

Caratterizzazione genetica delle popolazioni di *Zerynthia polyxena*

1. Introduzione

Prerequisito indispensabile per i progetti di conservazione che prevedano la traslocazione di individui da una popolazione ad un'altra è conoscere la diversità genetica delle diverse popolazioni coinvolte. E' infatti fondamentale evitare di traslocare individui fra popolazioni geneticamente molto differenti perché accoppiamento fra individui geneticamente molto distanti può portare a depressione da *outbreeding* (esoincrocio) fra i quali una scarsa *fitness*. Inoltre, nel caso di popolazioni piccole e minacciate è importante conoscere il grado di variabilità genetica all'interno e tra le popolazioni in modo da verificare lo stato di conservazione delle stesse. Popolazioni piccole e con scarsa variabilità genetica possono andare incontro a eventi stocastici e/o depressione da *inbreeding* (inincrocio) che le porterebbero velocemente all'estinzione.

La caratterizzazione genetica delle popolazioni di *Z. polyxena* è stata effettuata dal gruppo di Zoologia dell'Università del Piemonte Orientale. A seguito della valutazione della situazione delle popolazioni, della tipologia di campioni disponibili e dei dati disponibili in letteratura, si è scelto di procedere sia con l'analisi di marcatori mitocondriali e nucleari (rispettivamente *cytochrome c oxidase*, utilizzato per il DNA barcoding e il gene *wingless*) sia con un approccio di genotipizzazione su larga scala degli SNPs (mutazioni puntiformi diffuse nel genoma) con il metodo di Restriction-site Associated DNA sequencing (RADseq).

2. Materiali e metodi

2.1 Estrazione e quantificazione del DNA

Nel corso degli anni di progetto sono stati raccolti campioni di tessuto da 303 individui provenienti dalle diverse popolazioni presenti in bassa e media Val di Susa. I campioni biologici, uno o due zampe centrali dell'animale, sono stati asportati tramite una forbice entomologica, sterilizzata



UNIVERSITÀ DEL PIEMONTE ORIENTALE

ad ogni prelievo; questo metodo è il meno invasivo tra quelli utilizzati negli insetti e non compromette le funzioni vitali dell'individuo (Koscinski et al., 2011). Il prelievo è avvenuto durante le operazioni di cattura, marcatura e ricattura effettuate nell'anno 2020 nelle Aree 1, 2, 5, 6 e in sessioni di cattura nelle aree di Exilles, Eclause, Lonati, Molinere e nell'Area 4. I campioni, una volta raccolti, sono stati conservati in etanolo al 100% e posti in freezer a -20 °C. L'estrazione del DNA è stata effettuata tramite kit di purificazione a membrana silicea per campioni con piccole concentrazioni di DNA (NucleoSpin Tissue XS, Macherey Nagel). Il DNA di ogni campione è stato eluito in 65 µl di Buffer di eluizione. La concentrazione del DNA nei campioni estratti è stata determinata con uno spettrofotometro NanoQuant Plate™ TECAN e attraverso fluorometria con QuBit (Thermo Fisher Scientific). La concentrazione del DNA misurata variava tra 0.01 e 20 ng/µl. Un secondo campionamento è avvenuto nelle aree 1 e 2 nella primavera 2022, gli individui raccolti sono stati analizzati tramite approccio RADseq (paragrafo 2.3) per evidenziare se fossero avvenuti cambiamenti nella variabilità genetica di queste due popolazioni successivamente agli interventi di traslocazione e creazione del corridoio ecologico fra le due aree.

2.2 analisi del DNA barcoding e gene wingless

Una selezione di campioni biologici raccolti nelle diverse aree di studio è stata caratterizzata geneticamente attraverso il sequenziamento di un marcatore nucleare ed uno mitocondriale. Il primo marcatore è il marcatore universale *cytochrome c oxidase* (COI, 658 paia di basi) del DNA mitocondriale, marcatore molecolare utilizzato universalmente negli insetti e in altri animali come *barcode* per studi tassonomici, filogeografici e filogenetici. Il secondo marcatore è una porzione del gene nucleare *wingless* (*wg*, 349 paia di basi). Entrambi i marcatori vengono comunemente utilizzati negli insetti e questo permette di avere a disposizione dati di confronto disponibili in database online (GenBank e BOLD System). Per l'amplificazione dei marcatori sono stati utilizzati i protocolli e i primer da Hebert et al. (2004, primers: LepF1 and LepR1) per citocromo c ossidasi I e da Zinetti et al. (2013; primers: wg_F29 e wg_R382) per il gene *wingless*. Il sequenziamento è stato effettuato dalla ditta esterna Macrogen Europe (The Netherland).

Gli elettroferogrammi risultanti sono stati verificati con il software 4Peaks (<https://nucleobytes.com/>) e allineati con Mega X (Kumar et al., 2018), gli aplotipi (sequenze uniche) sono stati identificati con il tool online FaBox (<https://birc.au.dk/~palle/php/fabox/>). Il network Median Joining degli aplotipi è stato ottenuto con il software PopArt (<http://popart.otago.ac.nz>).



UNIVERSITÀ DEL PIEMONTE ORIENTALE

L'albero Neighbour Joining con distanza di Tamura Nei è stato calcolato con Mega X e radicato con una sequenza di *Zerynthia rumina* pubblicata sulla banca dati GenBank (NCBI www.ncbi.nlm.nih.gov).

