Promosso da:
MINISTERO DELL'AMBIENTE E DELLA SICUREZZA ENERGETICA
Italiadomani
Funded by the European Union NextGenerationEU

Analisi genomiche (genomi *de novo* e *resequencing*) su campioni di tessuti muscolari e/o prelievi ematici di specie di vertebrati di interesse conservazionistico e/o gestionale

Committente: MASE, Ministero dell'Ambiente e della Sicurezza Energetica
Assegnatario: Consorzio Futuro in Ricerca, Ferrara
Responsabile scientifico: Giorgio Bertorelle, Università di Ferrara
Co-responsabili scientifici: Andrea Benazzo, Roberto Biello, Silvia Fuselli, Emiliano Trucchi
Collaboratori scientifici: Sonia Celestini, Linda Cerioni, Valentina Giongo, Jacopo Martellossi, Agnese Roscioni, Nicola Zadra. Project manager: Amiel Masarap

BREVE MANUALE UTENTE di www.genodiv.it

Sintesi
Il sito Genodiv (www.genodiv.it) è uno strumento di supporto per tecnici, funzionari, e manager di istituzioni pubbliche e Parchi Nazionali impegnati nelle attività di conservazione e gestione della biodiversità. In particolare, il sito contiene i risultati di uno studio genomico su 29 specie di vertebrati distribuiti sul territorio nazionale. Queste informazioni possono contribuire ad orientare le scelte di policy, le priorità di monitoraggio e le eventuali azioni di tutela nei Parchi Nazionali e in aree al di fuori dei Parchi. Il sito è concepito come una struttura aperta, aggiornabile ed espandibile: la sua architettura consente infatti di integrare in futuro nuove specie, nuovi siti di campionamento e ulteriori risultati di ricerca.

Accesso
Attualmente, le credenziali per accedere al sito possono essere richieste alla propria istituzione di appartenenza (MASE, ISPRA o Parchi Nazionali).

Scopo del manuale
Questo documento spiega come usare il sito Genodiv e come orientarsi tra le sue sezioni principali. Il portale raccoglie le informazioni di base e i risultati ottenuti in due progetti finanziati dal MASE con fondi PNRR dedicati al campionamento, al sequenziamento e alle analisi genomiche di specie di vertebrati italiani di interesse conservazionistico e gestionale. Il manuale è rivolto principalmente a utenti non specialisti in genomica. Per facilitarne la consultazione, la navigazione del portale è stata progettata in modo semplice e intuitivo; inoltre, i principali termini tecnici e le metodologie impiegate sono descritti nelle sezioni dedicate al **Glossario** e ai **Metodi**, mentre tutti i risultati principali sono accompagnati da didascalie e testi esplicativi che ne facilitano l'interpretazione.

Sommario

1. Struttura del sito *3*

2. Home *3*

3. Tutte le specie *3*

3.1 Pagina generale della specie **3**

3.2 Macroaree **4**

3.3 Siti di campionamento per la macroarea **5**

3.4 Gli individui **5**

3.5 Genetica della specie **5**

3.7 Download dati **5**

3.8 IGV (*de novo*) **6**

4. Aree geografiche *7*

5. Progetto e gruppo di ricerca *7*

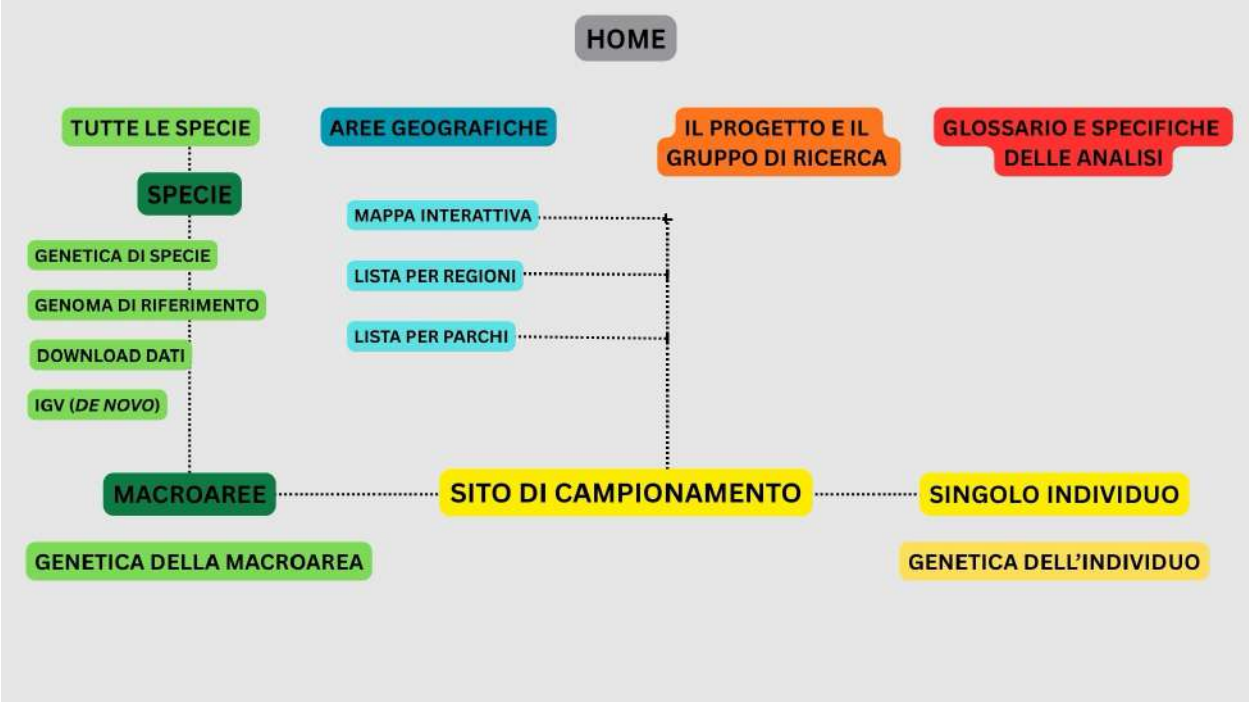
6. Glossario e specifiche delle analisi *7*

