefratelli. However, looking at the first cluster, we can see that there are 2 central points, and that samples were separated by two gradations of colors, revealing the existence of two sub-population in the central cluster. Summarizing, the principal component analysis reveals the presence of four clusters: cluster 1, comprising the two populations in the north (Monte Limbara and Monte Albo); cluster 2, comprising six populations in the

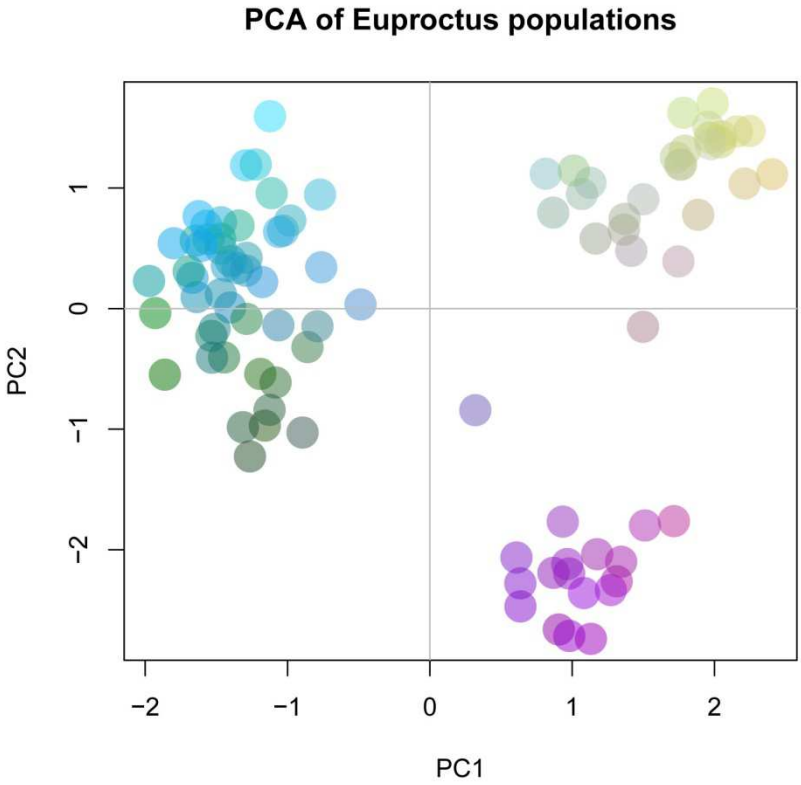


Figure 2. Principal Component Analysis results of *Euproctus platycephalus* populations with 637 SNPs loci. Blue and green dots in the upper-left corner of the graph represent all the populations comprised in the putative centre (SO, RP, VI, FO, MF,PE), yellow-grey dots in upper right corner represent the two populations in the putative north (LI, MA), while violet dots in the bottom right part of the graph represent the single population in the putative south (SE).

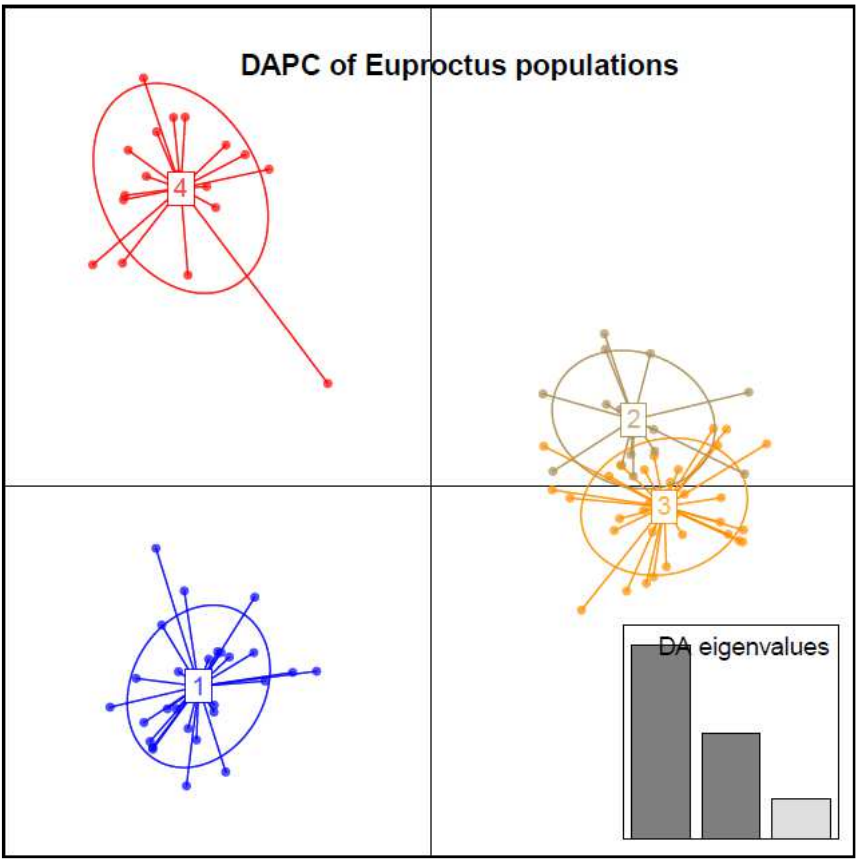


Figure 3. Discriminant Analysis of the Principal Component for *Euproctus platycephalus* populations with 637 loci. Cluster number 1 (blue color) represents the two populations in the north (LI, MA), cluster number 2 (tan color) represents two out six populations in the centre (MF, PE), cluster number 3 (orange color) represents four out six of the populations in the centre (SO, RP, VI, FO) and cluster number 4 (red color) represents the single population in the south (SE).

Table 1. Synopsis of Admixture results. For each panel the best K, and the clusterization obtained for K3 and K4 are reported.

Panel	Best K	K 3clusterization	K 4clusterization
637 loci	K 4	North-South Centre1 Centre2	North Centre1 Centre2 South
96 loci – Fstmax	K 4	North Centre South	North Centre1 Centre2 South
96 loci – Informativeness	K 3	North Centre South	North Centre1 Centre2 South
96 loci – Loadings	K 3	North Centre South	North Centre1 Centre2 South
96 loci – Shared	K 3	North Centre South	North Centre1 Centre2 South
96 loci - Mixed	K 4	North Centre South	North Centre1 Centre2 South

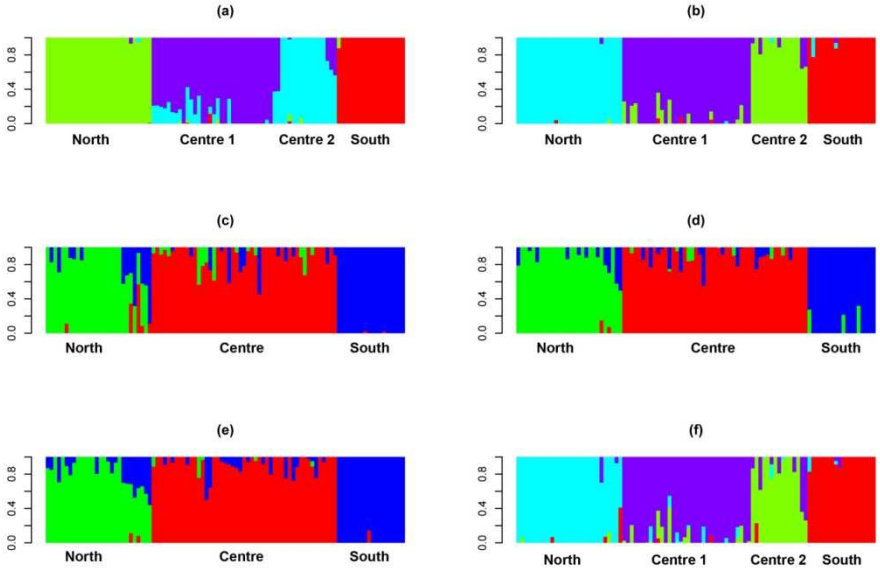


Figure 4. Admixture results of the best K for each panel. (a) Results for the whole panel with 637 loci, best K = 4; (b) Results for the 96Loci_Fst panel, best K = 4; (c) Results for the 96Loci_Info panel, best K = 3; (d) Results for the 96Loci_Load panel, best K = 3; (e) Results for the 96Loci_Shared panel, best K = 3; (f) Results for the 96Loci_Mixed panel, best K = 4.