2.2 Analisi degli SNPs

L'approccio Restriction-site associated DNA sequencing (RADseq) permette di ottenere un elevato numero di marcatori a singolo nucleotide (SNPs) per ogni campione che consentono di caratterizzare con grande dettaglio i genotipi degli individui e la struttura genomica delle popolazioni. Per *Zerynthia polyxena* in Val di Susa è stato messo a punto un approccio RAD sequencing predisposto per utilizzare basse quantità di DNA di partenza (200 ng), adattandolo anche a campioni con (100-150 ng). Tra le tecniche disponibili il protocollo che è stato valutato più idoneo è stato quello del single-digested restriction site-associated DNA sequencing (sdRADseq) utilizzando un protocollo modificato da Elshire et al. (2011) e pubblicato in Pellegrino et al. (2015). Per la preparazione delle librerie di sequenziamento (*sequencing library*) dai 40 ai 60 campioni sono stati digeriti con l'enzima di restrizione PstI-HF® (New England Biosystem) e marcati con una combinazione di *barcode e adapters*. Le librerie così ottenute sono state sottoposte a corse separate su Illumina MiSeq (600 cicli). Le *reads* risultanti sono state demultiplexate e allineate con procedimenti bioinformatici, grazie alla collaborazione con lo spin off SmartSeq, e filtrati successivamente secondo i protocolli di Pellegrino et al. (2016) utilizzando il software PLINK (Purcell et al., 2007). La Discriminant Analyses of Principal Component (DAPC), gli indici di fissazione (*Fst*, *Fis*), gli indici di eterozigosità sono stati calcolati utilizzando hierfstat e adegenet in R (<https://www.r-project.org>). La presenza di flusso genico è stata verificata con il software DivMigrate (Keenan et al., 2013).

3. Risultati

3.1 Sequenziamento COI e wg

In totale sono stati sequenziati 44 campioni identificando 4 aplotipi (sequenze differenti) per COI e 52 campioni identificando 5 aplotipi per il gene *wg*. I Network Median Joining (fig. 1 e 2) calcolati sugli aplotipi mostrano come gli aplotipi più abbondanti siano COI_4 e wg2, queste sequenze sono condivise da individui in tutte le popolazioni campionate. Le popolazioni dell'Area 4 e 5, inoltre, mostrano due aplotipi privati di COI (COI_1 e COI_2). Nel complesso l'analisi di questi

due marcatori, uno nucleare ed uno mitocondriale, evidenzia la mancanza di una strutturazione fra le popolazioni in cui sono presenti molti aplotipi condivisi.

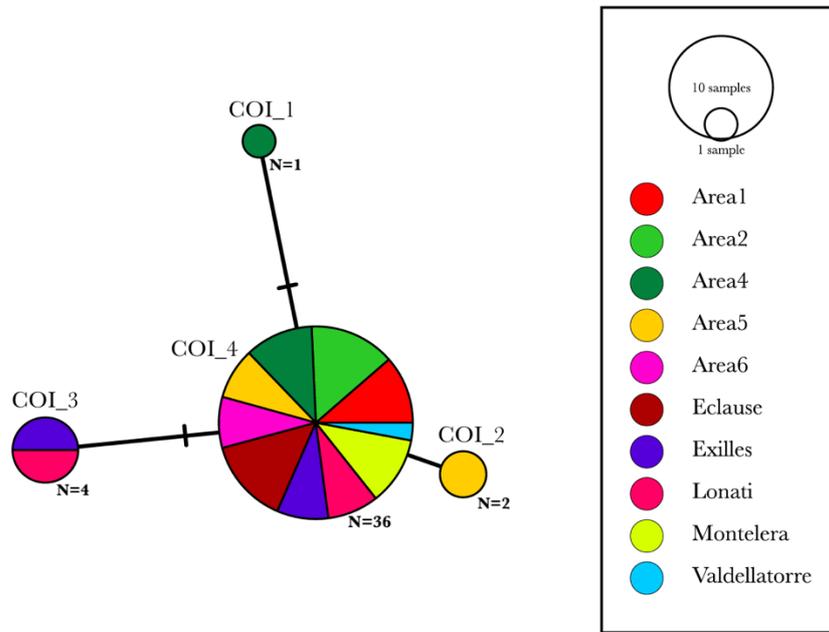


Figura 1 – Median Joining Network degli aplotipi calcolato su COI del mtDNA.

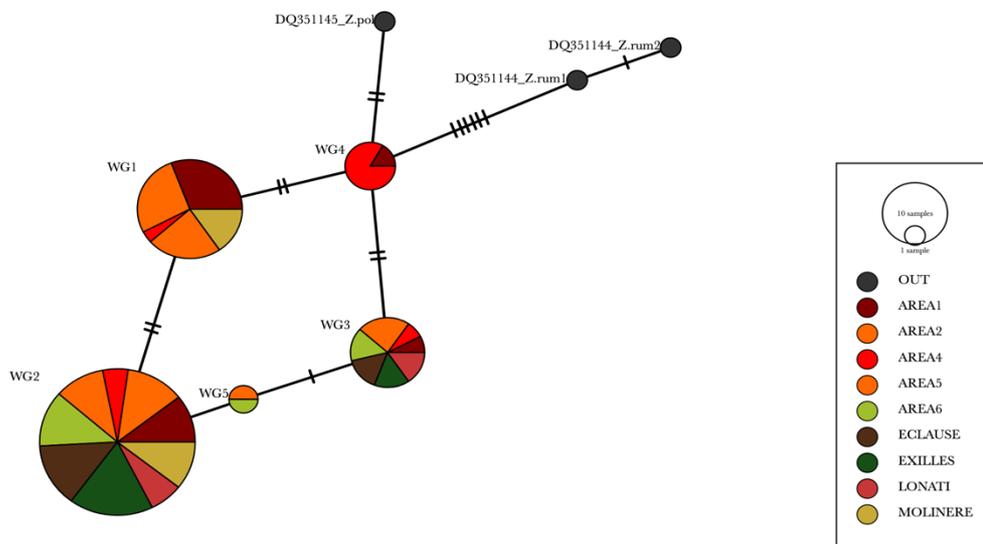


Figura 2 – Median Joining Network degli aplotipi calcolato su una porzione di 349 bp del gene wg.

