



Borsa di studio attivata ai sensi di quanto disposto dal D.M. n. 1061 del 10/08/2021

Titolo del progetto: Raccolta di dati genetici e creazione di una banda dati georeferenziata a supporto della stima della variabilità genetica di organismi animali per la progettazione di azioni di conservazione della biodiversità delle aree europee

La borsa sarà attivata sul seguente corso di dottorato accreditato per il XXXVII ciclo:

BIOLOGIA AMBIENTALE ED EVOLUZIONISTICA

Responsabile scientifico: Prof. Alessio De Biase

Area per la quale si presenta la richiesta: GREEN

Numero di mensilità da svolgere in azienda: 6

Numero di mensilità da svolgere all'estero: 6 presso CIBIO, Portogallo

Azienda: Agriconsulting (www.agriconsulting.it)

Progetto di ricerca:

La biodiversità, la variabilità tra gli organismi viventi, sta diminuendo a livello globale, e questo ha richiesto forti impegni di conservazione da parte della maggior parte dei paesi del mondo, compresa l'Italia. La Commissione Europea ha adottato una nuova Strategia per la Biodiversità per il 2030 come componente centrale del Green Deal Europeo, ed è in corso lo sviluppo degli obiettivi post-2020 della Convenzione sulla Diversità Biologica (CBD). L'UE e i suoi Stati membri saranno presto vincolati ad impegni ambientali a livello continentale e globale, compreso l'ampliamento della copertura della rete Natura 2000 delle aree protette dall'attuale 18% al 30%.

L'azione di conservazione richiede una buona conoscenza della distribuzione della biodiversità e dei siti importanti per la persistenza della biodiversità. Tuttavia, a causa della scarsità di dati, le valutazioni di conservazione di solito rappresentano la biodiversità attraverso pochi gruppi tassonomici noti di vertebrati e solo a livello di specie. Eppure, la congruenza dei modelli spaziali della biodiversità tra taxa e livelli è solo parziale e uno dei problemi che i biologi della conservazione devono affrontare è la disponibilità di dati primari di distribuzione e di diversità genetica intraspecifica degli organismi biologici.

La diversità genetica intraspecifica è alla base della capacità di popolazioni e specie di evolversi e adattarsi in risposta a pressioni selettive. Di solito è strutturato spazialmente e le aree di elevata diversità genetica sono spesso sovrapposte tra specie diverse (Hewitt 1996). I recenti progressi nell'analisi dei dati (anche in chiave genomica) e nella sintesi per le valutazioni di conservazione consentono alla comunità scientifica di caratterizzare la biodiversità e il suo cambiamento in modo più completo. Recentemente, sono state messe a punto metodologie per identificare aree di conservazione prioritarie che tengono conto della diversità filogenetica e intraspecifica, integrando concetti di filogenesi, filogeografia, statistica spaziale e prioritizzazione della conservazione spaziale (Carvalho et al 2017). Purtroppo, la scarsità di dati genetici correttamente georeferenziati non consente di stimare la diversità intraspecifica per un grande numero di organismi e, nel contesto zoologico, soprattutto per gli invertebrati in senso lato.

L'obiettivo del progetto è creare un database sulla distribuzione e la diversità genetica degli animali europei, contenente informazioni sulla diversità intraspecifica e sulla presenza di specie. Popoleremo il database con dati provenienti da database esistenti, dati bibliografici, collezioni non ancora digitalizzate su vertebrati e invertebrati. Esamineremo le conoscenze filogeografiche, e quelle derivanti da studi di genetica di popolazione e filogenesi molecolare, esistenti su taxa selezionati per valutare i modelli di variabilità al di sotto del livello di specie. I dati ottenuti da studi filogeografici condotti per l'area geografica e i taxa di nostro interesse, saranno analizzati sia in un quadro filogeografico sia come dati spazialmente espliciti secondo l'approccio classico alla genetica delle popolazioni. Nelle

nostre intenzioni il database dovrebbe rappresentare un punto di partenza per la raccolta di dati che possano essere integrati nelle analisi volte alla prioritizzazione delle azioni di conservazione secondo l'approccio integrato discusso in precedenza. A complemento di questo primo importante obiettivo, il secondo obiettivo è l'analisi dei dati relativi a taxa selezionati in funzione della densità di dati disponibili per modellizzare la diversità animale in Italia, o per una più ampia estensione geografica europea. Mapperemo modelli spaziali di linee intraspecifiche e diversità di specie e li utilizzeremo per identificare siti importanti per la persistenza della diversità animale utilizzando la metodologia Key Biodiversity Areas, uno standard sviluppato dall'Unione internazionale per la conservazione della natura e delle risorse naturali (IUCN 2016).