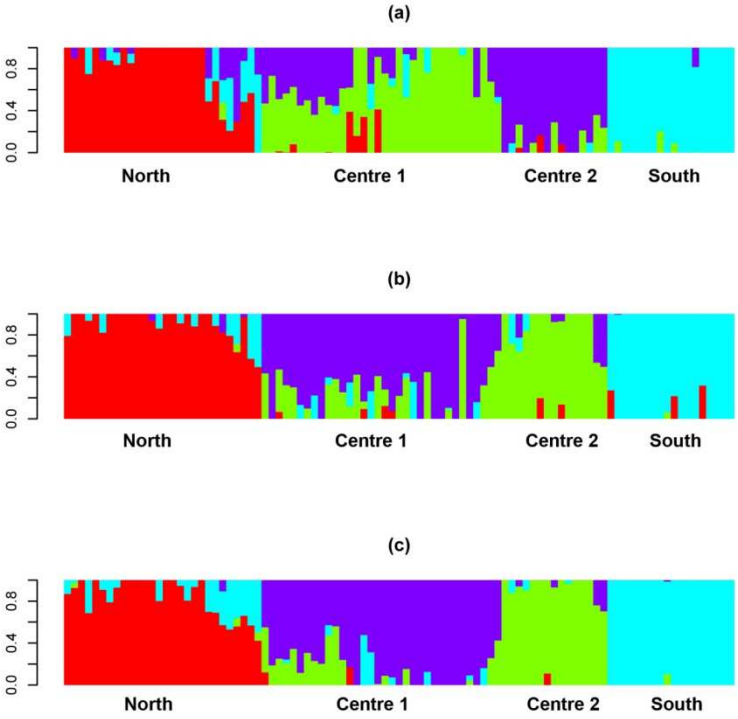


Figure 5. Admixture results with K = 4 for: (a) the 96Loci_Info panel; (b) the 96Loci_Load panel; (c) the 96Loci_Mixed panel.

centre (Pischina Urtaddala, Su Cunnu ‘e s’ebba, Codula Orbisi, Roa Paolinu, Villagrande, Foresta Ortuabis); cluster 3, represented by the two south-easternmost populations in the centre (Monte Ferru and Perdasdefogu); cluster 4, represented by the single population in the south (Settefratelli). DAPC grouping, reported in Figure 3, shows the same outcomes of the PCA analysis, dividing samples in 4 groups. Group 1 represents the population in the south, group 2 and 3 the two populations in the centre, and group 4 the population in the north. According to the values of the cross validation error, Admixture (run with 637 loci) returns 4 as best K value, reflecting the same subdivisions obtained by PCA and DAPC (Figure 4a).

The correlation test among Fst, Informativeness and Loading loci returned the following results: for Fst Vs Informativeness R = 0.4, for Informativeness Vs Loadings R = 0.03 and for Fst Vs Loadings R = -0.05.

Table 1 summarizes the results of Admixture analyses performed with the whole SNPs dataset and with the five 96 SNPs sub-panels. For three of them - 96Loci_Info, 96Loci_Loadings and 96Loci_Shared - the best K value resulted 3, while for 96Loci_Fst and 96Loci_Mixed the best number of populations was 4. Looking at the clusters, we can observe that the first separation (K = 3) identified by the panel of 637 loci is among the two populations in the centre and the group north-south. All the other panels identified the groups of K 3 as North – Centre – South. Taking into consideration the column regarding K 4 results, all clusters show the same subdivision for each panel (see also Figure 4). All the four suggested populations finally resulted in Hardy-Weinberg equilibrium. Using the 637 loci panel as reference, and comparing the K 4 related results, we chose 96Loci_Fst as the best sub-panel.

Discussion

Our study represents the first application of the GBS technique on *Euproctus platycephalus*. As mentioned, the analysis of *Euproctus* genome presented an intrinsic difficulty not only for its huge dimension, but also because of its state of non-model species. However, we benefited from the GBS’s ability in reducing genome complexity, obtaining sufficient overlap in sequence coverage. Our results not only show that GBS is a useful method for SNPs discovery also in large genome amphibian species, but especially proved their informativeness and ability in identifying genetic patterns based on our data. Since all the statistical analyses performed with the whole SNPs dataset (637 loci) returned the same number of genetic clusters (n = 4), this outcome suggest a fair reliability of SNPs markers.

We suggest that among the three tested indexes to select 96 loci, for our purpose, the Fst would be the most informative. In fact the clusterization obtained with the 96Loci_Fst panel (K = 4) reflects that obtained with the whole SNPs panel (Figure 4a,b), thus being a valuable subset of SNPs containing most of information on broad genetic distinction among groups.

On the other hand, the remaining two other indexes, Informativeness and Loadings, suggested the presence of only three clusters (K = 3, Figure 4c,d) in which, differently from the Fst clustering, all central populations are grouped together.

Comparing the Admixture results of K 4 for Fst (Figure 4b), Informativeness and Loadings (Figure 5), it appears that also the latter two indexes returned the same clusterization, but admixture among individuals is more evident. We observed similar results for the fourth panel (96Loci_Shared; Figure 4e and FigureX). As for the fifth panel, 96Loci_Mixed, Admixture result (Figure 4f) reflects almost exactly the one obtained for the panel 96Loci_Fst, as the first 32 loci selected for this panel are those with the highest Fst value (Material and methods section). For this reason we propose that the more efficient and practical way to select informative loci in *E. platycephalus* would be to score them directly according to Fst values.

More in detail, average Fst value for all the 637 loci was 0.072, while it was 0.264 for the 96 selected loci, ranging from a maximum of 0.756 to a minimum of 0.175. Although no genome scan was applied so far to identify Fst outliers, as suggested by Funk *et al.* (2012), we very likely already included them in our selected SNPs. In order to deepen the present outcomes, and identify not only the ESUs but also MUs and adaptive groups, future statistical analyses have to be performed.

Loci with the highest Fst values possibly have adaptive value, or may reflect structuring due to genetic drift. Interestingly in this study the two groups of central populations, detected only on the base of SNPs with the highest Fst value (or with the whole SNPs panel), very likely are subjected to different ecological settings. In fact the populations of the first group (Centre1) are located in the Gennargentu Mountains characterized by continental climate condition, while one of the two populations of the other group (Centre2) is located in a Mediterranean habitat very close to the sea. Assessing if groupings based on highest Fst values in *E. platycephalus* correlate with those subjected to different ecological conditions would be of outmost importance in the case of individuals' reintroduction. This because filling out a declining population with individuals belonging to a population adapted to a very different environment, would produce a worst situation, leading to outbreeding depression (Allendorf & Luikart 2007).

Due to official schedules imposed by the *Euproctus* conservation program, this prelimimary identification of ESUs was already used to collect individuals from field belonging to these main genetic groups and start captive breeding activities aimed to future reintroduction programs. Then, further steps will be needed to distinguish neutral loci from the adaptive ones, in order to better understand the adaptive potential of the Sardinian Brook Salamander populations, delineate ESUs and Management Units (MUs), and improve planning of conservation strategies.

CHAPTER 5

Individual heterozygosity and demographic estimates by GBS in a non-model amphibian species:preliminary results

Background

The conservation of genetic variability is of crucial importance for the survival of natural populations. In particular, this feature is fundamental for small and isolated populations, as a decrease in genetic diversity can be an indicator of increasing inbreeding. One of the most used measure to estimate the amount of genetic variation within different populations is heterozygosity (Allendorf & Luikart, 2007). Heterozygosity is usually computed as the sum of proportion heterozygous at all loci / total number of loci sampled. It can be a useful parameter for understanding how natural selection is acting on contemporary populations (Szulkin *et al.* 2010). One of the most used approaches has been to seek to relate individual heterozygosity with variation in characters that are potentially related to fitness (Chapman *et al.* 2009). These kinds of approaches take the name of Heterozygosity-Fitness-Correlations studies (HFC; Balloux *et al.* 2004). Since quantifying the relationship between molecular and phenotypic variation is often complex, another less-demanding method is to study the statistical association between molecular genotypes and traits under selection (Szulkin *et al.* 2010).

Recent advantages in genetic techniques, such as the development of Next Generation Sequencing (NGS), together with the increased focus on long-term population studies, have lead to a great improvement in HFC related works. The major advance offered by NGS is the ability to produce an enormous volume of data cheaply (Metzker 2010): as the number of markers increases, also the accuracy of parameters' estimation can increase. Evolutionary theory would suggest that we should only expect correlations of genetic diversity with fitness-related traits because dominance variance is expected to be high for traits with a direct effect on fitness, and such traits have a more complex genetic architecture (Chapman *et al.* 2009). From an evolutionary perspective, one of the characters that limits the reproductive potential of an individual and should as such be opposed by natural selection is ageing (Zwaan 1999).

The target species of the present work, the Italian stream frog (*Rana italica*), is an anuran species endemic to peninsular Italy. It is distributed in the western side of the Italian peninsula, ranging from 100 m to over 1500 m a.s.l. The species is strictly bounded to rivers, creeks and streams with perennial water, usually located in woodlands. In the Latium region this species is facing a slow decline since the beginning of the'70s, mostly due to water pollution and habitat fragmentation (Bologna *et al.*, 2000). It is protected by the Bern Convention (Annex II) and by the Habitat Directive (92/43/CEE, Annex IV-D). A previous study, carried out with allozymes and mitochondrial markers, revealed for the Italian stream frog a genetic pattern matching the "southern richness – northern purity" pattern, along the whole

distribution area (Canestrelli *et al.* 2008). In particular, in the Latium region, the authors found a very low genetic variability at both markers.