Le distanze genetiche (*p-distances*) fra le popolazioni campionate calcolate su due marcatori mostrano valori bassi: esse sono minori dell'1% (da 0,2 % a 0,6 %) per il gene wg e compresi tra 0 e



UNIVERSITÀ DEL PIEMONTE ORIENTALE

0,1% per il marcatore COI. In particolare, la distanza genetica fra le Aree 1 e 2 è uguale 0 nel caso di COI, e 0,3% ($\pm 0,2$) misurata sulle sequenze di *wg*.

Le distanze genetiche calcolate all'interno delle popolazioni sono minori dell'1% e, per *wg*, variano da 0,2% nella popolazione di Exilles a 0,6% nell'Area 4. L'Area 1 e l'Area 2 mostrano distanze genetiche rispettivamente di 0,3 e 0,4 %. Nel caso di COI le distanze genetiche variano fra 0 e 0,1%.

3.2 Analisi degli SNPs anno 2020

In totale, per questa analisi di screening, sono stati analizzati 140 campioni, con un numero di *reads* (sequenze) totale di 55,7 milioni, variabile nei diversi campioni con una media di 680000 *reads* ciascuno.

L'allineamento delle *reads* e i primi filtri applicati hanno permesso di identificare 7202 SNPs che sono poi stati ulteriormente filtrati con il software PLINK, fino ad ottenere 136 campioni (Tabella 1) con un valore di genotipizzazione totale nel dataset residuo di 0.945.

Tabella 1: Elenco dei campioni genotipizzati per località e legenda delle aree.

Località	Area	Abbreviazione	Numero di individui genotipizzati	Numero di individui dopo filtri in PLINK
Chiomonte	Area 1	CH	11	11
Eclause	Area 5	EC	13	12
Exilles		EX	9	9
Forte Fenil	Area 5	FF	14	13
Forte Fenil-Pramand		FP	8	8
Giaglione	Area 2	GL	13	13
Ramats –	Area 4	LR	8	8
Le Combe	Area 6	LC	10	9
Maison-Baccon		MB	10	10
Milanere		MLE	10	9
Sant'Antonio	Area 4	SA	9	9
Val della Torre		VLT	8	8
Valgioie		VG	9	9
Val di Lanzo		VL	8	8
<i>Totale</i>			140	136

Gli indici di diversità calcolati sull'intero dataset sono:

- eterozigosità osservata (H_o) = 0.34
- eterozigosità attesa media delle popolazioni/diversità genetica media all'interno delle popolazioni (H_s) = 0.24
- diversità genetica totale media (H_t) = 0.25
- diversità genetica totale corretta (h_{tp}) = 0.25
- diversità genetica fra i campioni (D_{stp}) = 0.01
- indice di fissazione (F_{st}) = 0.03
- F_{st} corretto (F_{stp}) = 0.03
- coefficiente di inbreeding per tutti i loci (F_{is}) = -0.39
- Misura di differenziazione a livello di popolazione (D_{est}) = 0.01

I valori di eterozigosità osservata sono maggiori di quelli di diversità genetica media sia nell'intero dataset che nelle singole popolazioni; questa differenza può indicare presenza di *outbreeding* (incrocio di popolazioni precedentemente differenziate).

L'indice *F_{is}* è stato introdotto da Wright (1922) per quantificare la deviazione locale dall'equilibrio di Hardy Weinberg (HWE), misura la differenza fra eterozigosità attesa e osservata. In questo studio valori di *F_{is}* sono minori di 0, possibile segnale di un eccesso di eterozigoti. Un eccesso di eterozigoti può essere causato da accoppiamento disassortativo, *outbreeding* o da una selezione dipendente dalla frequenza che avviene in popolazioni dove gli eterozigoti vengono selezionati positivamente.

Tabella 2: indici di diversità genetica calcolati nelle diverse popolazioni. Hobs: Observed heterozygosity, H n.b.: unbiased heterozygosity, Hexp: Expected heterozygosity

Località	H expected Genetic diversity	Hn.b. unbiased heterozygosity	Hobs, Observed heterozygosity	P(0,95)	P(0,99)	Numero medio di alleli per locus
CH	0,2284	0,2392	0,3298	0,6868	0,8863	18.863
EC	0,254	0,2651	0,3557	0,7573	0,9017	19.017
EX	0,2438	0,2581	0,3446	0,8644	0,8644	18.644
FF	0,1462	0,1521	0,1968	0,4707	0,7253	17.253
FP	0,294	0,3136	0,4341	0,9159	0,9159	19.159

GL	0,2312	0,2405	0,3344	0,7247	0,8863	18.863
LC	0,2847	0,3015	0,4017	0,9337	0,9337	19.337
LR	0,1662	0,1773	0,243	0,6217	0,6217	16.217
MB	0,2669	0,281	0,3821	0,9195	0,9195	19.195
ML	0,2069	0,2192	0,2879	0,7644	0,7644	17.644
SA	0,2375	0,2515	0,335	0,8733	0,8733	18.733
VG	0,2699	0,2858	0,4023	0,8579	0,8579	18.579
VLT	0,2587	0,2759	0,3803	0,8283	0,8283	18.283
VL	0,2177	0,2323	0,3203	0,7152	0,7152	17.152

L'indice *Fst* misura la variabilità dovuta alla strutturazione genetica fra le popolazioni. Nel caso in studio le popolazioni hanno generalmente valori di *Fst* inferiori a 0.15, indicando così una bassa differenziazione, valori di *Fst* 0.15–0.2 indicano una differenziazione moderata (Tabella 3).

