Piattaforma GenPat

Manuale d'uso

Unità Bioinformatica

Indice

1. Home

1.1 Piattaforma GENPAT

1.1.1 Introduzione

Il Ministero della Salute, con Decreto del 30 maggio 2017 (G.U.R.I. n. 196 del 23 agosto 2017), ha attivato il *Centro di Referenza Nazionale per Sequenze Genomiche di microrganismi patogeni: banca dati e analisi di bioinformatica* presso la sede centrale dell'Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise (IZSAM).

La mission e gli obiettivi del Centro sono:

- realizzare una piattaforma nazionale per la raccolta e la conservazione delle sequenze genomiche di microrganismi patogeni, per l'esecuzione di analisi bioinformatiche e per l'archiviazione e la condivisione dei risultati;
- realizzare un sistema strutturato e permanente di referenti, all'interno dei singoli Istituti Zooprofilattici Sperimentali, ai fini di un miglior coordinamento delle attività che saranno poste in essere sul territorio nazionale;
- fornire assistenza tecnico-scientifica al Ministero della Salute ed alle autorità competenti;
- curare l'organizzazione di corsi di formazione, nell'ambito delle proprie competenze, per il personale del Servizio Sanitario Nazionale e per gli operatori di altri Enti competenti;
- promuovere le attività di ricerca nel settore di competenza;
- implementare ogni altra utile attività, attinente alle proprie competenze, ivi comprese la collaborazione e il coordinamento con altre amministrazioni ed Enti del settore.

Responsabile: Cesare Cammà - c.camma@izs.it

1.1.2 Gestione campioni ufficiali

In accordo con i relativi laboratori nazionali di riferimento, a partire da Lunedì 3 Giugno 2024, utenti autorizzati possono visualizzare metadati minimi e risultati di caratterizzazione genomica di tutti i ceppi di *L. monocytogenes* e *Campylobacter* isolati nel territorio nazionale, nell'ambito del controllo ufficiale e delle altre attività ufficiali, come definite dal Regolamento (UE) 2017/625 e dal D. Lgs. 2 febbraio 2021, n.27.

Maggiori informazioni possono essere trovate nella sezione dedicata Gestione campioni ufficiali.

1.1.3 Link rapidi

- Quick start
- Gestione campioni ufficiali
- Funzionalità Principali
- Video utili

1.2 Avvio rapido

1.2.1 Guida all'Uso

Nel seguente video vengono mostrate le principali attività eseguibili sulla piattaforma, tra cui:

- gestione della tabella dei campioni;
- utilizzo di filtri, selezione di campioni ed operazioni su di essi;
- lancio di analisi su campioni selezionati;
- sorveglianza sullo stato dell'analisi lanciata e visualizzazione dei risultati.

Nota: Per questioni di privacy e sicurezza dei dati, il video è stato realizzato nell'applicativo demo Cohesive, pertanto l'interfaccia può presentare alcune marginali differenze.

Attività principali eseguibili

1.2.2 Procedura di Upload

I 2 video seguenti mostrano le procedure di upload dei campioni nella piattaforma. Nel primo video viene caricato un file compresso in formato `.zip`, contenente i file `.fasta` con i relativi metadati; il secondo video mostra l'importazione di sequenze direttamente da **NCBI**.

Upload di file di sequenze

Upload di file di sequenze

Upload da NCBI

Upload da NCBI

1.2.3 Procedura di Download

E' possibile scaricare direttamente i campioni usando una raccolta di campioni da carrello o un tag attivo. Il seguente video fornisce una breve dimostrazione di una delle procedure di download disponibili; maggiori informazioni, insieme da altri tutorial, sono disponibili nella sezione del wiki sulla procedura di download.

Procedura di download

1.2.4 Profili Utente

Ad ogni utente può essere associato più di un profilo e, dunque, gli utenti possono appartenere a diversi **gruppi**.