Rana italica is object of a long-term ecological study, for which mark and recapture and skeletochronology data have already been collected. In order to combine ecological and genetic data, with the purpose of deepen the actual knowledge about the species, we aimed at investigating its actual genetic structure at a regional scale. To validate the results obtained by previous analyses, which found no genetic variability, we needed markers with a greater resolution power. Since no microsatellite markers were available, we used the novel technique of Genotyping By Sequencing (GBS). The GBS protocol is a multiplexed, high-throughput, and low-cost method to explore the genetic diversity in populations (Elshire *et al.* 2011). It employs a reduced representation library (RRL) strategy (Altshuler *et al.* 2000) to target a fraction of the genome for sequencing, thereby decreasing cost and increasing the SNP-calling accuracy. GBS is the simplest of the RRL approaches developed thus far (Davey *et al.* 2011), and has already seen extensive application in a several taxa, i.e., in barley and wheat (Poland *et al.* 2012), maize (Elshire *et al.* 2011; Hansey *et al.* 2012). The available genomic resources for amphibians are limited, with a single genome available (Hellsten *et al.* 2010) few transcriptomes, and virtually no population genomic datasets for this group.

Summarizing, the preliminary aims of this work were to provide the results of the first application of GBS technique on the target species, and then discover how many SNPs are needed to obtain a reliable estimate of individual heterozygosity.

Materials and methods

Sampling and DNA extraction

Samples were collected from 7 different sites in Latium region (Italy), between spring 2010 and spring 2012. We collected samples as follow: 14 individuals from Veio, 13 individuals from Insugherata, 13 individuals from Cannuccete, 14 individuals from Tolfa, 14 individuals d from Licenza, 13 individuals from Monterano and14 individuals from Cineto (Figure 1). Tissues samples were collected by toe clipping of adult individuals, and were immediately stored in ethanol 95% at -20 °C, for further genetic analyses. Tissues were digested using a proteinase K solution (56 °C - overnight), and the total genomic DNA was extracted using a DNeasy® Blood & Tissue Kit (Qiagen – Cat. No. 69506). A plate of 96 samples (with one blank) was send to the Cornell Institute for Genomic Diversity to conduct GBS (Elshire *et al.* 2011).

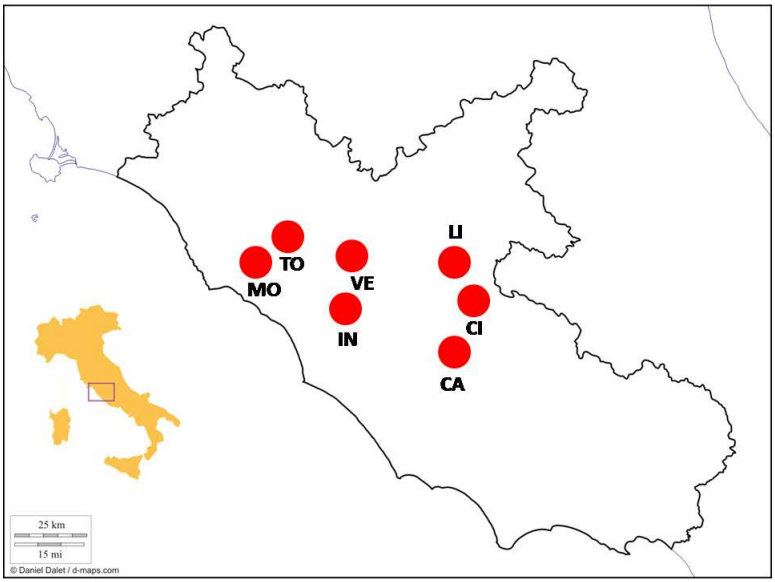


Figure 1. Location of sampling sites in Latium region. MO = Monterano, TO = Tolfa, VE = Vejo, IN = Insugherata, LI = Licenza, CI = Cineto, CA = Cannuccete.

Genotyping by sequencing protocol and SNPs calling

GBS (Elshire *et al.* 2011) is a simple technique for constructing reduced representation libraries for the Illumina sequencing platform and is conceptually similar to RAD sequencing (Hohenlohe *et al.* 2010). Briefly, DNA from each individual was separately digested using the restriction enzyme EcoT22I (New England Biolabs, Ipswitch, MA), a 6bp cutting enzyme previously shown to work well for populations of unknown structure and highly heterozygous materials (Chen *et al.* 2013). The fragmented DNA was then ligated to a barcoded adaptor and a common adaptor, with appropriate sticky ends, by adding T4 ligase (New England Biolabs). The digestion and ligation were carried out in a 96-well plate. The wells each contained DNA from a different individual and a barcoded adaptor unique to that well. One control well did not contain any DNA. After ligation, samples were combined (5 µL each) and purified using a commercial kit (QIAquick PCR Purification Kit; Qiagen, Valencia, CA) to form a library. The library was then subjected to a PCR, using long primers that matched the barcoded and common adaptors. The PCR has two functions. One is to perform a size-selection step, as the PCR preferentially amplifies fragments of an ideal

length for Illumina sequencing. The second is that the long primers add a length of sequence to the fragments in the library. These sequences bind to the Illumina flow cell and are also used to prime subsequent DNA sequencing reactions. After PCR, the library was cleaned again using a Qiagen QIAquick PCR Purification kit. Libraries were purified as above and fragment sizes evaluated on an ExperionH automated electrophoresis station (BioRad, Hercules, CA). Single-end sequencing of one 96-plex library per lane was performed on an Illumina HiSeq instrument with 100 bp read chemistry.

Illumina data files were filtered to individual genotypes using the Universal Network Enabled Analysis Kit (UNEAK) pipeline (Lu *et al.* 2013), which is available as part of TASSEL 3.0 (Bradbury *et al.* 2007), by using standard parameters.

SPNs filtering and heterozigosity analisys

Filtering of the VCF file resulting from the UNEAK pipeline was performed with the SNP & Variation Suite Resources (SVS 8.0.1, 2013-11-14; Golden Helix, Inc., Bozeman, MT) software. First of all, individual with high failure rates (i.e. call rates < 0.30 = missingness >70%) were filtered from the database. The genotypes' table was then filtered for SNPs with call rate < 0.80 (i.e. those markers with more than 20% of missing data) and minimum minor allele frequency (MAF) was set to 0.05.

In order to investigate for the presence of a genetic structure among groups, we used software TASSEL (Bradbury *et al.* 2007) to perform a neighbour joining (NJ) analysis, and the R package ADEGENET (Jombart 2008) to perform a Principal Component Analysis (PCA). The number of genetic was also tested with software ADMIXTURE (Alexander *et al.* 2009). In order to further test the previous resulted clusterizations, deviation from Hardy-Weinberg equilibrium was tested firstly considering all samples as belonging to a single population, and then dividing samples according to the results of NJ, PCA and ADMIXTURE.

Estimates of individual heterozygosity were performed using the R package Rhh (Alho *et al.* 2010). Rhh provides three estimates of individual multilocus heterozygosity: standardized heterozygosity (SH - (Coltman *et al.* 1999), internal relatedness (IR - (Amos *et al.* 2001) and homozygosity by loci (HL - (Aparicio *et al.* 2006). The standardized heterozygosity index was formerly developed to measure the heterozygosity of individuals on an identical scale, presuming that not always all the analyzed individuals are genotyped together and with the same panel of markers. In particular, SH is given by the proportion of heterozygous typed loci divided the mean heterozygosity of typed loci (Coltman *et al.* 1999). Internal relatedness, instead, returns an estimate of the connection among the analyzed individuals. This index is very similar to the *r*-values, as it shows negative values when the individuals are not inbreed, and positive values when the individuals are very likely born from related parents (Amos *et al.* 2001). The IR value can vary between -1 and 1; the first event can occur only when all loci have two alleles and individuals are heterozygous for all of them, while the second one happens when all loci are homozygous, independently from allelic frequencies. Homozygosity by loci improves the estimate made by the

internal relatedness, taking into account the contribution of each locus to the final homozygosity index, instead of the contribution of each allele (Aparicio *et al.* 2006). In addition, Rhh package calculates the heterozygosity-heterozygosity correlation, which can be considered as an index of the reliability of the set of markers (Alho *et al.* 2010; Balloux *et al.* 2004). Briefly, the correlation is calculated by creating random pairs of half sets of markers, and then estimating the individual multilocus heterozygosity for each set. The mean values between each pair of sets returns the heterozygosity-heterozygosity correlation (Alho *et al.* 2010). When the value of this correlation is positive and significant the estimates reflect a signature of inbreeding.

In order to identify the minimum number of loci needed to obtain a reliable estimate, we used different SNPs subsets to calculate individual heterozygosity. In particular, we estimated the overall individual heterozygosity with 12 subsets of loci. The smallest subset was of 5 loci, while the biggest one was represented by the whole SNPs panel (Figure 2). Briefly, we created a custom R script for the random selection of loci. Taking as example the subset of 5 loci, we used the script to select 5 different loci for each sample, and then we estimated the individual heterozygosity for each one. These steps were repeated for each SNPs subset, and the distribution of our results was plotted in box-and-whisker diagrams (Figure 2).