7. Conclusioni e consigli *7*

1. Struttura del sito

Il menu principale del sito comprende: Home, Tutte le specie, Aree geografiche, Il progetto e il gruppo di ricerca, Glossario e specifiche delle analisi.

La navigazione è gerarchica: da una visione generale si passa al dettaglio di specie, macroarea, sito di campionamento e singolo individuo.



2. Home

La pagina Home è il punto di ingresso del portale. Collegandosi all'indirizzo www.genodiv.it si accede a questa pagina, dove è possibile effettuare l'autenticazione e conoscere in estrema sintesi l'obiettivo del progetto.

3. Tutte le specie

La pagina "Tutte le specie" mostra i riquadri delle 35 specie o sottospecie analizzate (29 specie target e 6 specie o unità tassonomiche di confronto). Una tabella infografica mostra l'avanzamento complessivo del progetto nelle sue 3 principali fasi: estrazione DNA, sequenziamento e analisi bioinformatiche. Cliccando sul riquadro riferito a ciascuna specie o sottospecie si arriva alla **Pagina generale della specie**, contenente una fotografia, il nome della specie e una percentuale di avanzamento delle analisi.

3.1 Pagina generale della specie

Questa pagina raccoglie le informazioni generali della specie, la mappa interattiva dei siti di campionamento, la sezione dedicata alle **Macroaree**, ognuna rappresentata da un riquadro specifico, e il riepilogo dello stato di avanzamento delle analisi.

Da questa pagina è possibile scaricare il "**Report finale**" della specie, che include un riassunto dei risultati principali, la loro interpretazione in chiave evoluzionistica e conservazionistica, nonché le possibili implicazioni di carattere gestionale. La pagina funge inoltre da punto di accesso alle sezioni di "**Genetica**" della specie, "**Genoma di riferimento**", "**Download dati**" e "**IGV**" (figura sottostante). Quest'ultima sezione è disponibile solo per le 6 specie per cui il progetto prevedeva anche l'assemblaggio del genoma *de novo*.

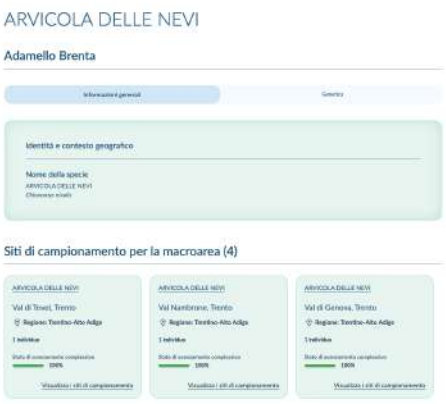
Pagina generale della specie, indicate le sezioni interne a cui è possibile accedere.



3.2 Macroaree

Nella **Pagina generale della specie** sono riportate con specifici riquadri le **macroaree** relative alla specie stessa. Una macroarea può comprendere uno o più **Siti di campionamento** ed è stata definita integrando risultati genetici con informazioni geografiche. Accedendo alla pagina della macroarea è possibile visionare la scheda **Genetica** specifica per la macroarea. Questa pagina riassume i risultati delle analisi bioinformatiche relative agli individui appartenenti a quell'area geografica. Questa scheda può mancare nel caso in cui la macroarea comprenda un singolo individuo; in questo caso visionare la pagina **Genetica** dedicata al singolo individuo. Ogni pagina di macroarea mostra inoltre tutti i siti di campionamento inclusi nella macroarea con appositi riquadri.