Gli utenti potranno visualizzare campioni o effettuare operazioni relativamente al gruppo a cui appartengono e limitatamente a ciò che è consentito agli utenti di tale gruppo. I gruppi possono quindi assumere, dal punto di vista dell'utente, un ruolo di filtro generale.

Il video che segue mostra come cambiare gruppo, evidenziando come questo influisca sul menù di navigazione principale e sui campioni che potranno essere visualizzati.

Nota: Per questioni di privacy e sicurezza dei dati, il video è stato realizzato nell'applicativo demo Cohesive, pertanto l'interfaccia può presentare alcune marginali differenze.

Cambiare gruppo utenti

1.3 Funzionalità

1.3.1 Caratteristiche principali

Flessibilità dello schema dei database

Il sistema possiede degli schemi di gestione dei database estremamente flessibili e non è necessaria alcuna fase di sviluppo software per effettuare operazioni come la creazione di nuove tabelle o di nuovi campi nei moduli.

Il prossimo video mostra come, oltre a definire attributi addizionali, sia possibile creare da zero un nuovo database personalizzato, attraverso l'utilizzo di:

- *classi*, ovvero le tabelle dei database;
- *attributi* di classe, ovvero le colonne delle tabelle (presenti per ogni tipo, incluse le liste con valori chiusi, singoli o multipli, chiavi esterne, files o formule) ed il relativo layout;
- *domini*, ovvero tipi di relazioni, anche con eventuali attributi specifici aggiuntivi;
- *tipi di lookup*, ovvero liste con valori chiusi che possono essere associate ad un attributo di classe;
- *report*, ovvero la specializzazione di una classe in "sottoclassi" (anche a più livelli), aggiungendo attributi e domini specifici, condividendo gli attributi di superclasse;
- *viste* basate su filtri o query SQL;
- comportamenti dell'interfaccia utente specifici per ciascuna classe: validazione dei dati, automatismi durante l'aggiornamento dei dati, menu contestuali, guida in linea, ecc.

Nota: Per questioni di privacy e sicurezza dei dati, alcuni dei video presenti nella pagina sono stati realizzati nell'applicativo demo Cohesive, pertanto l'interfaccia può presentare alcune marginali differenze.

Flessibilità dei database

Modifica generale

Questa opzione consente di cambiare il valore di uno o più attributi di una selezione di cards dalla classe corrente. Tale funzione può essere attivata dal menù contestuale della classe, abilitando la selezione multipla.

Modifica generale

Importazione / Esportazione dati

La piattaforma supporta l'importazione e l'esportazione dei dati nei formati `CSV`, `XLS`, `XLSX` nonchè come database relazionali.

Il sistema si basa sulla configurazione di specifici "template", per garantire la gestione di operazioni complesse in maniera semplificata. Per una descrizione dettagliata e videoguide sull'argomento, si visiti la pagina dedicata all'Importazione / Esportazione dei dati di questa guida.

Tabelle

La piattaforma possiede una tabella dei campioni integrata nell'interfaccia: essa elenca le *entries* della piattaforma e permette di interagire con esse. Le tabelle supportano strumenti di personalizzazione e l'utilizzo di filtri, oltre a molte funzioni per esportare, scaricare, stampare o salvare le *entries* nel le Carrello. Le informazioni in merito ad una entry di interesse possono essere consultate espandendo la scheda informativa ad essa relativa. Una guida dettagliata, completa di video, è disponibile nella sezione dedicata alle tabelle.

Tag (o Motivo)

I *tag* ("motivi") sono una funzionalità della piattaforma che permette all'utente di riunire una selezione di campioni sotto una stessa "etichetta". Quando un tag è attivo, solo i campioni appartenenti a quel tag potranno essere visualizzati nella tabella dei campioni: in questo modo i tag possono agire da filtri.

Ulteriori informazioni sui tag sono disponibili sia nella pagina dedicata, sia nella sottosezione sulla gestione dei tag, dove sarà possibile trovare video dimostrativi su come creare, attivare, modificare o eliminare un tag.