Costruire una banca dati di dati genetici per la pianificazione della conservazione

Metteremo a punto un database contenente informazioni sulla diversità intraspecifica e sulla presenza di specie di animali vertebrati e invertebrati presenti in Europa. I dati filogeografici e altri dati spazialmente espliciti della letteratura saranno utilizzati per popolare il database. Effettueremo una ricerca bibliografica e una valutazione critica della letteratura sulle popolazioni italiane ed europee dei taxa animali selezionati. Inoltre, verranno interrogati i database che memorizzano dati genetici (es. NCBI/EMBL-EBI/DDBJ, BOLD, GGBN Data Portal, GD²) o filogeografici e filogenetici (es. The Dryad Digital Repository; TreeBase; Map of Life). Solo una parte di questi dati è georeferenziata (Scotch et al 2011; Pope et al 2015, Gratton et al 2017) e si procederà alla creazione di un database di sequenze georeferenziate, principalmente di marker mitocondriali, ottenute per l'area di studio. Inoltre, sulla base di approcci sviluppati di recente (Gratton et al 2017), svilupperemo un software che ci permetta di cercare automaticamente sequenze georeferenziate nei repository e aggiornare il database ogni volta che sia necessario. Sono già stati raccolti dati genetici mitocondriali per tutti gli Anfibi e Rettili in Italia, a seguito di un'attenta revisione degli studi filogeografici disponibili che coinvolgono l'erpetofauna italiana (es. Mattocchia et al 2011; Bisconti et al 2011). Le analisi relative al secondo obiettivo del progetto, si concentreranno principalmente sui taxa per i quali i dati raccolti saranno sufficienti a modellizzare la distribuzione e la diversità intraspecifica. Questo ci consentirà di condurre analisi e confronti per mettere in luce le carenze dei dati che rappresentano sempre più un impedimento alla prioritizzazione delle azioni di conservazione e dunque contribuire a comunicare alla comunità scientifica le priorità d'indagine per raggiungere gli obiettivi di conservazione della biodiversità a livello europeo.

Confronto della diversità intraspecifica all'interno e all'esterno dei siti prioritari

Ai fini del secondo obiettivo del progetto, come applicazione pilota del database costruito nel passaggio precedente, procederemo a valutare lo stato della diversità genetica all'interno delle aree attualmente considerate prioritarie per le politiche di conservazione rispetto ad aree non prioritarie ma confrontabili che le precedenti. Considereremo il database genetico sopra descritto come il nostro universo di campionamento. Successivamente si procederà al campionamento di tutte le specie che ricadono nelle aree prioritarie e di controllo; tali specie dovrebbero però essere rappresentate all'interno del database genetico da un numero adeguato di esemplari (10-15; ma si veda anche l'approccio della rarefazione) per ciascuna delle popolazioni all'interno dell'universo di campionamento. Per ogni singola specie si calcoleranno analiticamente stimatori della diversità genetica (ad esempio, diversità nucleotidica, stimatore di Watterson di , diversità aplotipica) sia nelle aree prioritarie che in quelle di controllo. Quindi determineremo le differenze significative tra le stime della diversità genetica eseguendo test di permutazione (Alexander et al 2016) e per tenere conto di possibili dimensioni disuguali dei campioni genereremo curve di accumulo di aplotipi che, in modo simile a curve di rarefazione della ricchezza di specie (Gotelli & Colwell 2001), rappresenteranno il grado di comportamento asintotico come funzione sia del numero di campioni campionati che del numero medio cumulativo di aplotipi accumulati (Phillips et al 2019).

Per riassumere la diversità genetica multi-taxa in diverse aree geografiche, applicheremo un approccio geostatistico (Monestiez et al 1994; Guillot et al 2009; Tarroso et al 2015) per costruire una superficie spaziale della diversità genetica multispecie utilizzando il kriging convenzionale (es. Ribeiro & Diggle 2001) o un approccio incentrato sulla filogeografia (Tarroso et al 2019). Standardizzeremo la diversità nucleotidica () in ogni località per ogni taxon (che varia da 0 a 1) per portare queste metriche alla stessa scala tra taxa e tra diversi marcatori genetici con distinti tassi di evoluzione (Fonseca et al 2019). Successivamente, interpoleremo le stime della diversità dei nucleotidi tra le località (kriging) per ciascuna specie e sommeremo le mappe risultanti per tutte le specie insieme per ottenere la superficie

spaziale finale della diversità genetica (Fonseca et al 2019). Questo approccio aiuta a bypassare la scarsità di campioni utilizzati negli studi filogeografici o genetici spazialmente espliciti, o disponibili nei database genetici pubblici, e consentirà di generare mappe spaziali continue. Le superfici genetiche ottenute sono utili per valutare le aree prioritarie rilevate dal punto di vista evolutivo e genetico, per la prioritizzazione e pianificazione della conservazione spaziale (Vandergast et al 2008; Laity et al 2015; Carvalho et al 2017).