Pagina di macroarea, con siti di campionamento



3.3 Siti di campionamento per la macroarea

Cliccando su "Visualizza i siti di campionamento" si apre la relativa pagina. Qui sono presenti una descrizione generale del sito, la macroarea di appartenenza, l'eventuale Parco Nazionale di cui fa parte, una mappa interattiva con la posizione degli individui, i riquadri dei singoli individui e un'infografica finale sullo stato di avanzamento delle analisi. Alcuni individui sono indicati come back-up: sono campioni tissutali disponibili ma che sono stati esclusi dal sequenziamento genomico finale in seguito a controlli di qualità, congruenza genetica, o semplicemente perché provenivano da siti già sufficientemente coperti da altri campioni.

3.4 Gli individui

Per consultare i dati relativi ad ogni singolo individuo campionato, si clicchi su "Approfondisci individuo" sul relativo riquadro. Ogni individuo campionato è identificato da un codice univoco composta da un prefisso riportante un'abbreviazione del nome scientifico della specie/sottospecie e un codice numerico. Da notare che per le lucertole (vivipara e carniolica) e per l'orbettino (italiano ed europeo) alcuni individui sono stati classificati nelle relative specie successivamente al campionamento, sulla base delle analisi genetiche. Il prefisso individuale non corrisponde quindi necessariamente alla specie di appartenenza. Nella pagina dell'individuo sono riportati i dati biologici disponibili e le informazioni di campionamento. Da questa pagina si accede anche alla pagina di genetica del singolo individuo, dove vengono presentati i risultati delle analisi riferite a quel campione.

3.5 Genetica della specie

Dalla **Pagina generale della specie** abbiamo visto che si può accedere alla sezione di **Genetica** che contiene i principali risultati ottenuti per la specie. Questa sezione presenta i risultati delle analisi biomolecolari e bioinformatiche riferite alla specie nel suo complesso. Qui vengono riassunti e descritti sottoforma di grafici, tabelle e didascalie i principali risultati ottenuti. Per le sottospecie e complessi di specie analizzati congiuntamente (scoiattolo nero e meridionale; astore sardo e continentale; sterpazzolina albistrata, comune, di Moltoni e iberica; orbettino europeo ed italiana, lucertola carniolica e vivipara), la scheda **Genetica** della specie è condivisa fra specie/sottospecie.

3.6 Genoma di riferimento

La sezione **Genoma di riferimento** riporta le informazioni principali sul genoma riferimento utilizzato come base per le analisi di *resequencing*, permettendo di visionarne caratteristiche e indici di qualità. Per i genomi disponibili pubblicamente sul portale del *National Center for Biotechnology Information* (NCBI), sono riportati il link alla risorsa originale e le statistiche correlate. Per i genomi generati *de novo* nell'ambito del progetto, sono invece disponibili statistiche aggiuntive derivanti da specifiche analisi bioinformatiche.

3.7 Download dati

La sezione **Download dati** consente agli utenti autorizzati di scaricare i principali dati pre-processati, quali:

- VCF filtrati ed utilizzati per le analisi di genomica di popolazione sia per polimorfismi a singolo nucleotide (SNP) che per polimorfismi di piccole inserzioni e delezioni (INDEL).
- Valori di differenziazione genomica F_{ST} per finestre genomiche per ogni confronto a coppie tra macroaree. Questi valori sono riportati in un unico file in formato "CSV", con colonne separate da tabulazioni, in cui le prime tre indicano le coordinate genomiche e le successive i valori di F_{ST} per ogni confronto. L'ultima colonna riporta sempre gli identificativi di eventuali geni (presenti anche nei file di annotazione e nelle sequenze nucleotidiche e amminoacidiche scaricabili dal portale) che ricadono nella relativa finestra.
- Posizione delle regioni mascherate durante la fase di filtraggio dei siti varianti, corrispondenti a regioni ripetute e/o a bassa mappabilità in formato "bed".
- Annotazione genica in formato "GFF3".
- Sequenze tradotte in amminoacidi dei trascritti codificanti predetti dall'annotazione genica in formato "fasta".
- Sequenze nucleotidiche dei trascritti codificanti predetti dall'annotazione genica in formato "fasta".
- Genoma di riferimento in formato "fasta".

3.8 IGV (*de novo*)

Per le specie per cui il genoma è stato generato *de novo*, la sezione **IGV** ospita un'istanza web del genome browser *Integrative Genomics Viewer*. Tramite questa sessione interattiva è possibile esplorare nel dettaglio il genoma di riferimento insieme a:

- Annotazione genica
- Annotazione degli elementi ripetuti.
- Le varianti SNP con i relativi genotipi per singolo individuo analizzato, nonché il loro eventuale effetto su sequenze codificanti e siti di splicing; per ogni individuo è inoltre associata la relativa macroarea di appartenenza tramite il codice colore presente nelle pagine "Genetica" della specie.
- I valori di differenziamento genomico lungo finestre genomiche F_{ST} per ciascun confronto tra macroaree.

