



LIFE19 NAT/IT/000883



Committente



Con il contributo dello strumento finanziario Life dell'UE

PROGETTO LIFE19 NAT/IT/000883 LIFE INSUBRICUS

“Urgent actions for long-term conservation of *Pelobates fuscus insubricus* in the distribution area”



Piano reintroduzione/ripopolazione di *Pelobates fuscus insubricus* e Studio di Fattibilità Action A3.1 Analisi genetiche

Co-financed by:



Partners:



Deadline 10/2022 - Actual 22/03/2023



Indice

1	SUMMARY	3
2	Introduzione	5
2.1	Obiettivi dell'azione.....	5
3	Metodologia di lavoro	5
3.1	Area di studio.....	5
3.2	Campionamento biologico	6
3.3	Indagini molecolari	6
4	Risultati e discussione.....	8
4.1	Population genomics	8
4.2	mtDNA genetic structure.....	17
5	Bibliografia.....	19
6	REFERENTI.....	20

1 SUMMARY

The objective of action A3.1 was first of all to arrive at a genome-wide characterization of the natural population of the Insubric Pelobate fosco in northwestern Italy, which includes more than 95 % of the entire surviving population of this endemic species. Thus, the purpose of this analysis was to have in mind what was the actual ex-ante situation, to verify the genetic compatibility among the possible breeding pools in the putative source sites, and consequently to collect all the useful elements to modulate with knowledge the strategy of population strengthening and to the repopulation in the strict sense.

Based on the expressed purposes, 192 samples were thus successfully collected and processed (gDNA extraction) distributed as follows:

- IT2010011 "Paludi di Arsago" (ARS) (47 samples); IT2020007 "Pineta pedemontana di Appiano Gentile" (PIN) (11 samples), in Lombardy;
- IT1130004 "Bertignano Lake and pond near the road to Roppolo" (BER) (8 samples); IT1150001 "Ticino Valley" (CAM) (32 samples); IT1110021 "Ivrea Lakes" (LGH) (11 samples); IT1110047 "Scarmagno - Torre Canavese (Ivrea right moraine)" (MDX) (44 samples); IT1110035 "Poirino-Favari Ponds" (POI) (17 samples); and IT1120013 "Ritano Islet" (RIT) (22 samples) in Piedmont.

The different size of the amount of samples analyzed for each site depended on the relative size of the surviving Pelobate population, which in some cases was literally reduced to a dozen or so adult breeding individuals during the monitoring and sample collection period (2021-2022).

All the extracts obtained showed good yields in terms of gDNA concentration (> 20 ng/μL) and high purity, characteristics that certainly and positively affected the goodness of the analysis performed. Specifically, the sequencing analysis using the ddRAD approach resulted in a high number of SNPs (15,422) on which to perform the analyses.

It is emphasized that the numerosity of the marker panel exceeds by several orders of magnitude what has ever been produced so far in genetic studies on Italian populations of *P. fuscus* and is comparable, if not greater, to similar studies recently conducted in other European countries on the genus *Pelobates* (e.g., Dufresnes et al. 2019).

The bioinformatic elaborations carried out in summary revealed the following useful aspects for project purposes:

- The 8 macropopulations of *P. fuscus* constitute as many conservation units, being characterized by a unique genomic profile without any sharing of genotypes. This finding was partly predictable considering the strong process of fragmentation of natural habitats that has disrupted the Po Valley context in the last century, irreversibly isolating the remaining populations. In fact, the estimated gene flow instead shows a pattern of isolation among the macropopulations compatible with the geographic distances between them, at any rate, it excludes flow of individuals currently taking place among them.
- Despite the isolation of the populations, the rates of kinship and inbreeding were generally low. Populations such as those of POI, CAM, and ARS were found in this regard to have greater genetic diversity even from the mitochondrial point of view.

Ultimately, the salient results, when compared with the information available from previous bibliography, were obtained through a haplotype network reconstruction that showed 12 haplotypes from the 8 macropopulations under investigation, 4 of which were completely novel for Italy and the Palearctic.

2 Introduzione

2.1 Obiettivi dell'azione

Gli obiettivi dell'azione A3.1 sono inquadrabili in due principali ambiti relativi alla strutturazione genetica delle popolazioni residue di *P. fuscus* nell'area di studio (Piemonte – Lombardia):

- 1) caratterizzazione a scala genomica della popolazione naturale presente (situazione *ex-ante*) per la verifica della compatibilità genetica tra i possibili pool di riproduttori presenti nei putativi siti sorgente e la popolazione presente nella area di progetto da ripopolare (*population genomics*);
- 2) alcune informazioni storiche, relative alla variabilità genetica di popolazioni italiane di *P. fuscus* sono depositate in banche dati pubbliche (GenBank NCBI) sotto forma di sequenze nucleotidiche di marcatori molecolari standard. Si rende necessario un aggiornamento di tali conoscenze per evidenziare eventuali fenomeni di isolamento, frammentazione o deriva genetica delle stesse popolazioni, intercorsi negli ultimi due decenni (*mtDNA genetic structure*).

3 Metodologia di lavoro

3.1 Area di studio

Le indagini atte a valutare tramite approccio genomico lo stato di conservazione attuale di *P. fuscus* hanno interessato quasi tutte le stazioni riproduttive accertate nel corso dell'ultimo decennio circa sul territorio lombardo-piemontese per un totale di 8 macropopolazioni indagate (Fig. 1).

Nello specifico i siti Natura 2000 indagati sono stati i seguenti:

IT2010011 “Paludi di Arsago” (**ARS**) e IT2020007 “Pineta pedemontana di Appiano Gentile” (**PIN**) in Lombardia, IT1150001 “Valle del Ticino” (**CAM**), IT1120013 “Isolotto del Ritano” (**RIT**), IT1110035 “Stagni di Poirino-Favari” (**POI**), IT1130004 “Lago di Bertignano e stagno presso la strada per Roppolo” (**BER**), IT1110021 “Laghi Ivrea” (**LGH**), IT1110047 “Scarmagno - Torre Canavese (morena destra Ivrea)” (**MDX**) in Piemonte.

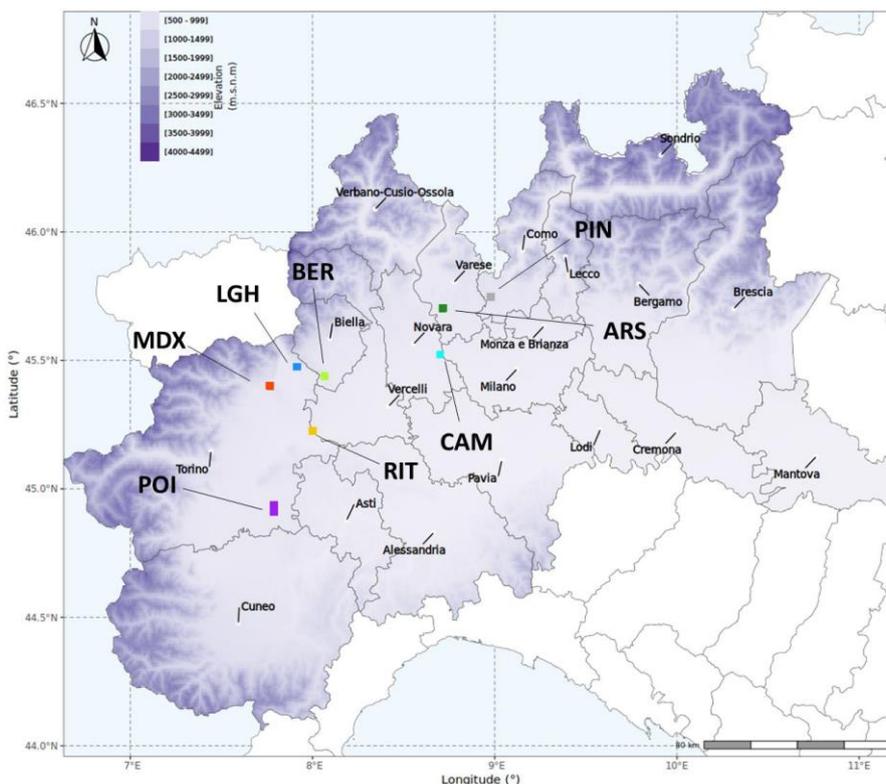


Figura 1: Carta di insieme del territorio interessato dalle attività di monitoraggio genomico *ex-ante*: si tratta di quasi tutte le stazioni riproduttive attualmente conosciute per il Pelobate fosco in subrico in Piemonte e Lombardia. I colori, rappresentativi di ogni località verranno poi riutilizzati nelle successive rappresentazioni grafiche relative alle analisi bioinformatiche condotte sui marcatori genomici e mitocondriali.