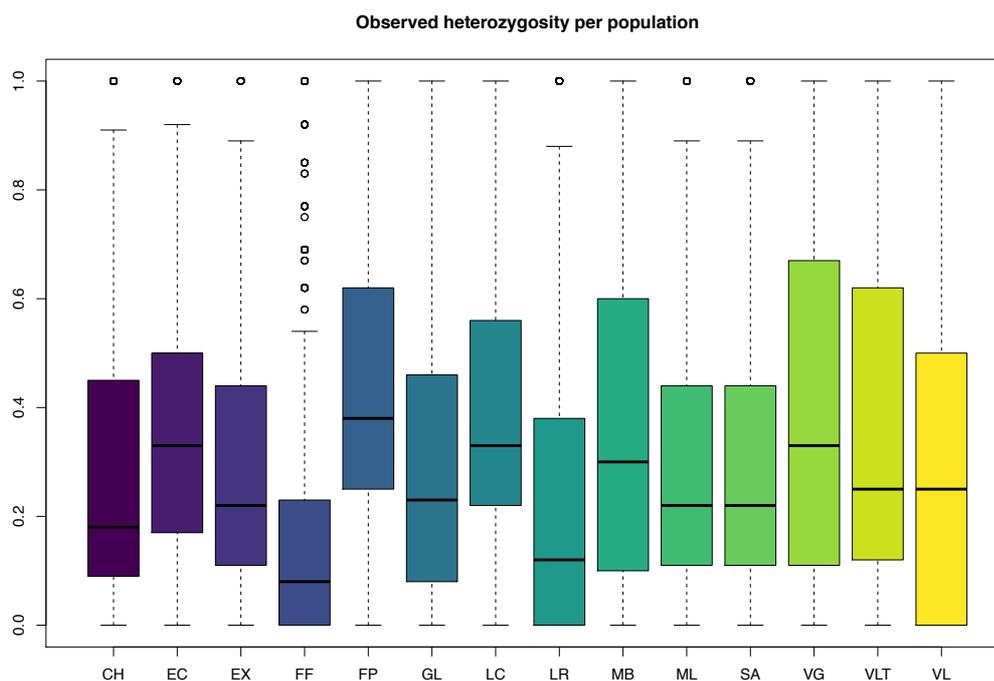


Figura 3: Eterozigotità osservata in ogni popolazione (si veda tabella 1 per la legenda delle località).

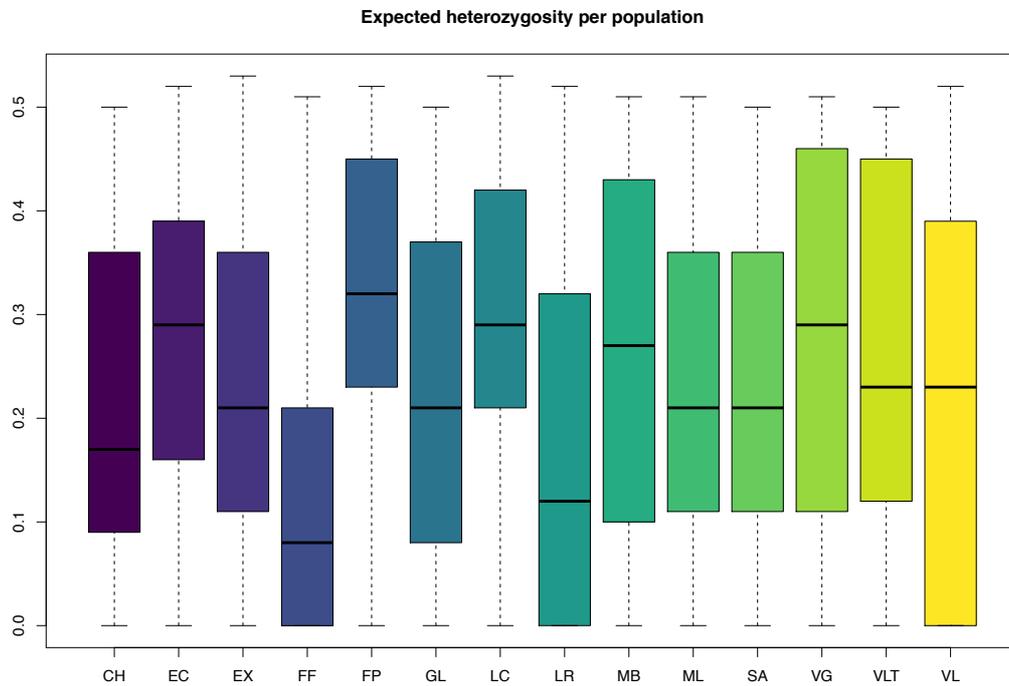


Figura 4: Eterozigotità attesa, ovvero diversità genetica all'interno di ogni popolazione (si veda tabella 1 per la legenda delle località).