Le sessioni di IGV possono essere espanse per una visualizzazione a schermo intero premendo sul tasto "Espandi a schermo intero" presente a destra sopra la visualizzazione integrata nel portale; una volta entrati a schermo intero, è possibile uscire dalla modalità utilizzando il tasto "Chiudi schermo intero". Le sessioni sono personalizzabili dall'utente; tuttavia, a ogni ricaricamento della pagina, la visualizzazione verrà ripristinata alle impostazioni di default. Da notare che, per questioni di velocità di navigazione, si consiglia di visualizzare annotazioni geniche, elementi ripetuti e SNP solo per regioni genomiche relativamente piccole, eseguendo degli ingrandimenti. Per le annotazioni geniche ed elementi ripetuti, i valori di default sono impostati automaticamente da IGV sulla base della loro densità, mentre i siti varianti per SNP vengono visualizzati per regioni genomiche inferiori a 50 kilobasi; tali soglie possono essere modificate dall'utente, ma si sconsiglia di superare il doppio dei valori di default.

Informazioni dettagliate riguardo ogni dato genomico visualizzato possono essere ottenute cliccando col tasto sinistro del mouse sul relativo dato. Per i valori di F_{ST}, ad ogni finestra genomica (50Kb) è associato il relativo valore di differenziazione; per i geni le informazioni principali riguardano l'identificativo e il numero di esoni. Per gli elementi ripetuti è riportata l'annotazione in termini di Classe/Superfamiglia secondo gli standard RepeatMasker/DFAM (https://www.dfam.org/home). Per i siti varianti per SNP sono presenti sia informazioni generali sul sito (allele di riferimento e alternativo, annotazione dell'effetto) premendo sulla traccia blu scuro in cima a quella degli SNP, sia informazioni individuali sul genotipo, indicate sia da codice colore (grigio=omozigote per l'allele di riferimento, azzurro=omozigote per l'allele alternativo, blu=eterozigote) sia per esteso cliccando sul rettangolo corrispondente all'individuo.

Per approfondire il funzionamento del software, è possibile consultare la funzione "Help" integrata nell'interfaccia oppure visitare il repository ufficiale: https://github.com/igvteam/igv-webapp. Ulteriori informazioni riguardo specifiche impostazioni delle sessioni per ogni specie sono presenti sul portale sulla relativa scheda **IGV**.

4. Aree geografiche

La pagina **Aree geografiche** permette una lettura territoriale del progetto. La mappa interattiva consente di cercare e aprire tutti i siti di campionamento. La stessa logica vale per la lista per regioni e per la lista per Parchi Nazionali: cliccando sui riquadri si apre la scheda del sito, con il nome della specie associata. Nei Parchi Nazionali, la visualizzazione mostra tutti i siti presenti nel parco selezionato.

5. Progetto e gruppo di ricerca

La pagina "Il progetto e il gruppo di ricerca" presenta il progetto, i suoi obiettivi e il gruppo che ha svolto le attività di campionamento, laboratorio, analisi bioinformatica e interpretazione dei risultati.

6. Glossario e specifiche delle analisi

Questa sezione aiuta la lettura del sito. Il glossario spiega i termini tecnici usati nelle altre pagine, mentre le specifiche delle analisi riportano metodi, software e passaggi bioinformatici in forma più dettagliata. È la sezione da consultare quando si vuole interpretare correttamente un risultato o comprendere il significato di un termine genetico, evoluzionistico, o di biologia della conservazione.

7. Conclusioni e consigli

Tutte le informazioni e i risultati contenuti nel sito sono organizzate sulla base di tre principali livelli gerarchici: specie, macroarea (che può comprendere uno o più siti di campionamento geograficamente vicini), individuo. Per ciascun livello sono disponibili informazioni descrittive e i risultati delle analisi genomiche.

Se l'interesse principale riguarda una determinata specie, si consiglia di iniziare la consultazione dalla pagina **Tutte le specie**, accedere alla scheda della specie desiderata e quindi alla sezione

Genetica, dove sono riportati i risultati delle analisi genomiche. Dalla stessa scheda della specie è inoltre possibile accedere alla sezione **Download dati** per scaricare i dati disponibili. Qualora si desideri approfondire una specifica popolazione o un singolo campione, è possibile navigare verso le relative pagine dedicate alle macroaree o ai singoli individui, che riportano informazioni e risultati riferiti a quel livello.

Se invece l'interesse riguarda una particolare area geografica o un Parco Nazionale, è consigliabile iniziare la consultazione dalla pagina **Aree geografiche**, che permette di visualizzare tutte le specie campionate e i relativi risultati disponibili per quell'area o Parco.