3.2 Campionamento biologico

Per poter procedere con la caratterizzazione dei principali indici di *population genomics* e *mtDNA genetic structure*, le 8 macropopolazioni di *P. fuscus* indagate sono state campionate utilizzando i seguenti criteri di massima:

- il numero di campioni biologici raccolti doveva essere non inferiore a 30 soggetti per macropopolazione o comunque proporzionale alla e rappresentativo della popolazione effettiva presente nel sito;
- i campioni raccolti dovevano essere rappresentativi delle diverse classi di età (da girino ad adulto “maturo”) e sesso;
- la raccolta dei campioni biologici doveva essere sincrona (i.e., condotta durante la stessa stagione riproduttiva) per tutte le macropopolazioni indagate;
- il campionamento doveva essere il meno invasivo possibile ma anche efficace per consentire una resa in termini di gDNA estratto di alta qualità sufficiente per poter procedere con le analisi dei marcatori genomici e mitocondriali. A tal fine, in seguito al rilascio di apposite autorizzazioni da parte del Ministero dell'Ambiente

e della Sicurezza Energetica (Registro Ufficiale MATTM 0063395 del 14-06-2021), si è proceduto alla raccolta di toe clip su individui adulti e di frammenti di coda di girini in stadio avanzato di sviluppo.

I campioni sono stati raccolti mediante strumenti appositamente sterilizzati tra un prelievo su un individuo e il successivo, e immediatamente riposti in tubi sterili tipo Eppendorf™ contenenti etanolo >96% (molecular biology grade), siglati in modo univoco e conservati in freezer (-20°C) fino alla consegna al laboratorio incaricato FEM2-Ambiente (Milano). In tabella Allegato 1 viene fornito l'elenco dei campioni utilizzati rispettivamente per le analisi di *population genomics* e di *mtDNA genetic structure*.

3.3 Indagini molecolari

Le analisi molecolari sui campioni raccolti sono state affidate alla società esterna FEM2-Ambiente s.r.l. a seguito di determinazione di aggiudicazione definitiva del Responsabile U.O.9 n. 160 del 29/04/2021 e successivo contratto di appalto del servizio di analisi genetiche, DNA-ambientale e analisi sanitarie del progetto LIFE19 NAT/IT/000883 LIFE INSUBRICUS – “URGENT ACTIONS FOR LONG-TERM CONSERVATION OF *PELOBATES FUSCUS INSUBRICUS* IN THE DISTRIBUTION AREA”- AZIONE A.3 SUB.ATTIVITÀ 3.1. ANALISI GENETICHE e 3.2 ANALISI SANITARIE. CIG 8652825912 CUP C16J19000420003.

Il dott. Valerio Mezzasalma ricopre il ruolo di responsabile delle attività di laboratorio, mentre il Prof. Andrea Galimberti è il responsabile scientifico dell'attività di ricerca incaricato da FEM2 Ambiente nell'ambito del succitato incarico.

2.3.1 Estrazione del gDNA e quantificazione

I campioni raccolti selezionati per le analisi molecolari Allegato 1 sono stati reidratati in soluzione fisiologica, sminuzzati e sottoposti al protocollo di estrazione specifico per tessuti animali con il kit commerciale DNeasy Blood & Tissue (Qiagen) seguendo le istruzioni fornite dal produttore. Gli estratti di gDNA sono stati poi quantificati mediante fluorimetro Qubit 4 (ThermoFisher) e in caso di parametri di concentrazione o di qualità dell'estratto inferiori alle soglie indicate dal Service di sequenziamento i campioni “problematici” sono stati nuovamente sottoposti al protocollo di estrazione apportando modifiche atte a migliorare la resa del processo.

2.3.2 Sequenziamento genomico

Le libraries ddRAD sono state allestite dal service IGA technology s.r.l. (Udine, Italy), utilizzando un protocollo customizzato sulla base di quello proposto da Peterson et al (2012). In breve, a partire dalle quantificazioni del gDNA ottenuto tramite il processo di estrazione purificazione, queste sono state normalizzate a valori di concentrazione uniformi e sottoposte a un doppio processo di restrizione tramite gli enzimi SphI e BstYI. I frammenti di gDNA così ottenuti sono stati purificati tramite il kit AMPureXP beads (Agencourt) e legate ad adattatori barcode che permettono, in fase di analisi una univoca associazione frammento-campione di partenza. In seguito ad un ulteriore processo di pooling, indexing, purificazione e check di quantificazione, le libraries sono state processate tramite sequenziamento con chimica V4 2 × 125 bp in un sequenziatore HiSeq2500 (Illumina, San Diego, CA) che permette di ottenere una copertura a 3M reads/campione.

2.3.3 Amplificazione e sequenziamento marcatori mitocondriali

Per quanto riguarda l'analisi della strutturazione genetica a livello mitocondriale (*mtDNA genetic structure*), si è proceduto innanzitutto a un'indagine bibliografica e di data mining nella banca genetica internazionale GenBank NCBI per individuare il marcatore genetico più indicato per effettuare confronti tra l'asset genetico attuale e dati precedentemente depositati per le stesse o per altre popolazioni italiane o Paleartiche di *P. fuscus*. L'indagine ha permesso di identificare chiaramente una regione di 571 bp del gene mitocondriale *cytb* come miglior candidato per effettuare tale analisi.

Cinque campioni per ciascuna delle 8 macropopolazioni campionate nell'ambito dell'azione progettuale in oggetto sono stati selezionati (in modo rappresentativo per l'intero periodo di raccolta e del sesso) per l'amplificazione e sequenziamento della regione *cytb*. In breve, a partire dagli stessi estratti di gDNA ottenuti come esplicitato nella sezione 2.3.1 è stata effettuata un'amplificazione mediante PCR seguendo lo stesso protocollo di Crottini et al. (2007) ma con reagenti, stechiometria di reazione e approccio di sequenziamento illustrati in Mazzamuto et al. (2016).

2.3.4 Analisi bioinformatiche

Population genomics

Le analisi bioinformatiche a partire dalle *reads* grezze di sequenziamento fino all'identificazione dei genotipi sono state condotte dal service IGA technology s.r.l. (Udine, Italy). In breve, le reads sono state sottoposte a un processo di *demultiplexing* tramite il software Stacks v 2.0 e allineate al genoma di riferimento di *Pelobates cultripes* (PRJEB48197). Contestualmente sono state condotte analisi di *filtering* sulle sequenze ottenute per assicurare un'elevata qualità dei genotipi sui quali inferire i principali indici di *population genomics* come illustrato in Magris et al. (2022). Tra questi, particolare attenzione è stata rivolta a:

- *Multidimensional Scaling Analysis*: basata sulla distanza genetica tra campioni interpretata sulla base dell'origine geografica degli stessi. Lo scopo è verificare se cluster genetici corrispondono a una comune provenienza in termini spaziali.
- *Admixture Analysis*: utilizzata per investigare la struttura delle macropopolazioni assunto un valore K del numero di macropopolazioni possibili (in questo caso da 2 a 8).
- *Kinship e inbreeding*: indici di inbreeding e parentela tra membri di una stessa macropopolazione.
- *Fst, Eterozigosità (attesa e osservata), Allelic richness, Ne (popolazione effettiva)*: indici standard nell'ambito della genetica di popolazione atti a stimare il flusso genico tra macropopolazioni e la strutturazione genetica per ogni macropopolazione indagata.
- *Maximum likelihood estimation (MLE)* per l'analisi di *Identity by Descent (IBD)*: analisi atta a identificare tra gli individui di ogni macropopolazione le relazioni di parentela di primo grado (genitori, figli, fratelli) e di

secondo grado ('fratellastri' – un genitore in comune -, avuncoli, 'nonni-nipoti'). Si rimanda a Galván-Femenía (2021) per ulteriori approfondimenti.

mtDNA genetic structure

Le sequenze di *cytb* ottenute tramite sequenziamento Sanger sono state corrette manualmente usando il software Bioedit e allineate usando il tool MAFFT con quelle reperite in banca genica a partire da precedenti studi condotti da Eggert et al. (2006), Crottini et al. (2007), Litvinchuk et al. (2013). L'allineamento ottenuto è stato quindi collassato in aplotipi tramite il tool FaBOX e questi sono stati assemblati sulla base della macropopolazione/gruppo geografico di appartenenza (Allegato 2) in ricostruzioni di Haplotype Network (Median-Joining Algorithm) mediante il software PoPart. In particolare sono state effettuate due ricostruzioni indicative dei seguenti parametri:

- Confronto tra aplotipi derivanti dalle 8 macropopolazioni indagate nell'ambito del presente progetto vs. aplotipi già identificati per le stesse popolazioni (e altre italiane) da Crottini et al. (2007) e Litvinchuk et al. (2013), vs. aplotipi provenienti dal resto della regione paleartica.
- Confronto tra i soli aplotipi italiani.