Bibliografia

- Alexander A et al 2016 *Molecular Ecology* 25:2754
Bisconti R et al 2011 *Molecular Ecology* 20:5313
Carvalho SB et al 2017 *Nature ecology & evolution* 1:0151
Gotelli NJ & Colwell RK 2001 *Ecology Letters* 4:379
Gratton P et al 2017 *Journal of Biogeography* 44:475
Guillot G et al 2009 *Molecular Ecology* 18:4734
Hewitt GM 1996 *Biological journal of the Linnean Society* 58:247
Laity T et al 2015 *Science of the Total Environment* 534:131
Mattoccia M et al 2011 *Biological Journal of the Linnean Society* 104:903
Monestiez P et al 1994 *Theoretical and Applied Genetics* 88:33
Phillips JD et al 2019 *Ecol Evol.* 9:2996
Pope LC et al 2015 *Molecular ecology* 24:3802
Scotch M et al 2011 *Journal of biomedical informatics* 44:S44
Tarroso P et al 2019 *Molecular Ecology Resources* 19:1081
Vandergast AG et al 2008 *Biological Conservation* 141:1648

Titolo del progetto (inglese): Gathering genetic data and building of a georeferenced database to support the evaluation of the genetic variation of animals to design biodiversity conservation actions in European areas.

Progetto di ricerca (inglese):

Biodiversity, the variability among living organisms, is decreasing globally, and this has required strong conservation commitments from most countries of the world, including Italy. The European Commission has adopted a new Biodiversity Strategy for 2030 as a central component of the European Green Deal, and the development of the post-2020 goals of the Convention on Biological Diversity (CBD) is underway. The EU and its Member States will soon be bound to environmental commitments at continental and global level, including extending the Natura 2000 network coverage of protected areas from the current 18% to 30%.

Conservation action requires a good knowledge of the distribution of biodiversity and of the sites important for the persistence of biodiversity. However, due to the paucity of data, conservation assessments usually represent biodiversity across a few known taxonomic groups of vertebrates and only at the species level. Yet the congruence of spatial models of biodiversity between taxa and levels is only partial and one of the problems that conservation biologists face is the availability of primary data on the distribution and intraspecific genetic diversity of biological organisms.

Intraspecific genetic diversity underlies the ability of populations and species to evolve and adapt in response to selective pressures. It is usually spatially structured, and the areas of high genetic diversity are often overlapped between different species (Hewitt 1996). Recent advances in data analysis (including genomic methodologies) and synthesis for conservation assessments allow the scientific community to characterize biodiversity and its change in a more comprehensive way. Recently, methodologies have been developed to identify priority conservation areas that take into account phylogenetic and intraspecific diversity, integrating concepts of phylogeny, phylogeography, spatial statistics and prioritization of spatial conservation (Carvalho et al. 2017). Unfortunately, the deficiency of genetic data, correctly georeferenced, does not allow to estimate the intraspecific diversity for a large number of organisms and, in the zoological context, especially for invertebrates.

The aim of the project is to create a database on the distribution and genetic diversity of European animals, containing information on intraspecific diversity and the presence of the species. We will populate the database with data from existing databases, bibliographic data, not yet digitized collections on vertebrates and invertebrates. We will examine the phylogeographic knowledge, and that deriving from population genetics and molecular phylogeny studies, existing on selected taxa to evaluate the models of variability below the species level. The data obtained from phylogeographic studies conducted for the geographical area and the taxa of our interest, will be analyzed both in a phylogeographic framework and as spatially explicit data according to the classical approach to population genetics. In our intentions, the database should represent a starting point for the collection of data that can be integrated in the analyzes aimed at prioritizing conservation actions, according to the integrated approach discussed above. To complement this first important objective, the second objective is the analysis of data for taxa selected according to the available density of data enough to model animal diversity in Italy, or for a wider European geographical extension. We will map spatial patterns of intraspecific lineages and species diversity and use them to identify important sites for the persistence of animal diversity using the Key Biodiversity Areas methodology, a standard developed by the International Union for Conservation of Nature and Natural Resources (IUCN, 2016).