Preliminary results and discussion

Illumina sequencing of 95 individuals on one lane resulted in 269 695 656 reads. Of these, 181 237 253 were good barcoded reads (Glaubitz *et al.* 2014), that were used in the UNEAK pipeline. The UNEAK pipeline identified 20 399 biallelic SNP loci, with an average coverage per locus of 7.12×, and an average coverage per individual of 6.88x. When loci with more than 20% of missing data and minor allele frequency (MAF) < 0.05, 854 loci were retained, with an average coverage per locus of 15.14x and an average coverage per individual of 14.72x. After the application of individual filtering, 14 samples were excluded from further analyses.

Results of the neighbor joining analysis show that samples are divided in two main clusters, with a further subdivision of one of them, as represented in figure 3. In particular, the populations of Tolfa, Monterano, Vejo and Insugherata represent a first cluster, with a further split between the group Tolfa-Monterano versus the group Vejo-Insugherata. The second cluster is constituted by the populations of Cineto, Licenza and Cannuccete. PCA grouping shows similar result on respect to the previous analysis, as reported in figure 4. The first cluster at the top-left corner of the graph (gree-blue dots) represents the group of Cineto-Licenza-Cannuccete. The second cluster on the right of the graph (orange-lilac dots) comprises all the other populations, that are Tolfa, Monterano, Vejo and Insugherata. ADMIXTURE results suggest that the most reliable number of populations is two, as shown in figure 5. After having tested the whole SNPs panel for the Hardy-Weinberg equilibrium, we found that 622 loci were in equilibrium (72%), while 232 were not (28%). Applying the same test on samples divided them in two populations we found that in the first population 42 loci were monomorphic, so they have been excluded from the analysis, and that 730 loci

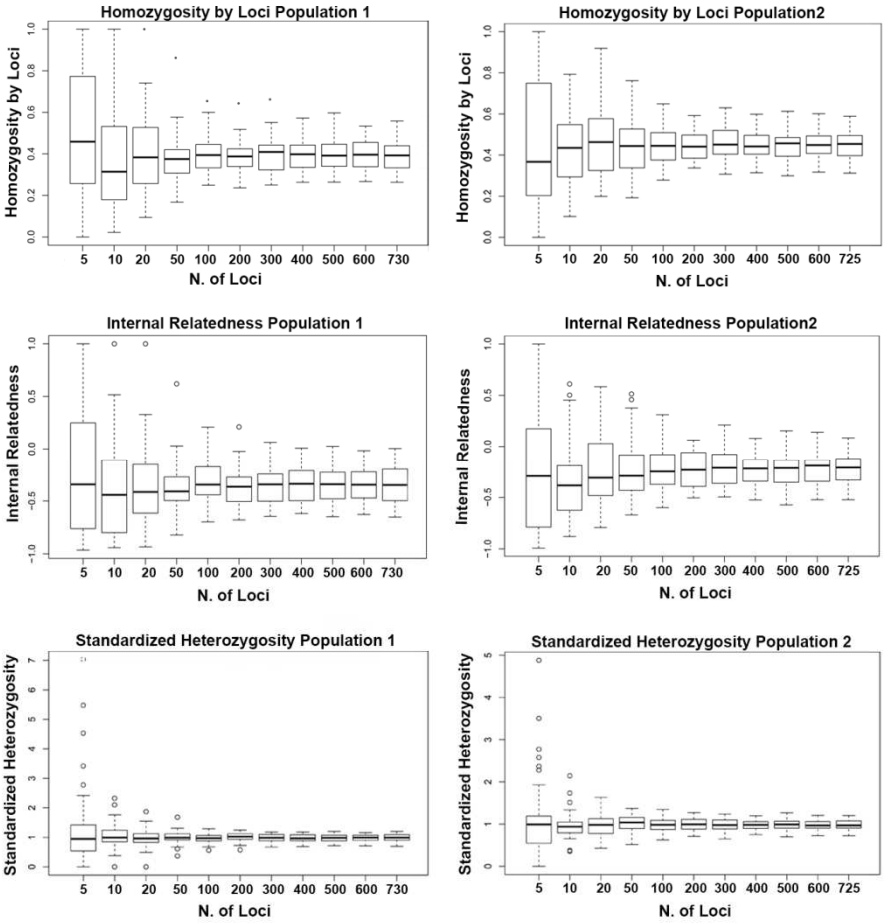


Figura 2. Box-plots graphs summarize the results of the individual heterozygosity analyses for the two populations of *Rana italica.* On the left column results for the first population (Tolfa-Monterano-Vejo-Insugherata) are reported. On the right column are shown the results for the second population (Cineto-Licenza-Cannuccete).

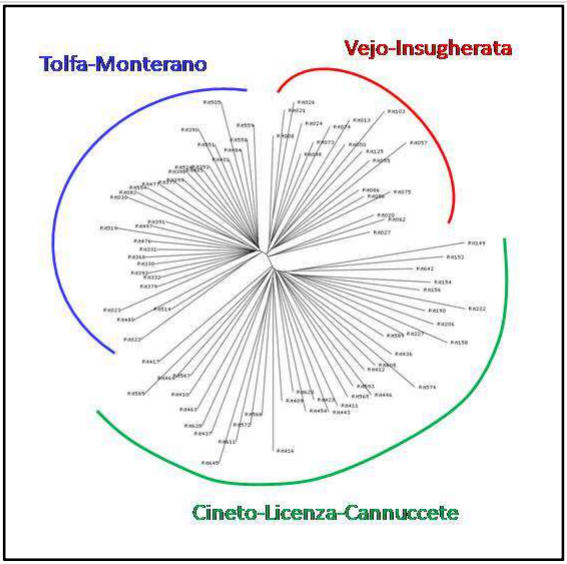


Figura 3. Neighbour joining tree of *Rana italica* populations. Samples are divided in three clusters: blue color highlights the samples belonging to the group Tolfa-Monterano, red color highlights the samples belonging to the group Vejo-Insugherata and the green color highlights samples belonging to the group Cineto-Licenza-Cannuccete.

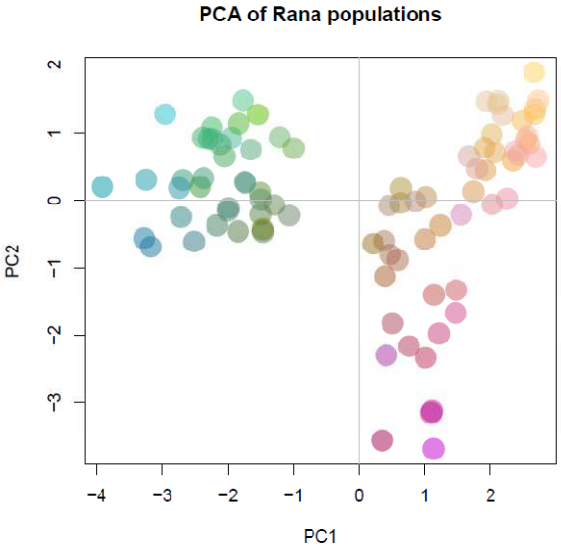


Figure 4. Principal Component Analysis results of *Rana italica* populations with 854 SNPs loci.

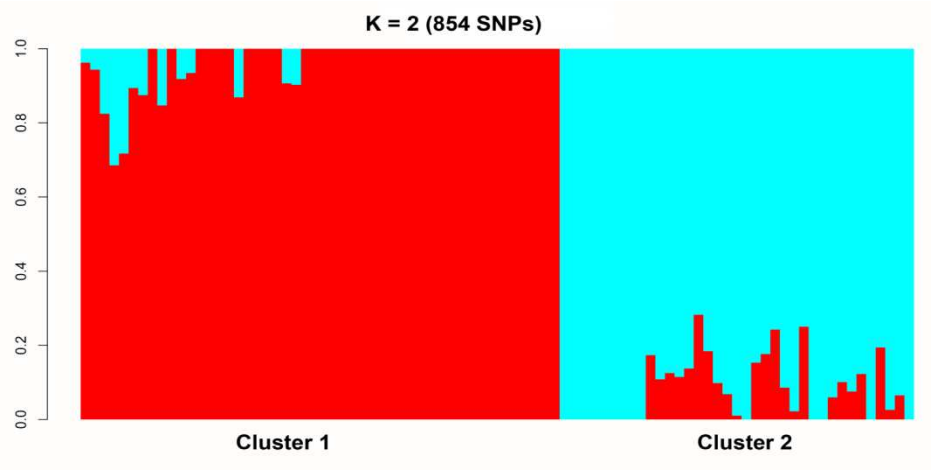


Figura 5. Admixture results for K 2 with 854 SNPs loci. Cluster 1 represents the populations of Tolfa, Monterano, Vejo and Insugherata, while Cluster 2 represents those of Cineto, Licenza and Cannuccete.

were in equilibrium (90%) while 82 were not (10%). In the second population 30 loci were monomorphic (and they have been excluded from further analyses), and the loci in equilibrium and not in equilibrium resulted 725 (88%) and 99 (12%), respectively. Performing a chi-quadro test the differences between the number of loci in equilibrium considering one or two populations resulted statistically significant ($p<0.01$). Outcomes of the individual heterozigosity estimates are reported in figure 2. In both populations the three multilocus heterozygosity estimates (SH, HL, IR) followed the same trend, reaching the asymptote around the value of 200 loci.