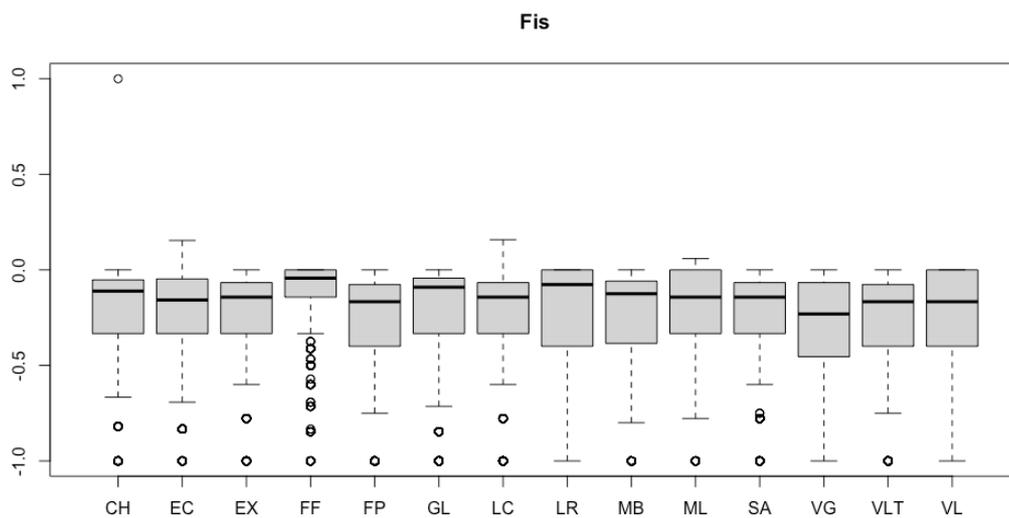


Figura 5: coefficiente di inbreeding per ogni popolazione (si veda tabella 1 per la legenda delle località).

Tabella 3: Indice di fissazione F_{st} fra le popolazioni.

	CH	EC	EX	FF	FP	GL	LC	LR	MB	ML	SA	VG	VL
CH													
EC	0,0424												
EX	0,0382	0,0273											
FF	0,0620	0,0684	0,0625										
FP	0,0617	0,0402	0,0448	0,1069									
GL	0,0242	0,0428	0,0350	0,0597	0,0531								
LC	0,0382	0,0356	0,0353	0,0746	0,0352	0,0332							
LR	0,0399	0,0636	0,0523	0,0419	0,0976	0,0428	0,0675						
MB	0,0386	0,0290	0,0318	0,0782	0,0360	0,0397	0,0328	0,0617					
ML	0,0666	0,0563	0,0563	0,0757	0,0838	0,0669	0,0751	0,0731	0,0598				
SA	0,0268	0,0370	0,0328	0,0560	0,0540	0,0278	0,0349	0,0419	0,0332	0,0649			
VG	0,0927	0,0612	0,0688	0,1333	0,0675	0,0848	0,0744	0,1158	0,0597	0,0593	0,0804		
VL	0,0897	0,0614	0,0682	0,1259	0,0785	0,0861	0,0784	0,1113	0,0632	0,0499	0,0846	0,0450	
VL	0,0867	0,0597	0,0644	0,1027	0,0820	0,0822	0,0817	0,0972	0,0662	0,0460	0,0792	0,0530	0,0492

La Discriminant Analysis of Principal Components DAPC calcolata sul dataset (fig. 5) mostra una distribuzione abbastanza omogenea dei campioni e delle popolazioni delle Aree 1, 2, 4, 5 e 6, le popolazioni quindi non sembrano essere fortemente strutturate secondo la distribuzione geografica.

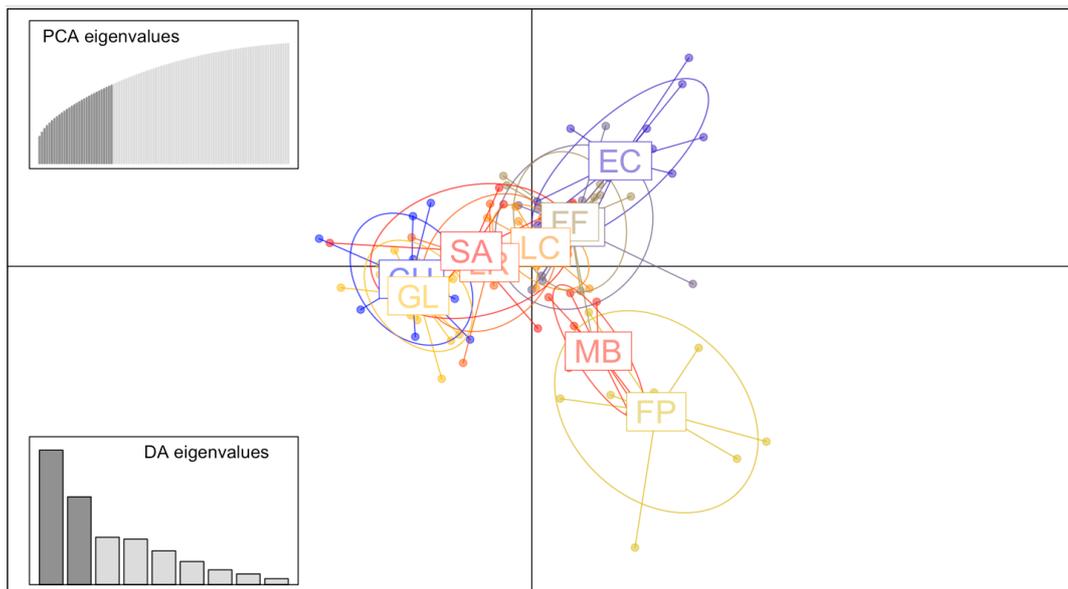


Figura 6: DAPC calcolata sul dataset di SNPs dei campioni provenienti delle aree 1, 2, 4, 5, 6 (vedi figura 1 per legenda delle popolazioni).