4 Risultati e discussione

4.1 Population genomics

Sulla base dei criteri esplicitati in precedenza sono stati processati con successo (estrazione gDNA) 192 campioni così distribuiti: ARS (47 campioni), PIN (11 campioni), BER (8 campioni), CAM (32 campioni), LGH (11 campioni), MDX (44 campioni), POI (17 campioni) e RIT (22 campioni) (Allegato 1). Tutti gli estratti ottenuti hanno mostrato buone rese in termini di concentrazione del gDNA (> 20 ng/uL) ed elevata purezza.

L'analisi di sequenziamento con approccio ddRAD, successivamente all'iter bioinformatico di filtering e validazione ha permesso di ottenere un elevato numero di SNPs (15'422) sui quali effettuare le analisi indicate nella sezione Metodi. Si sottolinea come la numerosità del pannello di marcatori supera di svariati ordini di grandezza quanto mai prodotto finora in studi genetici sulle popolazioni italiane di *P. fuscus* ed è comparabile, se non maggiore, a studi simili condotti di recente in altri paesi Europei sul genere *Pelobates* (es. Dufresnes et al. 2019).

Multidimensional Scaling Analysis

L'analisi (MDS) mette in relazione, in uno scenario multidimensionale, il raggruppamento dei genotipi degli individui campionati per ogni macropopolazione in termini di distanze genetiche, con la provenienza geografica degli stessi. Dai risultati ottenuti, sul pannello di 15'422 SNPs (Figura 2), si può osservare come le 8 macropopolazioni costituiscano altrettanti cluster genetici indipendenti, separati da distanze genetiche più (es. RIT vs. CAM) o meno (es. ARS vs. PIN) elevate. Un individuo della popolazione LGH presenta valori di distanza intermedi con la popolazione RIT, aspetto investigato più in dettaglio dagli output della *Admixture analysis*.

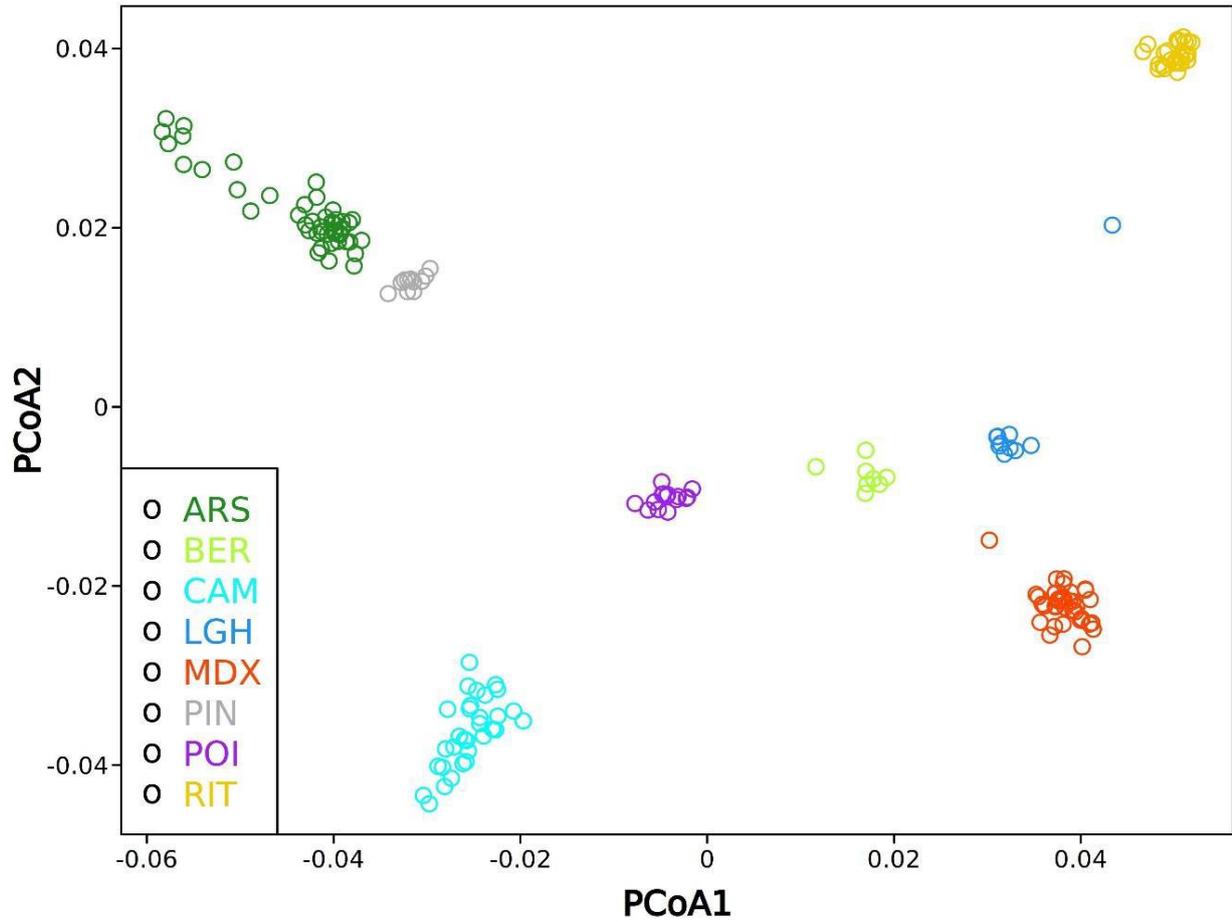


Figura 2: Multidimensional Scaling Analysis basata su IBS pairwise distances – 15'422 SNPs, 193 campioni (parametri di filtering degli SNPs: $DP \geq 10.85\%$, $MAF = 0.01$, $Id\ threshold = 0.2$). Ogni colore rappresenta una diversa macropopolazione di provenienza, come riportato in Figura 1.

Admixture Analysis

I risultati ottenuti consentono di esaminare più in dettaglio la putativa strutturazione genomica delle macropopolazioni indagate. Come si può osservare dal grafico in Figura 3 (pannello in alto), e dal relativo test di cross-validazione (pannello in basso), la configurazione più probabile e più supportata è quella che prevede un numero di macropopolazioni di *P. fuscus* nell'area di studio uguale a 8. Questo non solo corrisponde all'effettiva condizione rilevata su campo ma conferma ulteriormente l'indipendenza delle macropopolazioni indagate come singoli elementi di conservazione. È interessante notare, come già ipotizzato dall'analisi MDS, vi sia l'effettiva conferma della presenza nel sito LGH di un individuo con genotipo introgresso LGH-RIT. Questa peculiarità evidenzia l'elevato grado di sensibilità di un approccio genomico di questo tipo utile al fine di monitorare in futuro l'esito di interventi di ripopolamento o traslocazione di individui che hanno come obiettivo il rafforzamento delle macropopolazioni.