The main outcomes of this preliminary work have been the discovery of a genetic structure among the studied populations and the identification of the minimum number of loci needed to perform a reliable estimate of multilocus individual heterozygosity. Further analyses will be addressed at investigating the genetic clusterization in detail, in order to both confirm the resulting groups. Moreover, future research directions will comprise Heterozygosity-Fitness-Correlations analyses, in order to investigate for the presence of some kind of relationships between individual age and individual heterozygosity. Finally, demographic estimates will be performed by using the novel resulting SNPs, and we'll tested our populations for the presence of gene flow, trying also to apply a landscape genomic approach.

CONCLUSIONS

Stealing the words of Brito & Edwards (2009), and rearranging them for the occasion: "these are exciting times to be working in the fields of" conservation genetics and genomics. The human being has always been curious about the world around it, but often the tools to understand natural phenomena were not appropriate to satisfy this need. On the contrary, we are now living in an unprecedented historic period where technological improvements made us able to investigate almost every aspect of natural life. In particular, the field of genetics has been revolutionized by recent advancements. Indeed, the advent of Next Generation Sequencing techniques has lead to the ability of sequencing whole genomes in a short time and with a low cost, providing the chance to obtain thousand of novel genome-wide markers for virtually every species. Actually, the main issue for a researcher in this field is no longer the possibility or not of obtaining data, but, on the contrary, questioning on which is the best approach/technique to use.

Having the availability of such a huge number of genetic techniques, I tried to approach an extensive topic, such as that of conservation genetics, from different points of view, and with different genetic/genomic tools.

Along this PhD project the analyses of the DNA *sensu lato*, proved to be a unique tool to get insights at different research scales. In a conservative perspective, genetic and genomic markers revealed to be useful to answer vey wide questions, concerning for example evolutionary biology, with processes developing even at a very short temporal and spatial scale.

Starting from a small scale such as single-population level, microsatellite markers proved to be very informative and reliable for investigating the reproductive strategy of *Salamandrina perspicillata*, and to perform parentage analyses. By means of this tool, I was able to clarify the mechanism beyond the observed pattern of polyandry, and to find evidences of a female choice, based on female genetic dissimilarity from males, this providing indirect benefit to the offspring (increased heterozygosity). From a conservative point of view these findings are essential to understand how individuals can increase their chance of transmitting genetic benefits to the future generations.

On the other hand single nucleotide polymorphisms (SNPs) showed to be very powerful markers for studying natural populations' genetic structure on a wider geographic scale. They allowed me to clarify the pattern of genetic diversity among several populations of both *Euproctus platycephalus* and *Rana italica.*

Switching from *S.perspicillata* to *R.italica*, and moving my investigation to a larger scale, I was able to estimate a very important parameter in conservation genetics, such as individual heterozygosity, at a regional scale (Latium region). Individual heterozygosity not only is a measure *per se* of genetic variability within and among populations, and indirectly of inbreeding; but also it can be used to perform Heterozygosity-Fitness-Correlations. I'm planning to apply this kind of analysis in order to investigate for a possible correlation between the heterozygosity level within different populations and individual age.

The following step was an analysis at an even wider geographic scale, the whole distribution range (Sardinia Island) of an endangered species, *Euproctus platycephalus*. In this case, due to an urgent conservation need, I focused on identifying Evolutionary Significant Units. Since this analysis is based on the use of both neutral and adaptive loci, this first step is crucial for maximize the potential of such units for adapting to future environmental changes. In the future I will be able to use those loci separately to detect also Management Units and adaptive groups.

An additional significant result of this work was the production of the first genome-wide markers (SNPs) for *Euproctus platycephalus* and *Rana italica*,and the simultaneous validation of Genotyping By Sequencing applicability also on large amphibian genomes.

In conclusion this study gave a contribution, at different scales and with different approaches, to the conservation of the three target species. Moreover, it globally can give the idea that the choice of the appropriate methodology is essential for approaching specific issues, achieving the identified targets and answering the key questions.

REFERENCES

Adams EM, Jones AG, Arnold SJ (2005) Multiple paternity in a natural population of a salamander with long-term sperm storage. Molecular Ecology 14: 1803-1810.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19: 1655-1664.

Alford RA, Richards SJ (1999) Global amphibian declines: A problem in applied ecology. Annual Review of Ecology and Systematics 30: 133-165.

Alho JS, Valimaki K, Merila J (2010) Rhh: an R extension for estimating multilocus heterozygosity and heterozygosity-heterozygosity correlation. Molecular Ecology Resources 10: 720-722.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nature Reviews Genetics 11: 697-709.

Allendorf FW, Luikart GH (2007) Conservation and the genetics of populations. John Wiley & Sons.

Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407: 513-516.

Amos W, Wilmer JW, Fullard K, Burg TM, Croxall JP, Bloch D, Coulson T (2001) The influence of parental relatedness on reproductive success. Proceedings of the Royal Society B-Biological Sciences 268: 2021-2027.

Andersson S (1994) Costs of sexual advertising in the Lekking Jackson's Widowbird. Condor 96: 1-10.

Angelini C, Vanni S, Vignoli L (2007) *Salamandrina terdigitata* e *S. Perspicillata*. In: Fauna d'Italia, Amphibia. Vol. XLII. B. Edizioni Calderini, Bologna.

Aparicio JM, Ortego J, Cordero PJ (2006) What should we weigh to estimate heterozygosity, alleles or loci? Molecular Ecology 15: 4659-4665.

Arnqvist G, Nilsson T (2000) The evolution of polyandry: multiple mating and female fitness in insects. Animal Behaviour 60: 145-164.

Baillie JEM, Griffiths J, Turvey ST, Loh J, Collen B (2010) Evolution Lost: status and trends of the world's vertebrates. Zoological Society of London. London.

Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? Molecular Ecology 13: 3021-3031.

Bateman PW (1998) Mate preference for novel partners in the cricket *Gryllus bimaculatus*. Ecological Entomology 23: 473-475.

Bateson P (1983) Optimal outbreeding. In Mate choice 257-277. Cambridge University Press.

Beebee TJ, Griffiths RA (2005) The amphibian decline crisis: a watershed for conservation biology? Biological Conservation 125(3): 271-285.

Beebee TJC (2005) Conservation genetics of amphibians. Heredity 95: 423-427.

Birkhead TR, Pizzari T (2002) Postcopulatory sexual selection. Nature Review Genetics 3: 262–273.

Blaustein AR, Kiesecker JM (2002) Complexity in conservation: lessons from the global decline of amphibian populations. Ecology Letters 5: 597-608.

Blaustein AR, Wake DB (1990) Declining amphibian populations - a global phenomenon. Trends in Ecology & Evolution 5: 203-204.

Blaustein AR, Wake DB, Sousa WP (1994) Amphibian declines - Judging stability, persistence, and susceptibility of populations to local and global extinctions. Conservation Biology 8: 60-71.

Bologna MA, Capula M, Carpaneto GM (2000) Atlante degli Anfibi e Rettili del Lazio. Roma. Fratelli Palombi Editori.

Borkowska A, Borowski Z, Krysiuk K (2009) Multiple paternity in free-living root voles (*Microtus oeconomus*). Behavioural Processes 82: 211-213.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Brito PH, Edwards SV (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica 135(3): 439-455.

Brizzi R, Delfino G, Selmi MG, Sever DM (1995) Spermathecae of *Salamandrina terdigitata* (Amphibia, Salamandridae) – Patterns of sperm storage and degradation. Journal of Morphology 223: 21-33.

Bruni G, Romano A (2011) Courtship behaviour, mating season and male sexual interference in *Salamandrina perspicillata* (Savi, 1821). Amphibia-Reptilia 32: 63-76.

Calboli FCF, Fisher MC, Garner TWJ, Jehle R (2011) The need for jumpstarting amphibian genome projects. Trends in Ecology & Evolution 26: 378-379.

Canestrelli D, Cimmaruta R, Nascetti G (2008) Population genetic structure and diversity of the Apennine endemic stream frog, *Rana italica* - insights on the Pleistocene evolutionary history of the Italian peninsular biota. Molecular Ecology 17: 3856-3872.

Carey C, Heyer WR, Wilkinson J, Alford RA, Arntzen JW, Halliday T, Hungerford L, Lips KR, Middleton EM, Orchard SA, Rand AS (2001) Amphibian declines and environmental change: use of remote-sensing data to identify environmental correlates. Conservation Biology 15: 903–913.

Carlsson J (2008) Effects of Microsatellite Null Alleles on Assignment Testing. Journal of Heredity 99: 616-623.