Il modello di *Isolation by distance* descrive il processo secondo cui le differenze genetiche fra popolazioni dovrebbero aumentare con la distanza geografica fra popolazioni. In figura 7 sono stati rappresentati i valori linearizzati di F_{st} in relazione alla distanza geografica lineare.

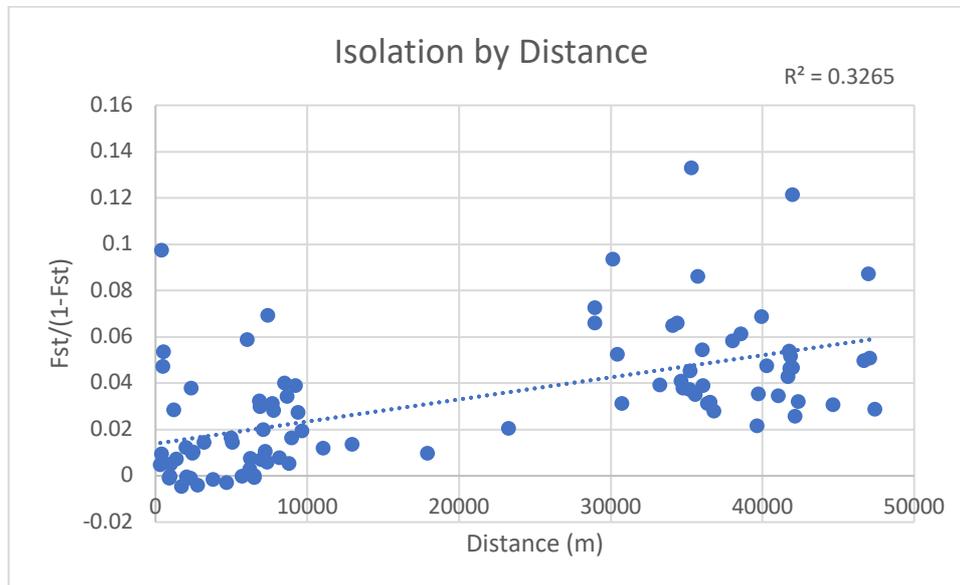


Figura 7: Grafico del test Isolation by distance

Il flusso genetico fra le popolazioni è stato stimato con il software DivMigrate. Il tasso di migrazione fra popolazioni viene calcolato confrontando la diversità genetica fra due popolazioni e un gruppo di migranti simulato. I tassi di migrazione sono stati misurati con distanza genetica D sul dataset comprendente tutte le popolazioni indagate (Fig. 8).

I modelli calcolati evidenziano come le migrazioni statisticamente significative avvengono dalla popolazione LR verso MB, EC (con migrazione relativa rispettivamente di 0.7 e 0.98) e verso VG, VLT, FP. Da VL avviene migrazione verso FP e da FF avviene migrazione verso VLT e VG. I flussi di migranti da e verso le altre popolazioni sono assenti o molto limitati.

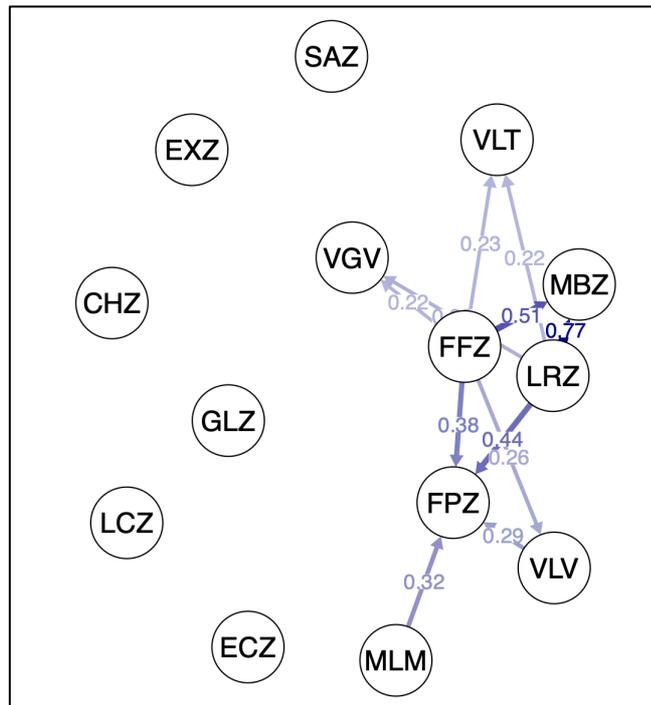


Figura 8: Flusso genetico fra le aree di campionamento stimato con il software DivMigrate. Network della migrazione relativa direzionale significativa calcolato con metodo Nm (Alcala et al., 2014, the effective number of migrants)

Riguardo alla diversità tra gli individui che verranno messe in connessione con opere di traslocazione e creazione del corridoio ecologico, le popolazioni dell'Area 1 e Area 2 risultano geneticamente compatibili, i risultati ottenuti da entrambe le analisi di sequenze di marcatori nucleari e mitocondriali e di mutazioni puntiformi (SNPs) non hanno infatti rilevato la presenza di una struttura genetica peculiare (aplotipi privati, distanze e differenziazione genetica) per una o entrambe le popolazioni.