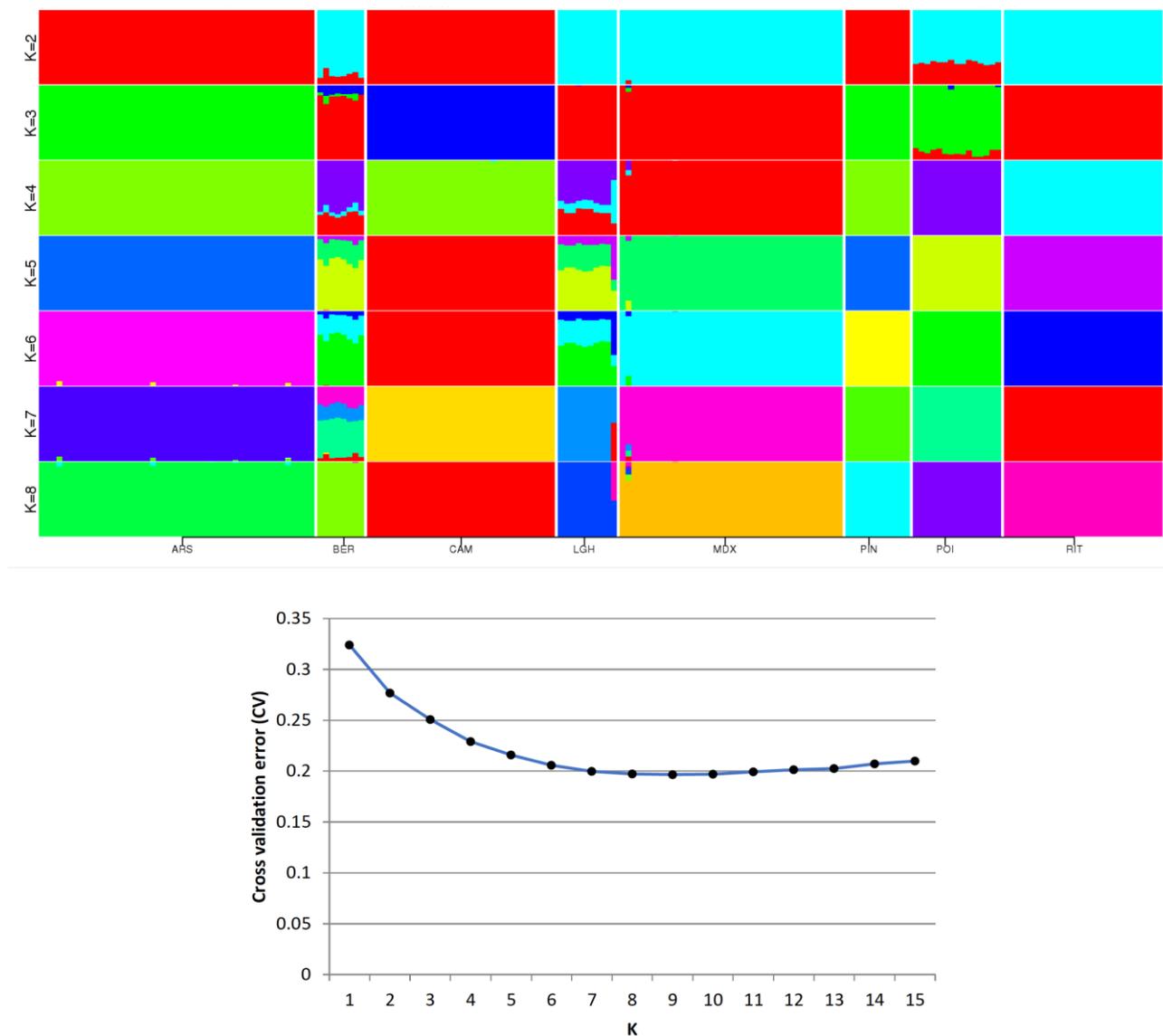


Figura 3: Admixture analysis con differenti valori di K (dove K è il putativo numero di macropopolazioni, da 2 a 8) (Pannello in alto). Risultati del test di cross-validazione: il valore più basso (K=8), rappresenta la configurazione più appropriata per l'interpretazione dei dati di Admixture (Pannello basso). Test condotto su 15'422 SNPs, 193 campioni (parametri di filtering degli SNPs: DP>=10.85%, MAF = 0.01, Id threshold = 0.2).

Kinship e inbreeding

I test di parentela e inbreeding sono mostrati in Figura 4. Come si può notare, i valori più elevati di questi indici sono osservabili nei genotipi di individui provenienti dalle macropopolazioni BER, LGH e RIT. Nel complesso, le macropopolazioni in esame presentano ancora un buon grado di diversità genetica anche se, le ridotte dimensioni di alcune e le barriere geografiche ad ostacolo di flussi continui potrebbero portare ad un progressivo innalzamento di tali indici. **NB:** Al fine di comprendere meglio gli sviluppi dello scenario di conservazione per le macropopolazioni oggetto di studio, si suggerisce di monitorare in modo regolare nel tempo i suddetti indici a partire dal termine del progetto. Per una visione più accurata dei pattern di parentela si rimanda ai risultati dell'analisi di *Identity by Descent (IBD)*.

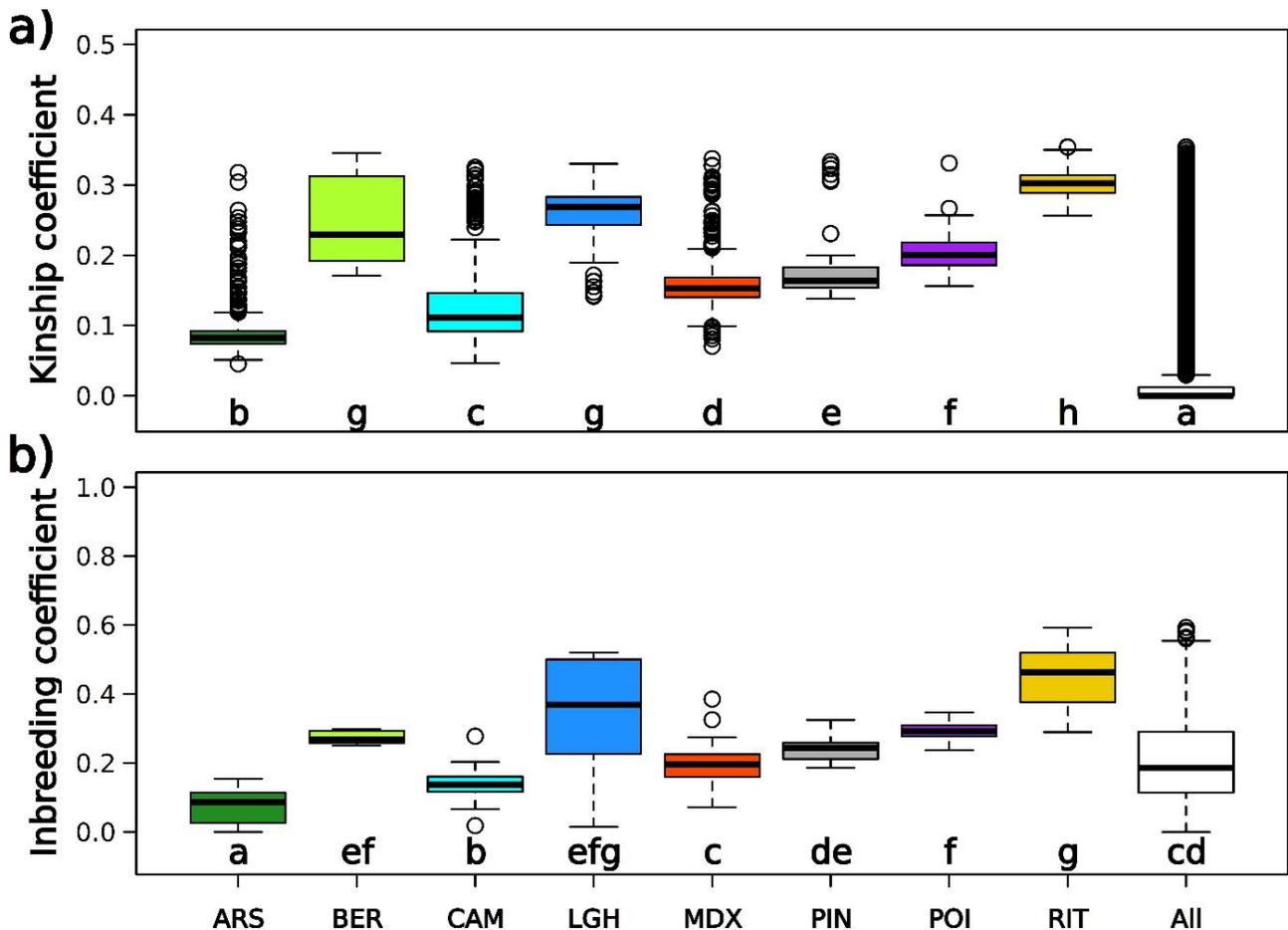


Figura 4: Coefficienti di Parentela (in alto) e inbreeding (in basso) per le 8 macropopolazioni di *P. fuscus* indagate e sulla popolazione complessiva (All). Test condotto su 15'422 SNPs, 193 campioni (parametri di filtering degli SNPs: $DP \geq 10.85\%$, $MAF = 0.01$, $Id\ threshold = 0.2$). La significatività statistica dei confronti dei valori di ciascun indice tra macropopolazioni è stata attribuita tramite il test di Wilcoxon (le macropopolazioni che condividono una stessa lettera non sono significativamente differenti).