Caspers BA, Krause ET, Hendrix R, Kopp M, Rupp O, Rosentreter K, Steinfartz S (2014) The more the better–polyandry and genetic similarity are positively linked to reproductive success in a natural population of terrestrial salamanders (*Salamandra salamandra*). Molecular ecology 23(1): 239-250.

Caughley G (1994) Directions in conservation biology. Journal of Animal Ecology 63: 215-244.

Chandler CH, Zamudio KR (2008) Reproductive success by large, closely related males facilitated by sperm storage in an aggregate breeding amphibian. Molecular Ecology 17: 1564-1576.

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009) A quantitative review of heterozygosity-fitness correlations in animal populations. Molecular Ecology 18: 2746-2765.

Chen C, Mitchell SE, Elshire RJ, Buckler ES, El-Kassaby YA (2013) Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. Tree Genetics & Genomes 9: 1537-1544.

Cohen LB, Dearborn DC (2004) Great frigatebirds, *Fregata minor*, choose mates that are genetically similar. AnimalBehaviour 68: 1229–1236.

Collins JP, Storfer A (2003) Global amphibian declines: sorting the hypotheses. Diversity and Distributions 9: 89-98.

Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. Evolution 53: 1259-1267.

Cook JA, MacDonald SO (2001) Should endemism be a focus of conservation efforts along the North Pacific Coast of North America? Biological Conservation 97: 207-213.

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. Heredity 93: 504-509.

D'Amen M, Bombi P (2009) Global warming and biodiversity: Evidence of climate-linked amphibian declines in Italy. Biological Conservation 142: 3060-3067.

Darwin CR (1871) The descent of man, and selection in relation to sex. London: John Murray.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12: 499-510.

De Donato M, Peters SO, Mitchell SE, Hussain T, Imumorin IG (2013) Genotyping-by-Sequencing (GBS): A Novel, Efficient and Cost-Effective Genotyping Method for Cattle Using Next-Generation Sequencing. Plos One 8.

Della Rocca F, Vignoli L, Bologna MA (2005) The reproductive biology of *Salamandrina terdigitata* (Caudata, Salamandridae). Herpetological Journal 15: 273-278.

Dizon AE, Lockyer C, Perrin WF, Demaster DP, Sisson J (1992) Rethinking the stock concept – a phylogeographic approach. Conservation Biology 6: 24-36.

Dufresne F, Jeffery N (2011) A guided tour of large genome size in animals: what we know and where we are heading. Chromosome Research 19: 925-938.

Eldredge N (1998) Life in the balance: humanity and the biodiversity crisis. Princeton University Press. Princeton. NJ.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. Nature 452: 169-175.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. Plos One 6.

Feder ME, Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. Nature Reviews Genetics 4: 651-657.

Frankham R (2010) Challenges and opportunities of genetic approaches to biological conservation. Biological Conservation 143: 1919-1927.

Frankham R, Ballou JD, Briscoe DA (2004) Primer of conservation genetics. Cambridge University Press, Cambridge.

Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Molecular Ecology 10: 2741-2752.

Freeland JR, Petersen SD, Kirk H (2011) Molecular Ecology, 2nd Edition Wiley-Blackwell.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution 27: 489-496.

Garner TWJ, Larsen KW (2005) Multiple paternity in the western terrestrial garter snake, *Thamnophis elegans*. Canadian Journal of Zoology-Revue Canadienne De Zoologie 83: 656-663.

Garner TWJ, Schmidt BR (2003) Relatedness, body size and paternity in the alpine newt, *Triturus alpestris*. Proceedings of the Royal Society B-Biological Sciences 270: 619-624.

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. Plos One 9.

Goodman D (1987) How do any species persist? Lessons for conservation biology. Conservation Biology 1: 59-62.

Gopurenko D, Williams RN, McCormick CR, DeWoody JA (2006) Insights into the mating habits of the tiger salamander (*Ambystoma tigrinum tigrinum*) as revealed by genetic parentage analyses. Molecular Ecology 15: 1917-1928.

Gowaty PA (1992) Evolutionary biology and feminism. Human nature (Hawthorne, N.Y.) 3: 217-249.

Gregory TR (2011). Animal Genome Size Database. http://www.genomesize.com

Griffith SC, Owens IPF, Thuman KA (2002) Extra pair paternity in birds: a review of interspecific variation and adaptive function. Molecular Ecology 11: 2195-2212.

Hansey CN, Vaillancourt B, Sekhon RS, De Leon N, Kaeppler SM, Buell CR (2012) Maize (*Zea mays* L.) Genome Diversity as Revealed by RNA-Sequencing. Plos One 7.

Hauswaldt JS, Polok A, Angelini C, Steinfartz S (2012) First microsatellite loci for spectacled salamanders (*Salamandrina perspicillata* and *S. terdigitata*) endemic to the Apennine peninsula. Conservation Genetics Resources 4: 399-402.

Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, Ovcharenko I, Putnam NH, Shu S, Taher L, Blitz IL, Blumberg B, Dichmann DS, Dubchak I, Amaya E, Detter JC, Fletcher R, Gerhard DS, Goodstein D, Graves T, Grigoriev IV, Grimwood J, Kawashima T, Lindquist E, Lucas SM, Mead PE, Mitros T, Ogino H, Ohta Y, Poliakov AV, Pollet N, Robert J, Salamov A, Sater AK, Schmutz J, Terry A, Vize PD, Warren WC, Wells D, Wills A, Wilson RK, Zimmerman LB, Zorn AM, Grainger R, Grammer T, Khokha MK, Richardson PM, Rokhsar DS (2010) The Genome of the Western Clawed Frog *Xenopus tropicalis*. Science 328: 633-636.

Hoffmann M, Hilton-Taylor C, Angulo A, Boehm M, Brooks TM, Butchart SHM, Carpenter KE, Chanson J, Collen B, Cox NA, Darwall WRT, Dulvy NK, Harrison LR, Katariya V, Pollock CM, Quader S, Richman NI, Rodrigues ASL, Tognelli MF, Vie J-C, Aguiar JM, Allen DJ, Allen GR, Amori G, Ananjeva NB, Andreone F, Andrew P, Aquino Ortiz AL, Baillie JEM, Baldi R, Bell BD, Biju SD, Bird JP, Black-Decima P, Blanc JJ, Bolanos F, Bolivar-G W, Burfield IJ, Burton JA, Capper DR, Castro F, Catullo G, Cavanagh RD, Channing A, Chao NL, Chenery AM, Chiozza F, Clausnitzer V, Collar NJ, Collett LC, Collette BB, Fernandez CFC, Craig MT, Crosby MJ, Cumberlidge N, Cuttelod A, Derocher AE, Diesmos AC, Donaldson JS, Duckworth JW, Dutson G, Dutta SK, Emslie RH, Farjon A, Fowler S, Freyhof J, Garshelis DL, Gerlach J, Gower DJ, Grant TD, Hammerson GA, Harris RB, Heaney LR, Hedges SB, Hero J-M, Hughes B, Hussain SA, Icochea M J, Inger RF, Ishii N, Iskandar DT, Jenkins RKB, Kaneko Y, Kottelat M, Kovacs KM, Kuzmin SL, La Marca E, Lamoreux JF, Lau MWN, Lavilla EO, Leus K, Lewison RL, Lichtenstein G, Livingstone SR, Lukoschek V, Mallon DP, McGowan PJK, McIvor A, Moehlman PD, Molur S, Munoz Alonso A, Musick JA, Nowell K, Nussbaum RA, Olech W, Orlov NL, Papenfuss TJ, Parra-Olea G, Perrin WF, Polidoro BA, Pourkazemi M, Racey PA, Ragle JS, Ram M, Rathbun G, Reynolds RP, Rhodin AGJ, Richards SJ, Rodriguez LO, Ron SR, Rondinini C, Rylands AB, de Mitcheson YS, Sanciangco JC, Sanders KL, Santos-Barrera G, Schipper J,

Self-Sullivan C, Shi Y, Shoemaker A, Short FT, Sillero-Zubiri C, Silvano DL, Smith KG, Smith AT, Snoeks J, Stattersfield AJ, Symes AJ, Taber AB, Talukdar BK, Temple HJ, Timmins R, Tobias JA, Tsytsulina K, Tweddle D, Ubeda C, Valenti SV, van Dijk PP, Veiga LM, Veloso A, Wege DC, Wilkinson M, Williamson EA, Xie F, Young BE, Akcakaya HR, Bennun L, Blackburn TM, Boitani L, Dublin HT, da Fonseca GAB, Gascon C, Lacher TE, Jr., Mace GM, Mainka SA, McNeely JA, Mittermeier RA, Reid GM, Paul Rodriguez J, Rosenberg AA, Samways MJ, Smart J, Stein BA, Stuart SN (2010) The Impact of Conservation on the Status of the World's Vertebrates. Science 330: 1503-1509.

Hohenlohe PA, Bassham S, Etter PD, et al. (2010) Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. Plos Genetics 6.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. Plos Genetics 6.

IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. Downloaded on 21 January 2015.