3.3 Analisi su SNPs nelle Aree 1 e 2, campionamento 2022.

Le analisi effettuate tramite RAD sequencing sui campioni prelevati nella primavera 2022, a seguito delle opere di traslocazione e di creazione del corridoio ecologico, mostrano come gli indici di eterozigotità osservata e attesa siano comparabili o superiori a quelli misurati precedentemente: eterozigotità attesa 0.3104 per Area 1 e 0.2552 per Area 2; eterozigotità osservata 0.4677 per l'Area 1 e 0.3752 per l'Area 2 (figure 9, 10). Anche il coefficiente di *inbreeding* si mantiene alto (fig. 11). Gli indici mostrano quindi una fase di eccesso di eterozigoti compatibile con la presenza di flusso

genico fra due popolazioni. In figura 12 è rappresentato il risultato dell'analisi condotta con il software DivMigrate (Sundqvist et al., 2013) il livello di migrazione relativo tra le popolazioni campionate, l'analisi evidenzia un flusso genico unidirezionale dalla popolazione dell'Area 1 verso la popolazione dell'Area 2.

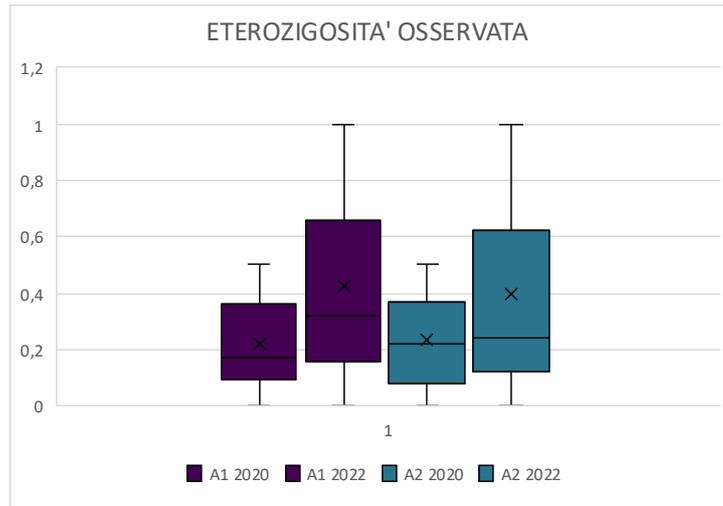


Figura 9: Confronto delle eterozigosità osservate per area calcolata su campionamento 2020 e

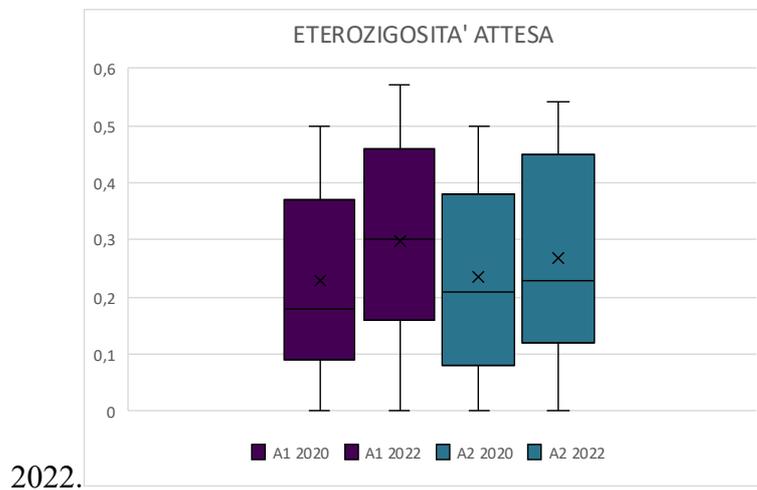


Figura 10: Confronto delle eterozigosità attese per area, calcolata su campionamento 2020 e 2022.

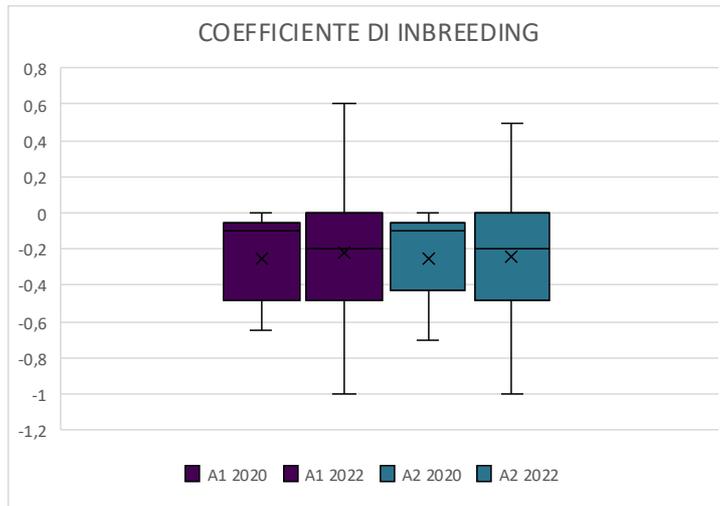


Figura 11: Confronto degli indici di inbreeding, per area calcolata su campionamento 2020 e 2022.

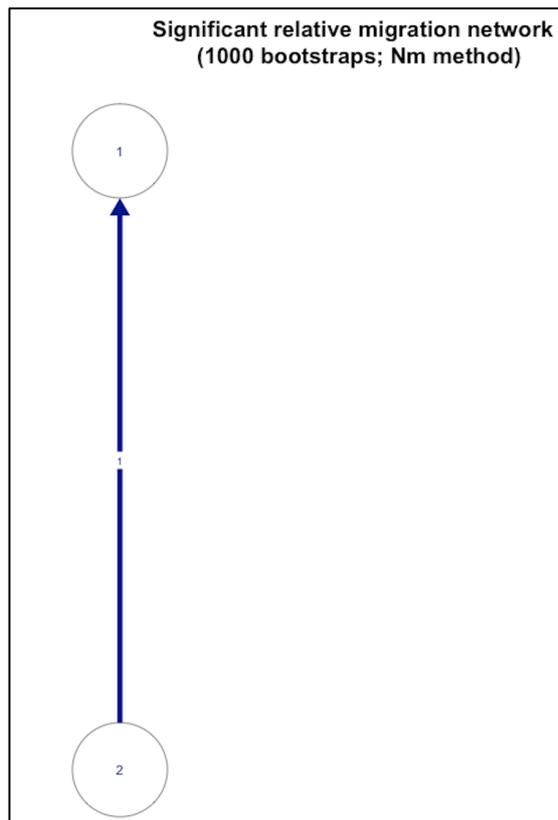


Figura 12: Flusso genetico fra le aree di campionamento stimato con il software DivMigrate.