Fst, Eterozigosità (attesa e osservata), Allelic richness, *Ne* (popolazione effettiva)

Il flusso genico tra macropopolazioni è riportato dalla matrice in Figura 5. Come si può osservare popolazioni più connesse (es. PIN e ARS) mostrano valori bassi dell'indice *Fst*, mentre popolazioni più isolate (es. BER e RIT) mostrano valori di *Fst* più elevati. Non sempre una ridotta distanza geografica corrisponde a un basso *Fst*, si veda ad es. il confronto LGH vs. BER. Pur trattandosi di una modellizzazione dei putativi flussi di individui tra macropopolazioni, si sottolinea come ulteriori aspetti di isolamento, deriva genetica o pressioni di altro tipo possono influenzare molto questo tipo di analisi.

	ARS	BER	CAM	MDX	POI	RIT	LGH	PIN
ARS		0.124	0.076	0.118	0.110	0.188	0.145	0.066
BER	0.038		0.144	0.141	0.192	0.278	0.205	0.190
CAM	0.022	0.040		0.132	0.130	0.222	0.166	0.118
MDX	0.033	0.034	0.035		0.159	0.197	0.146	0.166
POI	0.032	0.045	0.035	0.039		0.275	0.216	0.170
RIT	0.051	0.053	0.055	0.043	0.057		0.260	0.284
LGH	0.044	0.045	0.046	0.035	0.051	0.049		0.215
PIN	0.019	0.048	0.033	0.043	0.042	0.061	0.054	

Figura 5: Matrice dei valori di *Fst* (matrice a destra della diagonale) e di Jost's *D* (matrice a sinistra della diagonale). *Fst*= indice di fissazione allelica nei confronti tra macropopolazioni; Jost's *D*= differenziazione allelica.

La Tabella 1 riporta i parametri standard (*genomic population statistics*) per le varie macropopolazioni, tra cui Eterozigosità, Allelic Richness e Popolazione effettiva (*Ne*). Questi parametri riflettono e confermano quanto già evidenziato dagli indici precedenti. Ben evidenti sono i confronti tra macropopolazioni molto numerose e in buono stato di conservazione (es. ARS) e popolazioni più piccole e isolate (es. PIN). Si noti tuttavia che per quanto i panel allelici su cui è basato il calcolo di questi indici siano stati sottoposti a processi di normalizzazione, le disparità in termini di sampling size influiscono sui valori stessi.

Pop ID	Num_Indv	Obs_Het	Exp_Het	P_i	F_{is}	Allelic richness	N_e
ARS	47	0.13305	0.12337	0.1248	-0.02238	1.124064	97
BER	8	0.09084	0.0776	0.08327	-0.01455	1.080336	8.8
CAM	32	0.11369	0.10997	0.11189	-0.00616	1.110939	14
MDX	38	0.09743	0.09498	0.09632	-0.00052	1.095665	40.7
POI	15	0.08973	0.08663	0.08984	0.00144	1.088178	47
RIT	27	0.06337	0.05887	0.06005	-0.00668	1.059462	40.1
LGH	10	0.08489	0.0751	0.07932	-0.0115	1.077162	10.3
PIN	11	0.09978	0.09446	0.09919	-0.00151	1.09676	5.9

Tabella 1: Statistiche di popolazione basate sui dati genomici (15'422 SNPs, 193 campioni (parametri di filtering degli SNPs: $DP \geq 10.85\%$, $MAF = 0.01$, $Id\ threshold = 0.2$)

Maximum likelihood estimation (MLE) per l'analisi di Identity by Descent (IBD)

In Figura 6 (a-h), viene riportato in dettaglio, per ogni macropopolazione di *P. fuscus* indagata, la distribuzione dei confronti tra i genotipi degli individui in base al loro grado di parentela (simulazione più attendibile), come illustrato nello schema seguente:

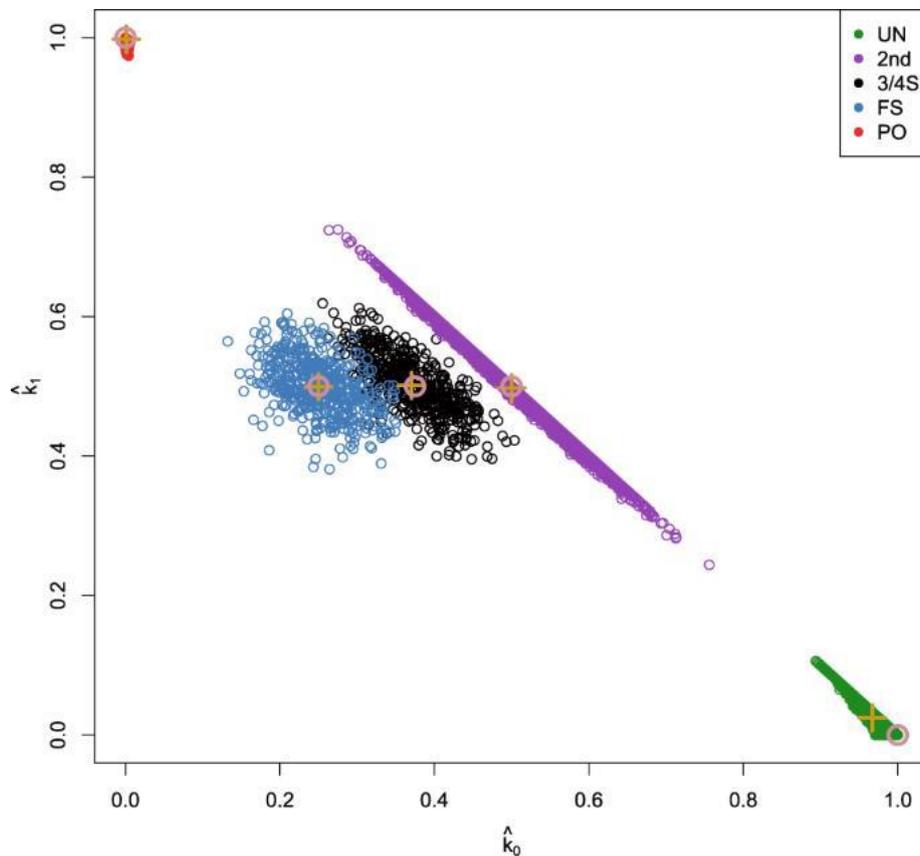


Figura 6: Distribuzione di IBD probability: schema di esempio da utilizzare per interpretare le successive ricostruzioni (a-h) relative alle 8 macropopolazioni di *P. fuscus* indagate. UN: unrelated; 2nd: second-degree relationships; 3/4S: three-quarter siblings. FS: full siblings; PO: parent-offspring. Immagine tratta da Galván-Femenía et al. (2021).

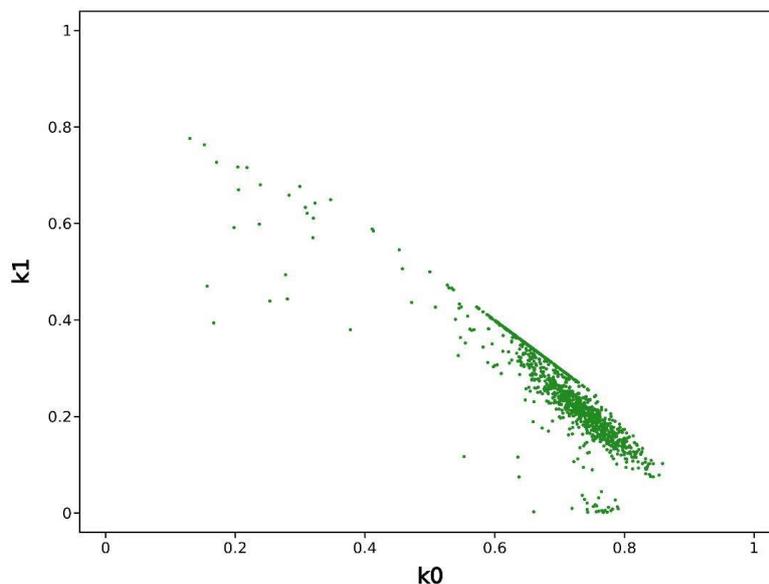


Figura 6a: ARS: la maggioranza degli individui presenta relazioni di 2nd degree / unrelated.