Jehle R, Sztatecsny M, Wolf JBW, Whitlock A, Hoedl W, Burke T (2007) Genetic dissimilarity predicts paternity in the smooth newt (*Lissotriton vulgaris*). Biology Letters 3: 526-528.

Jennions MD, Hunt J, Graham R, Brooks R (2004) No evidence for inbreeding avoidance through postcopulatory mechanisms in the black field cricket, *Teleogryllus commodus*. Evolution 58: 2472–2477.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403-1405.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27: 3070-3071.

Jones AG (2005) GERUD 2.0: a computer program for the reconstruction of parental genotypes from half-sib progeny arrays with known or unknown parents. Molecular Ecology Notes 5: 708-711.

Jones AG, Adams EM, Arnold SJ (2002) Topping off: A mechanism of first-male sperm precedence in a vertebrate. Proceedings of the National Academy of Sciences of the United States of America 99: 2078-2081.

Jones OR, Wang J (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10: 551-555.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodoehl PA (2013) diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution 4: 782-788.

Kiesecker JM, Blaustein AR, Belden LK (2001) Complex causes of amphibian population declines. Nature 410: 681-684.

King EDA, Banks PB, Brooks RC (2013) Sexual conflict in mammals: consequences for mating systems and life history. Mammal Review 43: 47-58.

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220: 671-680.

Lacy RC (1988) A report on population-genetics in conservation. Conservation Biology 2: 245-247.

Lande R (1988) Genetics and demography in biological conservation. Science 241: 1455-1460.

Lecis R, Genetica C (2004) The endemic Sardinian newt *Euproctus platycephalus*: local threats and population decline. Italian Journal of Zoology 71: 195-198.

Liebgold EB, Cabe PR, Jaeger RG, Leberg PL (2006) Multiple paternity in a salamander with socially monogamous behaviour. Molecular Ecology 15: 4153-4160.

Liu Y, Nyunoya T, Leng S, Belinsky SA, Tesfaigzi Y, Bruse S (2013) Softwares and methods for estimating genetic ancestry in human populations. Human Genomics 7.

Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang S-P, Wang Z, Chinwalla AT, Minx P, Mitreva M, Cook L, Delehaunty KD, Fronick C, Schmidt H, Fulton LA, Fulton RS, Nelson JO, Magrini V, Pohl C, Graves TA, Markovic C, Cree A, Dinh HH, Hume J, Kovar CL, Fowler GR, Lunter G, Meader S, Heger A, Ponting CP, Marques-Bonet T, Alkan C, Chen L, Cheng Z, Kidd JM, Eichler EE, White S, Searle S, Vilella AJ, Chen Y, Flicek P, Ma J, Raney B, Suh B, Burhans R, Herrero J, Haussler D, Faria R, Fernando O, Darre F, Farre D, Gazave E, Oliva M, Navarro A, Roberto R, Capozzi O, Archidiacono N, Della Valle G, Purgato S, Rocchi M, Konkel MK, Walker JA, Ullmer B, Batzer MA, Smit AFA, Hubley R, Casola C, Schrider DR, Hahn MW, Quesada V, Puente XS, Ordonez GR, Lopez-Otin C, Vinar T, Brejova B, Ratan A, Harris RS, Miller W, Kosiol C, Lawson HA, Taliwal V,

Martins AL, Siepel A, RoyChoudhury A, Ma X, Degenhardt J, Bustamante CD, Gutenkunst RN, Mailund T, Dutheil JY, Hobolth A, Schierup MH, Ryder OA, Yoshinaga Y, de Jong PJ, Weinstock GM, Rogers J, Mardis ER, Gibbs RA, Wilson RK (2011) Comparative and demographic analysis of orang-utan genomes. Nature 469: 529-533.

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE (2013) Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. Plos Genetics 9.

McNeely JA, Miller KR, Reid WV, Mittermeier RA, Werner TB (1990) Conserving the world's biological diversity. IUCN, World Resources Institute, Conservation International, WWF-US, and the World Bank, Washington, DC.

Metzker ML (2010) Applications of Next-Generation Sequencing. Sequencing technologies - the next generation. Nature Reviews Genetics 11: 31-46.

Miller CR, Joyce P, Waits LP (2002) Assessing allelic dropout and genotype reliability using maximum likelihood. Genetics 160: 357-366.

Møller AP, Jennions MD (2001) How important are direct fitness benefits of sexual selection? Naturwissenschaften 88: 401–415.

Moritz C (1994) Defining Evolutionarily Significant Units for conservation. Trends in Ecology & Evolution 9: 373-375.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology, 22: 2841-2847.

Neff BD, Pitcher TE (2005) Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. Molecular Ecology 14: 19-38.

Nunney L, Campbell KA (1993) Assessing minimum viable population size - demography meets population-genetics. Trends in Ecology & Evolution 8: 234-239.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. Trends in Genetics 26: 177-187.

Palsboll PJ, Berube M, Allendorf FW (2007) Identification of management units using population genetic data. Trends in Ecology & Evolution 22: 11-16.

Parker GA (1998) Sperm competition and the evolution of ejaculates: towards a theory base. Sperm competition and sexual selection 3: 54.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28: 2537-2539.

Pimm SL, Raven P (2000) Biodiversity - Extinction by numbers. Nature 403: 843-845.

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. Plos One 7.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: Causes, consequences and solutions. Nature Reviews Genetics 6: 847-859.

Pounds JA, Crump ML (1994) Amphibian declines and climate disturbance - the case of the Golden toad and the Harlequin frog. Conservation Biology 8: 72-85.

Puddu F, Viarengo M, Ermineo C (1988) Euprotto sardo. In Animali di Sardegna: gli Anfibi e i Rettili. Ed. Della Torre, Cagliari.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. Conservation Biology 17: 230-237.

Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. American Journal of Human Genetics 73: 1402-1422.

Rousset F (2008) GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8: 103-106.

Ryder OA (1986) Species conservation and systematic – the dilemma of subspecies. Trends in Ecology & Evolution 1: 9-10.

Schuster SC (2008) Next-generation sequencing transforms today's biology. Nature Methods 5: 16-18.

Sever DM (1991) Comparative anatomy and phylogeny of the cloacae of salamanders (Amphibia, Caudata). 1. Evolution at the family level. Herpetologica 47: 165-193.

Sever DM (2002) Female sperm storage in amphibians. Journal of Experimental Zoology 292: 165-179.

Sever DM, Brizzi R (1998) Comparative biology of sperm storage in female salamanders. Journal of Experimental Zoology 282: 460-476.

Sever DM, Halliday T, Waights V, Brown J, Davies HA, Moriarty EC (1999) Sperm storage in females of the smooth Newt (*Triturus-v-vulgaris* L.): I. Ultrastructure of the spermathecae during the breeding season. Journal of Experimental Zoology 283: 51-70.

Sindaco R, Doria G, Razzetti E, Bernini F (2006) Atlante degli Anfibi e dei Rettili d'Italia/ Atlas of Italian Amphibians and Reptiles. Societas Herpetologica Italica. Edizioni Polistampa. Firenze.

Steinfartz S, Stemshorn K, Kuesters D, Tautz D (2006). Patterns of multiple paternity within and between annual reproduction cycles of the fire salamander (*Salamandra salamandra*) under natural conditions. Journal of Zoology 268: 1-8.

Stuart SN, Chanson JS, Cox NA, Young BE, Rodrigues ASL, Fischman DL, Waller RW (2004) Status and trends of amphibian declines and extinctions worldwide. Science 306: 1783-1786.

Stuart SN, Chanson JS, Cox NA, Young BE, Rodrigues ASL, Fischman DL, Waller RW (2004) Status and trends of amphibian declines and extinctions worldwide. Science 306: 1783-1786.

Szulkin M, Bierne N, David P (2010) Heterozygosity-Fitness Correlations: a time for reappraisal. Evolution 64: 1202-1217.

Tabachnick BG, Fidell LS (2001) Using Multivariate Statistics. 4thedn, Allyn and Bacon, University of Michigan.

Tennessen JA, Zamudio KR (2003) Early male reproductive advantage, multiple paternity and sperm storage in an amphibian aggregate breeder. Molecular Ecology 12: 1567-1576.

Trivers R (1972). Parental investment and sexual selection. In:Sexual Selection and the Descent of Man 1871–1971 (Ed. by B.Campbell), pp. 136–179. Chicago, Illinois: Aldine.

Valiere N (2002) GIMLET: a computer program for analysing genetic individual identification data. Molecular Ecology Notes 2**:** 377-379.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4: 535-538.

Vanrooy P, Stumpel AHP (1995) Ecological impact of economic development on Sardinian herpetofauna. Conservation Biology 9: 263-269.

Vignoli L, Silici R, Brizzi R, Bologna MA (2010) In vivo sexual discrimination in *Salamandrina perspicillata*: a cross-check analysis of annual changes in external cloacal morphology and spermic urine release. Herpetological Journal 20: 17-24.

Vignoli L, Silici R, Bissattini AM, Bologna MA (2012) Aspects of olfactory mediated orientation and communication in *Salamandrina perspicillata* (Amphibia Caudata): an experimental approach. Ethology, Ecology and Evolution 24(2): 165-173.