Build a database of genetic data for conservation planning

We will develop a database containing information on intraspecific diversity and the presence of vertebrate and invertebrate animal species in Europe. Phylogeographic and other spatially explicit data from the literature will be used to populate the database. We will carry out a bibliographic search and a critical evaluation of the literature on the Italian and European populations of the selected animal taxa. Furthermore, databases that store genetic data (eg NCBI / EMBL-EBI / DDBJ, BOLD, GGBN Data Portal, GD²) or phylogeographical and phylogenetic data (eg The Dryad Digital Repository; TreeBase; Map of Life) will be queried. Only a part of these data is georeferenced (Scotch et al. 2011; Pope et al. 2015, Gratton et al. 2017) and a database of georeferenced sequences, mainly of mitochondrial markers, obtained for the study area will be created. Furthermore, based on recently developed approaches (Gratton et al. 2017), we will develop software that allows us to automatically search for georeferenced sequences in repositories and update the database whenever necessary. Mitochondrial genetic data have already been collected for all Amphibians and Reptiles in Italy, following a careful review of the available phylogeographic studies involving the Italian herpetofauna (eg Mattoccia et al, 2011; Bisconti et al 2011). The analyzes related to the second objective of the project will mainly focus on taxa for which the data collected will be sufficient to model the distribution and intraspecific diversity. This will allow us to conduct analyzes and comparisons to highlight the gaps in data that increasingly represent an impediment to the prioritization of conservation actions and therefore contribute to communicating the research priorities to the scientific community to achieve biodiversity conservation objectives at European level.

Comparison of intraspecific diversity inside and outside the priority sites

For the purposes of the second objective of the project, as a pilot application of the database built in the previous step, we will proceed to assess the state of genetic diversity within the areas currently considered priority for conservation policies with respect to non-priority areas but comparable to the previous ones. We will consider the genetic database described above as our "sampling universe". Subsequently, all the species falling within the priority and control areas will be sampled; however, such species should be represented in the genetic database by an adequate number of specimens (10-15; but see also the rarefaction approach) for each of the populations within the "sampling universe". For each single species, genetic diversity estimators will be analytically calculated (for example, nucleotide diversity, Watterson's estimator of θ , haplotypic diversity) both in the priority and control areas. We will then determine the significant differences between estimates of genetic diversity by performing permutation tests (Alexander et al. 2016) and to account for possible unequal sample sizes we will generate haplotype accumulation curves which, similarly to rarefaction curves of species richness (Gotelli & Colwell, 2001), will represent the degree of asymptotic behavior as a function of both the number of samples sampled and the cumulative average number of accumulated haplotypes (Phillips et al. 2019).

To summarize multi-taxa genetic diversity in different geographical areas, we will apply a geostatistical approach (Monestiez et al., 1994; Guillot et al., 2009; Tarroso et al. 2019) to construct a spatial surface of multispecies genetic

diversity using conventional kriging (eg Ribeiro and Diggle 2001) or an approach focused on phylogeography (Tarroso et al. 2019). We will standardize nucleotide diversity (π) at each location for each taxon (ranging from 0 to 1) to bring these metrics to the same scale between taxa and between different genetic markers with distinct rates of evolution (Fonseca et al. 2019). Next, we will interpolate the estimates of nucleotide diversity between locations (kriging) for each species and sum the resulting maps for all species together to obtain the final spatial surface area of genetic diversity (Fonseca et al. 2019). This approach helps to bypass the deficiency of samples used in spatially explicit phylogeographic or genetic studies, or available in public genetic databases, and will allow for the generation of continuous spatial maps. The genetic surfaces obtained are useful for evaluating the priority areas identified from an evolutionary and genetic point of view, for the prioritization and planning of spatial conservation (Vandergast et al. 2008; Laity et al. 2015; Carvalho et al. 2017).

Bibliography

- Alexander A et al. 2016 *Molecular Ecology* 25:2754–2772
- Bisconti, R et al 2011. *Molecular Ecology* 20: 5313-5327.
- Carvalho SB et al. 2017. *Nature ecology & evolution* 1: 0151.
- Fonseca, EM et al 2019. *Biodivers Conserv* 28:2877–2890
- Gotelli NJ & Colwell RK 2001. *Ecology Letters* 4: 379–391
- Gratton P. et al. 2017. *Journal of Biogeography* 44: 475-486.
- Guillot G et al. 2009. *Molecular Ecology* 18: 4734–4756.
- Hewitt G.M. 1996. *Biological journal of the Linnean Society* 58: 247-276.
- IUCN. (2016). *A Global Standard for the Identification of Key Biodiversity Areas, Version 1.0 (First edit)*. Gland, Switzerland: IUCN.
- Laity T. et al. 2015. *Science of the Total Environment* 534, 131-143.
- Mattocchia, Met al 2011. *Biological Journal of the Linnean Society* 104: 903-992.
- Monestiez P et al, 1994. *Theoretical and Applied Genetics* 88: 33-41.
- Phillips JD et al. 2019. *Ecol Evol.* 9: 2996-3010
- Pope L.C. et al. 2015. *Molecular ecology* 24: 3802-3809.
- Ribeiro PJ Jr & Diggle PJ 2001. *R News* 1:14–18
- Scotch M. et al. 2011. *Journal of biomedical informatics* 44: S44-S47.
- Tarroso P. et al. 2019. *Molecular Ecology Resources* 19: 1081–1094.
- Vandergast A.G. et al. 2008. *Biological Conservation* 141: 1648-1664.