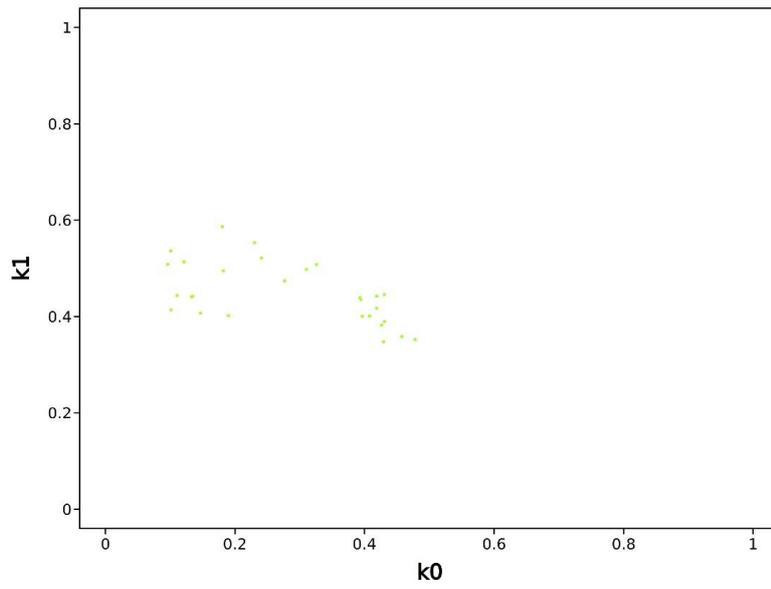


Figura 6b: BER: la maggioranza degli individui presenta relazioni di 2nd degree e full-sib.

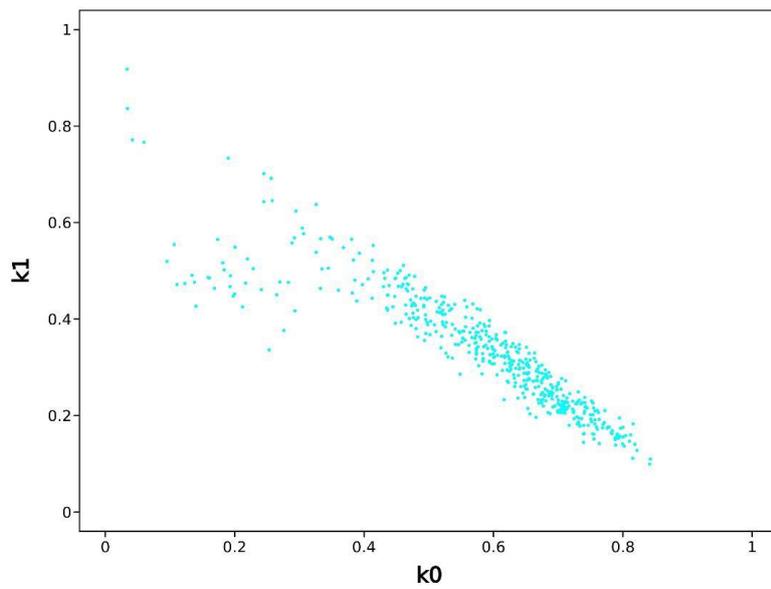


Figura 6c: CAM: la maggioranza degli individui presenta relazioni di 2nd degree e pochi sono full-sib

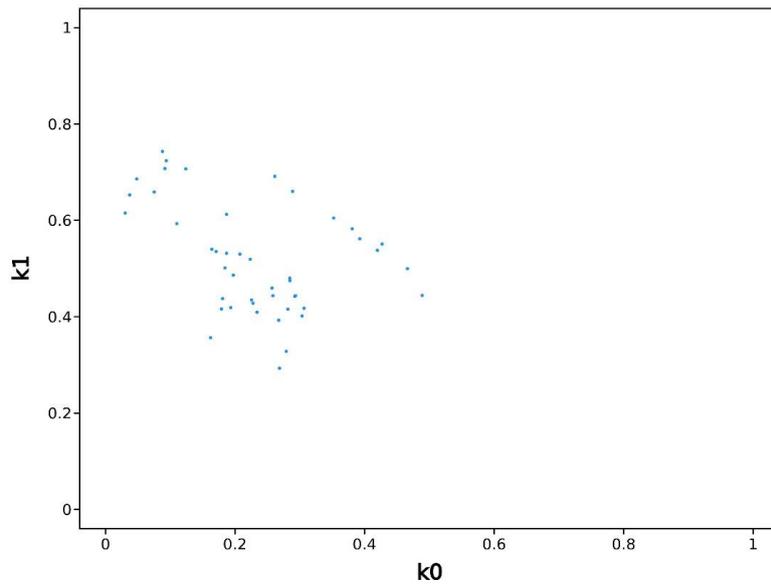


Figura 6d: LGH: la maggioranza degli individui presenta relazioni full-sib

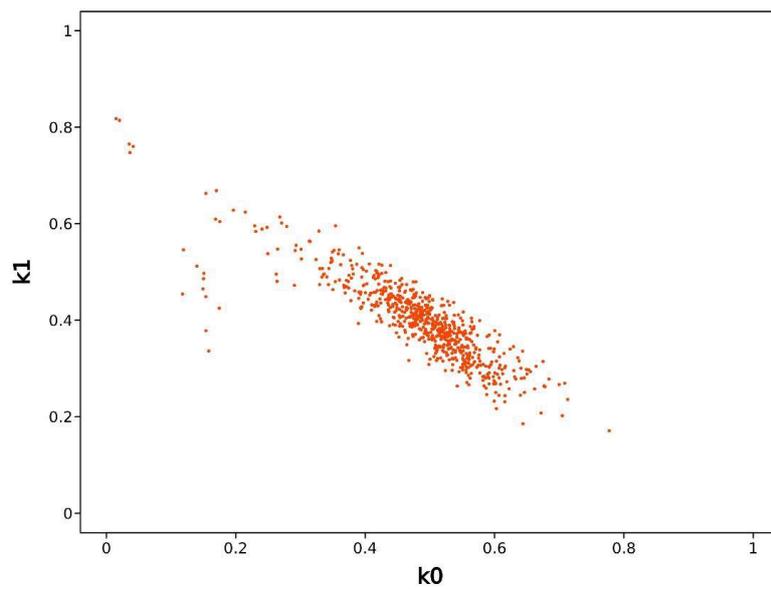


Figura 6e: MDX: la maggioranza degli individui presenta relazioni 2nd degree e alcuni confronti indicano relazioni full-sib. Cinque coppie di individui mostrano un elevato grado di inbreeding (es. figli di individui imparentati, probabilmente a loro volta fratelli).

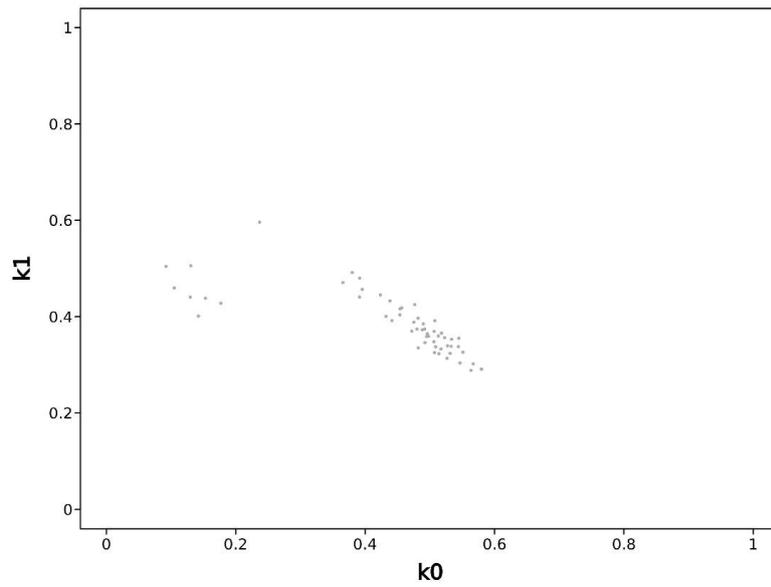


Figura 6f: PIN: la maggioranza degli individui mostra relazioni di 2nd degree e alcuni di full-sib.