Vogler AP, Desalle R (1994) Diagnosing units of conservation management. Conservation Biology 8: 354-363.

Wake DB, Vredenburg VT (2008) Are we in the midst of the sixth mass extinction? A view from the world of amphibians. Proceedings of the National Academy of Sciences of the United States of America 105: 11466-11473.

Wang JL (2004) Sibship reconstruction from genetic data with typing errors. Genetics 166: 1963-1979.

Waples RS (1991) Pacific salmon, *Oncorhynchus* spp., and the definition of 'species' under the Endangered Species Act. US National Marine Fisheries Service Marine Fisheries Review 53: 11-22.

White TA, Perkins SE, Heckel G, Searle JB (2013) Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. Molecular Ecology 22: 2971-2985.

Wigby S, Chapman T (2004) Sperm competition. Current Biology 14: R100–R103.

Young BE, Lips KR, Reaser JK, Ibanez R, Salas AW, Cedeno JR, Coloma LA, Ron S, La Marca E, Meyer JR, Munoz A, Bolanos F, Chaves G, Romo D (2001) Population declines and priorities for amphibian conservation in Latin America. Conservation Biology 15: 1213-1223.

Zeh JA, Newcomer SD, Zeh DW (1998) Polyandrous females discriminate against previous mates. Proceedings of the National Academy of Sciences of the United States of America 95: 13732-13736.

Zeh JA, Zeh DW (1994) Last-male sperm precedence breaks down when females mate with 3 males. Proceedings of the Royal Society B-Biological Sciences 257: 287-292.

Zeh JA, Zeh DW (1996) The evolution of polyandry I: intragenomic conflict and genetic incompatibility. Proceedings of the Royal Society of London Series B 263: 1711–1717.

Zeh JA, Zeh DW (1997) The evolution of polyandry II: postcopulatory defences against genetic incompatibility. Proceedings of the Royal Society of London Series B 264: 69–75.

Zhang P, Papenfuss TJ, Wake MH, Qu L, Wake DB (2008) Phylogeny and biogeography of the family Salamandridae (Amphibia: Caudata) inferred from complete mitochondrial genomes. Molecular Phylogenetics and Evolution 49: 586-597.

Zwaan BJ (1999) The evolutionary genetics of ageing and longevity. Heredity 82: 589-597.

ACKNOWLEDGEMENTS

In the end, here we are at the most important part of my thesis…these three years represented one of the most intense period of my life…they have been hard, they forced me to change places to live and consequently my life, but taught me a lot about myself…and gave me the chance to made a lot of new experiences and to meet a number of wonderful persons.

I would like to thank all those who contributed, also for a minimum part or in an indirect way, to the fulfillment of this thesis because, as the sea, it is made up a lot of single drops.

The first two persons I have to intensely thank are Marco Bologna and Leonardo Vignoli: without them, this thesis wouldn't have been a reality. They followed me since my very first steps as a "scientist", trusted in me and allowed me to independently carry out my research, but being always available for any doubt or need. In particular, I want to thank Leo for all the time spent together, especially in the field collecting animals, for have been my guide during all these years and for having helped me in improving myself both from a professional and personal point of view.

Most of the first part of my work benefited from the contribution of Daniele Macale, who followed me during the maintaining in captivity of "my" salamanders, and from whom I learnt a lot, both from a technical and human point of view. I owe a lot also to Giuliana, Max and Diego (Bioparco staff), which helped me with availability and passion.

Thanks to Zhao, for the shared time during his first year in Italy, and for having accepted to come with me to release all the metamorphs at Simbruini mountains. Thanks also to Maria Bobbio for her help in field sampling and in rearing the larvae of *Salamandrina*; it has been really a hard but nice work!

I was very lucky because I had the chance to share the "tugurio" and the department life with others students, most of which became not only collegues but friends: Agnese, Michela, Emanuela, Riccardo, Silvia, Marta, Irene, Lara, Marco, Andrea, Alessandro… Thank you guys for all the suggestions, the support and the nice moments!

My second life began when I moved to Bologna, and I would thank Dr. Ettore Randi for this opportunity. He received me in his lab, provided to me the best facilities for carrying out my research, and he was a great leader during the data analysis, always able to find a solution or a different approach. I really appreciated his constant will in improving the lab skills and knowledge.

If I had to thanks Dr. Randi for his hospitality, I first of all have to thanks Edo…it was thanks to you that I landed at Ispra, and later on it was always thanks to you that I discovered the lindy hop. I really miss our working hours, our talks, our taralli breaks...It was a pleasure to share this route with you, share our Spanish experience, our doubts….our Ph.D. life!

Now is the turn of my mentor, Franci. You adopted me since the very first time, sharing with me all your experience….with a huge patience! You taught to me all the lab's and software secrets, you fought with me against the evil 10H and the "folletto delle PCR", you came with me to my first international congress….you have been unique! Thanks Franci, this thesis is mine as much as it is yours!

In Ispra I found not only a wonderful place for working, but above all a second family. Elena, Romolo, you have been like an older sister and brother for me, always ready to help, support and incite me…and also to laugh a lot! Thanks for all the wonderful shared moments, from the dinners in the guesthouse to our dancing night! I'm happy you are now part of my life!

Silva, I don't have too much to say…you know, and our hug speaks for itself…thanks for all!

Aritz, you inspired me since the very first time, and I don't know how to thank you for all your help. Thanks for your enthusiasm for my project, for all your suggestions, for having shared the life in the guesthouse, for having showed to me how a real scientist should be…and for your friendship.

Luca, thanks for your good cheer and you energy, that made me smile also in very bad days, and also for you affection…you are my favorite Pipparsugo!

Marco, thanks for a lot of things! To having tried since the beginning to make me at ease, to having taught to me a lot about NGS, to having come with me and the blonde to the vet, for our time in Siviglia, for our swing, our talks, our hunting under the stars…and much more!

Davide, I have to thank you for all you help with R and the other bioinformatic issues…I would have been lost without you! And thanks also for all the nice walks with Thor, that I missed a lot!

I would also thank Fede, Chiara and Alessandra for their kindness, for having been always ready to answer to my infinite questions, for sharing the passion for cute animals…and for Qiagen!

Thanks to all the other people in the lab, Giulia, Letizia, Carlotta, Nadia, Patrizia, Manuela, Federica, Martina; to all the people who lived with me in the guesthouse, and made that period unforgettable: Paola, Mara, Francesco, Ezio, Alessandra and also to all the other Isprini as Camilla, Riccardo, Mariangela, Andrea.

A special thanks to Jacopo and Simona, for the beautiful relationship we have created, to Antea and Chiara for the very nice moments we spent together, and to Elisa, who is now a great friend of mine.

My third life took place in Madrid, where I had the chance to spend four unforgettable months, in the lab of Dr. David Vieites. I would sincerely thanks David for his welcome and his competence. This period has been incredible thanks to all the guys in the Museum of Natural Science. They completely shared with my their Spanish lifestyle: coffè breaks, office, ideas, cañas in the Asador, the preuvas, pestufinde, el gimnasio…and with some of them also the summer course in Barcelona! I hope to not forget someone, anyway thanks a lot to Marcos, Silvia, Melinda, Aurora, Paco, Regan, Chio, Marga, Javi, Raquel, Rigo, Chechu, Sergio, Aime, Daniel, Carlos, Alex, Rogga, Ramon, Rafa, Josè, Paloma, Eva, Elisa, Carol, Raul, Jorge Braun, Guillermo, Isaac… And thanks also to my roommate, first of all to Cristina, for having help me to learn Spanish, for all the time and the parties, for the nochevieja with her family, and to Adriaan, Luis and Lucia, for their kindness and company.

Then I would like to thank the two external referees that reviewed my thesis, Dr. Ekblom and Dr. Pertoldi, for their positive and comforting opinion and useful suggestion.

Thanks to Emiliano, who arrived in the corner but gave me a fundamental contribution in the final review, to Daniele, for all his valuable suggestions, and to Monica, for her empathy and company.

Finally, thanks to all those people who are not part of my “University Universe”, because it is mainly thanks to them that I’m who I am. So thanks to my family, that always supported me in my route, and in particular to my mother, for all her efforts during these years…mamma, we I argue a lot, but that at the end you are my best friend!

Thanks to all my old and new friends, for their unconditional love, and for being always by my side, especially in the worst moments! Arianna, Valentina, Ava, Chiaretta, Chiara, Alessandro, Bisi, Fede, Anna, Nicoletta, Marzia, Marco, Serena, Gaia, Davide, Giulia, Daniele...it’s a pleasure to have you in my life! Thanks to all the girls of my old and new volleyball team, to the new friends from the BBV and beach volley of Rome….this sport makes me feel always at home, wherever I am!

Thanks also to those I lost, because for better or for worse they left a track in my life.

And in the very end, thanks to my little Angie, who is an inseparable life partner, who knows me better than anyone else and fills my days of happyness….and as usual thanks to my angels, that give me the strenght…

Thank you!