Network della migrazione relativa direzionale significativa (bootstrap =1000) calcolato con metodo Nm (Alcala et al., 2014, the effective number of migrants).

Il modello evidenzia che la migrazione relativa direzionale statisticamente significativa, che si registra stimata sui campioni analizzati nel 2022 avviene dall'Area 2 verso l'Area 1.



UNIVERSITÀ DEL PIEMONTE ORIENTALE

4. Discussione e Conclusioni

Il progetto di caratterizzazione genetica e genomica delle popolazioni di *Zerynthia polyxena* della Val di Susa è stato effettuato attraverso l'analisi di marcatori mitocondriali e nucleari classici (sequenze nucleotidiche) e SNPs con un approccio RAD sequencing. I campioni, una o due zampe centrali, sono stati prelevati dagli individui in modo da creare meno disturbo possibile, durante le operazioni di cattura -marcatura e ricattura.

I risultati delle analisi sui marcatori nucleari e mitocondriali, gene *wingless* e *cytochrome c oxidase*, hanno rilevato rispettivamente 5 e 4 nuovi aplotipi (sequenze nucleotidiche), rispetto a quelli presenti in letteratura (Zinetti et al. 2013), evidenziando variabilità genetica abbinata a una mancanza di strutturazione poiché nelle popolazioni sono presenti molti aplotipi condivisi.

Le analisi genomiche condotte sulle popolazioni di *Zerynthia polyxena* della Val di Susa hanno evidenziato una alta diversità genetica all'interno delle popolazioni (alti tassi di eterozigotità osservata, media $H_o = 0.34$, si vedano grafici Fig. 3- 4, 9-10), bassi tassi di differenziazione genetica fra le popolazioni (F_{st} medio = 0,03) e di inbreeding ($F_{is} = -0,39$). Gli elevati valori di eterozigotità possono indicare fenomeni di *outbreeding* o una selezione dipendente dalla frequenza che avviene in popolazioni dove gli eterozigoti vengono selezionati positivamente.

Le popolazioni con stati di diversità e differenziazione genetica nonché *inbreeding* come quelli rilevati in *Zerynthia polyxena* sono quelle con valori ottimali per popolazioni che vanno incontro piani di gestione e conservazione (Hoban et al. 2022) ed hanno permesso di confermare la compatibilità degli individui presenti in Area 1 e in Area 2.

Le analisi genomiche hanno permesso anche di stimare il flusso genico fra le popolazioni, evidenziando come ci siano movimenti direzionali statisticamente significativi anche se limitati solo ad alcune popolazioni (fig. 8); le analisi condotte su Area 1 e 2 nel 2022 dopo gli interventi di conservazione rilevano un flusso statisticamente significativo fra le due popolazioni (figura 12).

5. Bibliografia

Alcala, N., J. Goudet, and S. Vuilleumier. 2014. On the transition of genetic differentiation from isolation to panmixia: What we can learn from G S T and D. *Theor. Popul. Biol.* 93:75–84. Elsevier Inc.



UNIVERSITÀ DEL PIEMONTE ORIENTALE

- Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:1–10.
- Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis. 2004. Identification of Birds through DNA Barcodes. *PLoS Biol.* 2:e312.
- Hoban, S., Archer, F. I., Bertola, L. D., Bragg, J. G., Breed, M. F., Bruford, M. W., ... & Hunter, M. E. 2022. Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. *Biological Reviews*, 97: 1511–1538.
- Keenan, K., P. Mcginnity, T. F. Cross, W. W. Crozier, and P. A. Prodöhl. 2013. DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* 4:782–788.
- Koscinski, D., Crawford, L.A., Keller, H.A. & Keyghobadi, N. 2011: Effects of different methods of non-lethal tissue sampling on butterflies. *Ecol. Entomol.* 36: 301–308.
- Pellegrino, I., L. Boatti, M. Cucco, F. Mignone, T. N. Kristensen, N. Mucci, E. Randi, A. Ruiz-Gonzalez, and C. Pertoldi. 2016. Development of SNP markers for population structure and phylogeography characterization in little owl (*Athene noctua*) using a genotyping-by-sequencing approach. *Conserv. Genet. Resour.* 8:13. Springer Netherlands.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. De Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.
- Sundqvist L, Keenan K, Zackrisson M, Prodöhl P, Kleinhans D. 2016. Directional genetic differentiation and relative migration. *Ecol Evol.*;6(11):3461-3475. doi: 10.1002/ece3.2096.
- Zinetti, F., L. Dapporto, A. Vovlas, G. Chelazzi, S. Bonelli, E. Balletto, and C. Ciofi. 2013. When the Rule Becomes the Exception. No Evidence of Gene Flow between Two *Zerynthia* Cryptic Butterflies Suggests the Emergence of a New Model Group. *PLoS One* 8.

Alessandria, 30 settembre 2022.


Dr.ssa Irene Pellegrino


Prof. Marco Cucco