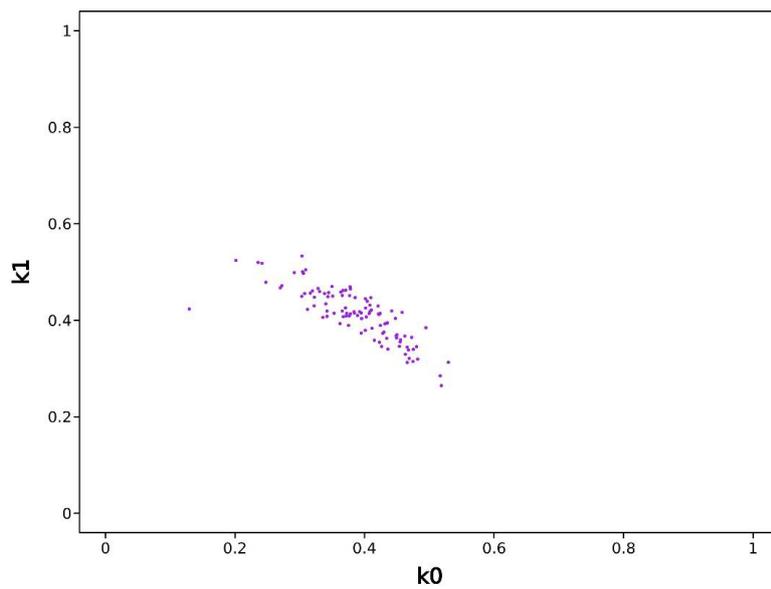


Figura 6g: POI: la maggioranza degli individui mostra relazioni di 2nd degree.

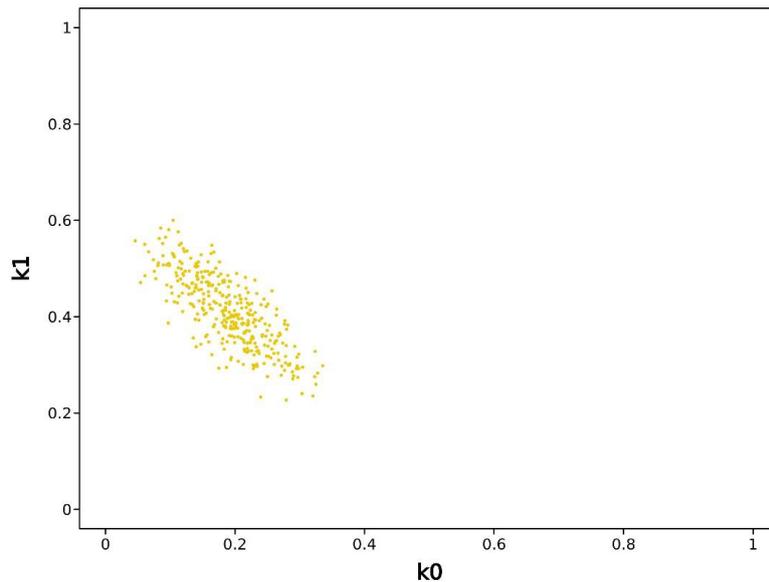


Figura 6h: RIT: tutti gli individui mostrano relazioni di full-sib.

4.2 mtDNA genetic structure

I 40 campioni selezionati dalle 8 macropopolazioni di *P. fuscus* indagate sono stati sottoposti con successo ad amplificazione e sequenziamento Sanger della regione *cytb* selezionata come esplicitato nella sezione metodi. Il confronto con i dati genetici precedentemente depositati nella banca genica internazionale (GenBank NCBI), ha permesso di ottenere un dataset complessivo di 198 sequenze totali, rappresentative di 17 paesi del Paleartico occidentale, delle quali 33 ottenute in precedenza da Crottini et al. (2007) e Litvinchuk et al. (2013) da macropopolazioni italiane, tra cui 4 (ARS, CAM, LGH, POI) in comune con i siti di monitoraggio e campionamento della presente attività progettuale e 3 non considerate in questa (Belangero e Pineta di Classe).

In totale sono stati individuati 43 aplotipi per l'intero dataset Paleartico Occidentale considerato. Di questi, 14 vedono rappresentanti nelle popolazioni italiane, tra cui 5 aplotipi (4 mai segnalati in precedenza!) provengono dal campionamento effettuato per il progetto LIFE in oggetto, uno dei quali in comune con quanto riscontrato in studi pregressi condotti sulle stesse popolazioni (es. Litvinchuk et al. 2013). Sette aplotipi descritti in tali studi, non sono più stati riscontrati nel campionamento attuale. Da chiarire se si tratti di sottocampionamento o di effettiva estinzione (Figura 7).

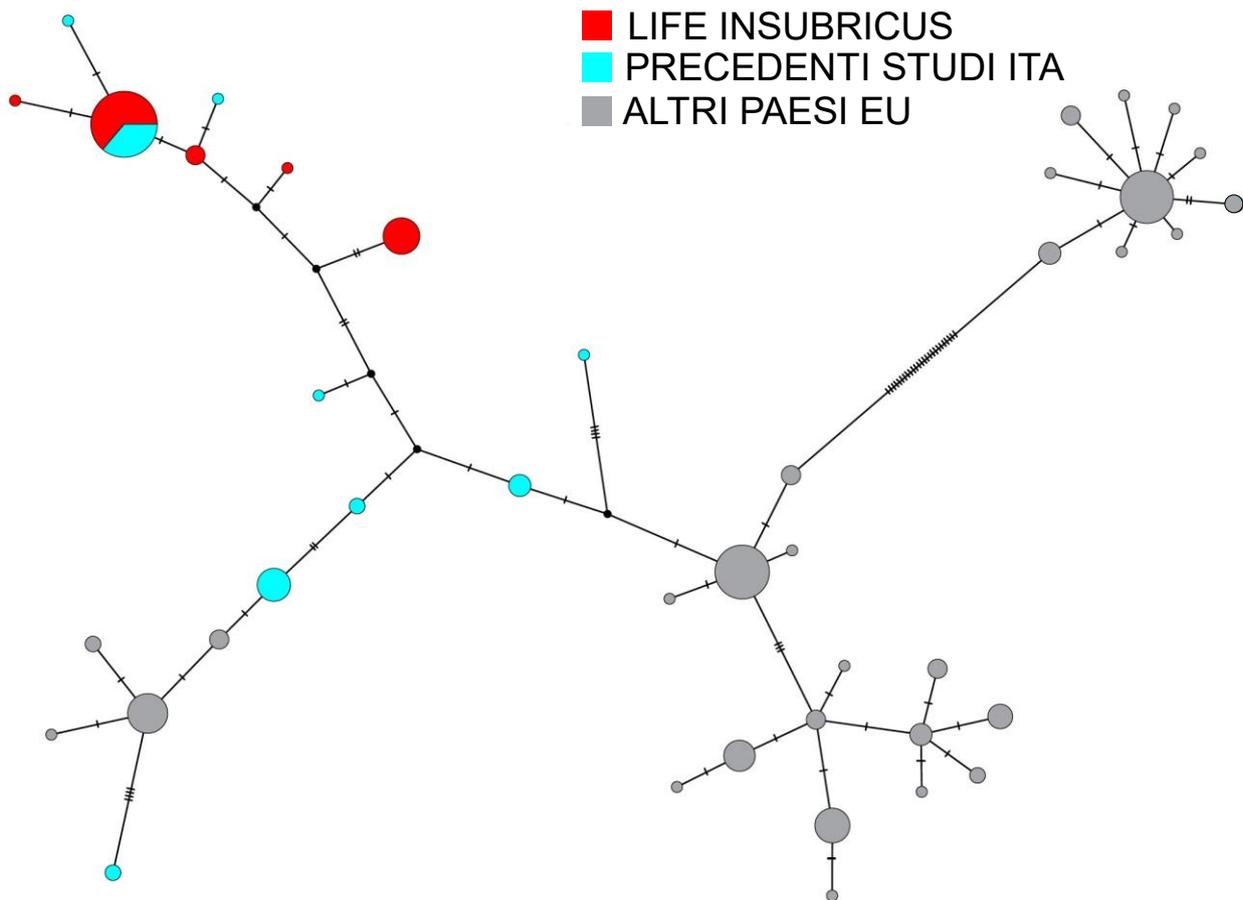


Figura 7: Median-joining network degli aplotipi cytb (43 aplotipi totali) del dataset complessivo che include oltre ai campioni raccolti nell'ambito della presente attività progettuale (rosso), altri campioni raccolti presso le stesse o altre località italiane (azzurro) nel contesto di studi precedenti, e del Palearctico Occidentale (grigio). Ogni cerchio rappresenta un aplotipo e la dimensione di ogni cerchio è proporzionale alla frequenza di quell'aplotipo. I punti neri corrispondono ai vettori mediani, mentre le barre rappresentano gli eventi di sostituzione.

Facendo un focus di maggiore dettaglio sulle sole popolazioni italiane si ottiene il network di aplotipi riportato in Figura 8. Come si può osservare, le macropopolazioni ARS, CAM e POI risultano le più diverse anche dal punto di vista del genoma mitocondriale dato che sono rappresentate da 6, 6, e 5 aplotipi rispettivamente. Le stesse macropopolazioni con l'aggiunta di BEL (non campionata in questo progetto), presentano anche degli aplotipi esclusivi (Figura 8).

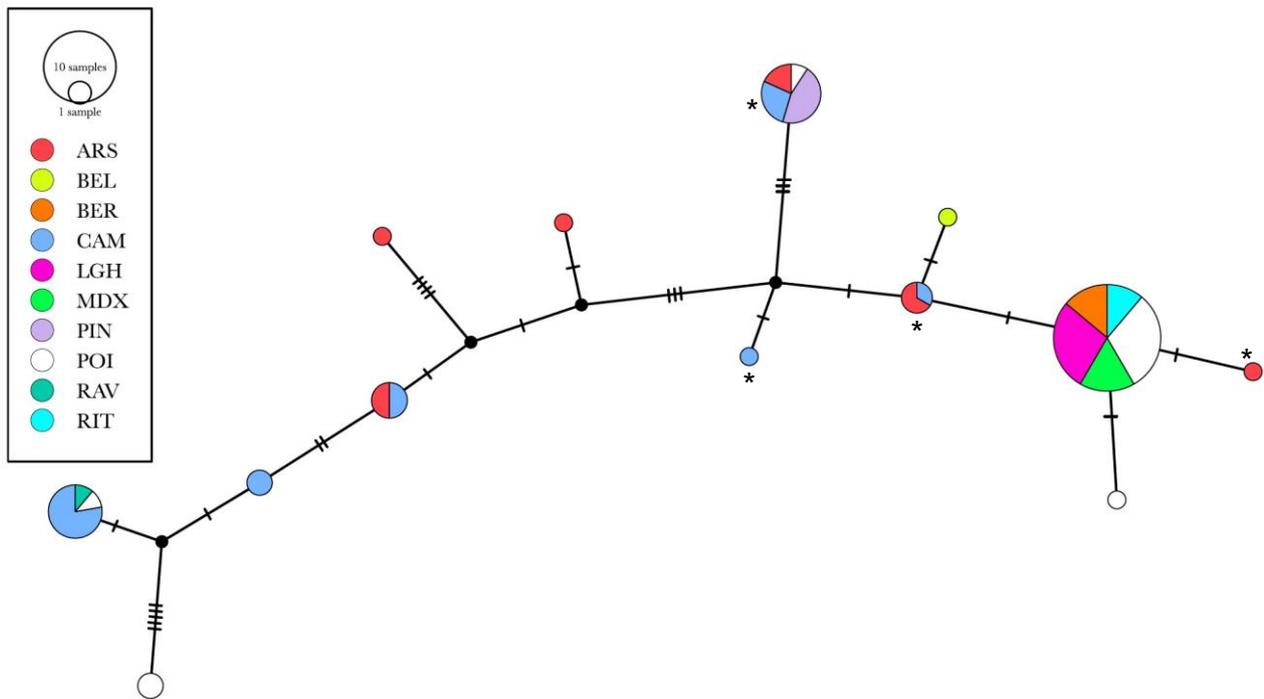


Figura 8: Median-joining network degli aplotipi italiani di *P. fuscus* per il marcatore mitocondriale *cytb* (13 aplotipi totali). L'analisi è stata condotta a partire dal dataset complessivo che include oltre ai campioni raccolti nell'ambito della presente attività progettuale anche quelli raccolti presso le stesse o altre località italiane in studi precedenti. Ogni cerchio rappresenta un aplotipo e la dimensione di ogni cerchio è proporzionale alla frequenza di quell'aplotipo. I punti neri corrispondono ai vettori mediani, mentre le barre rappresentano gli eventi di sostituzione. Il colore dei cerchi rappresenta la macropopolazione di provenienza. Il simbolo * indica gli aplotipi emersi dalle attività di monitoraggio condotte durante il presente progetto LIFE.

Allegato 1: Elenco campioni sottoposti alle analisi di *population genomics* e *mtDNA genetic structure*

Allegato 2: Tabella degli aplotipi e provenienza delle sequenze *cytb* utilizzate per le analisi di *mtDNA genetic structure*

5 Bibliografia

Crottini, A., Andreone, F., Kosuch, J., Borkin, L. J., Litvinchuk, S. N., Eggert, C., & Veith, M. (2007). Fossorial but widespread: the phylogeography of the common spadefoot toad (*Pelobates fuscus*), and the role of the Po Valley as a major source of genetic variability. *Molecular Ecology*, 16(13), 2734-2754.

Dufresnes, C., Strachinis, I., Suriadna, N., Mykytynets, G., Cogălniceanu, D., Székely, P., ... & Denoël, M. (2019). Phylogeography of a cryptic speciation continuum in Eurasian spadefoot toads (*Pelobates*). *Molecular Ecology*, 28(13), 3257-3270.

Eggert, C., Cogălniceanu, D., Veith, M., Dzukic, G., & Taberlet, P. (2006). The declining Spadefoot toad, *Pelobates fuscus* (Pelobatidae): paleo and recent environmental changes as a major influence on current population structure and status. *Conservation Genetics*, 7, 185-195.

Galván-Femenía, I., Barceló-Vidal, C., Sumoy, L., Moreno, V., de Cid, R., & Graffelman, J. (2021). A likelihood ratio approach for identifying three-quarter siblings in genetic databases. *Heredity*, 126(3), 537-547.

Litvinchuk, S. N., Crottini, A., Federici, S., De Pous, P., Donaire, D., Andreone, F., ... & Rosanov, J. M. (2013). Phylogeographic patterns of genetic diversity in the common spadefoot toad, *Pelobates fuscus* (Anura:

Pelobatidae), reveals evolutionary history, postglacial range expansion and secondary contact. *Organisms Diversity & Evolution*, 13, 433-451.

Magris, G., Marroni, F., D'Agaro, E., Vischi, M., Chiabà, C., Scaglione, D., ... & Morgante, M. (2022). ddRAD-seq reveals the genetic structure and detects signals of selection in Italian brown trout. *Genetics Selection Evolution*, 54(1), 8.

Mazzamuto, M. V., Galimberti, A., Cremonesi, G., Pisanu, B., CHAPUIS, J. L., Stuyck, J., ... & Martinoli, A. (2016). Preventing species invasion: A role for integrative taxonomy?. *Integrative zoology*, 11(3), 214-228.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, 7(5), e37135.

6 REFERENTI

Il presente report è stato redatto da:

Dott. Valerio Mezzasalma (Responsabile attività analitiche FEM2-Ambiente s.r.l.)



Prof. Andrea Galimberti (Responsabile scientifico delle analisi genomiche su *P. fuscus* – BTBS, Università degli Studi di Milano-Bicocca).



Hanno inoltre contribuito alle fasi analitiche di laboratorio del progetto il Dott. Fausto Ramazzotti (BTBS – UNIMIB) e i Dott.ssa Jessica Frigerio e Tommaso Gorini (FEM2-Ambiente srl).