Carrello

Il carrello è una funzione che permette lo stoccaggio temporaneo di campioni selezionati e che può essere usata per creare nuovi tag o per lanciare analisi. Il lancio analisi potrà infatti essere effettuato usando un tag attivo o i campioni presenti nel carrello. Il carrello presenta anche la modalità **Attiva Carrello**, che permette di operare sulla piattaforma filtrando tutti i campioni *e le analisi lanciate*, visualizzando solo ciò che è relativo ai campioni presenti nel carrello.

Nell'introduzione della pagina Wiki dedicata al carrello è possibile trovare informazioni e videotutorial sulle funzioni principali del carrello. Per una trattazione più approfondita, si faccia invece riferimento alla pagina Gestione Carrello della guida.

Integrazione Wiki nel Lancia Analisi

Nel sistema di lancio analisi della piattaforma, per ogni analisi e pipeline disponibile e selezionata, potrai trovare, insieme ad una breve descrizione, anche un link alla relativa documentazione.

Integrazione Wiki

Ripeti Analisi

In piattaforma sono presenti 2 azioni rapide per permettere all'utente di richiedere, per un nuovo set di campioni, il lancio di un'analisi precedentemente completata con successo. Le 2 funzioni permettono di mantenere i parametri già usati in precedenza per ripetere l'analisi su un nuovo dataset, fornito da tag attivo o da una selezione di campioni presente nel carrello.

Nella pagina dedicata sono disponibili ulteriori informazioni ed un videotutorial.

Dashboard

La visualizzazione diretta dei risultati delle analisi effettuate, in piattaforma, è consentita grazie all'utilizzo di alcune *dashboards* e alcune funzioni, tra le quali una versione integrata ed estesa di GrapeTree, **SPREAD** (maggiori informazioni sono disponibili nell'apposita sezione di questa guida).

Nel video seguente viene mostrato lo sviluppo nel tempo dell'epidemia di Sars-Cov-2 in Abruzzo, tramite *SPREAD*. L'analisi mostrata è stata realizzata usando i metadati dei campioni: per costruire l'albero vengono usate le date, raggruppate per anno/mese di campionamento, mentre per la collocazione sulla mappa sono utilizzati i relativi dati GeoJSON.

Timelapse di SPREAD

1.3.2 Caratteristiche delle tabelle

Gli **Elementi principali** della piattaforma, ovvero Campioni, codici Alias, Metadati e Risultati analisi, sono gestiti tramite dei sistemi di visualizzazione in formato tabulare integrati nella piattaforma.

Gli utenti possono interagire con le *entry* nella tabella di ciascun elemento principale e personalizzarla sulla base delle proprie preferenze e necessità. Le tabelle supportano l'utilizzo di filtri di base ed avanzati, oltre ad opzioni di esportazione, download e stampa.

> **Nota:** Si prega tener presente che le tabelle degli **Elementi principali** condividono una medesima gestione e simili funzionalità, pur considerando, tuttavia, la presenza di alcune opzioni specifiche previste nel menù contestuale della tabella di ciascun elemento.

Gestione della tabella e personalizzazione

Il video che segue mostra come una tabella possa essere gestita e modificata cambiando l'ordine e le dimensioni delle colonne, scegliendo quali di esse visualizzare e cambiando l'ordine ascendente o discendente per ogni colonna. Le modifiche apportate ad una tabella sono transienti, a meno che non vengano salvate tramite l'apposito pulsante; esse possono essere resettate rimuovendo le preferenze salvate, come mostrato nel video.

Video dimostrativo sulla personalizzazione della tabella

> **Nota** Le preferenze di griglia vengono salvate in base al gruppo selezionato al momento del salvataggio. Di conseguenza, se un'utenza è assegnata a più gruppi, è possibile avere configurazioni delle preferenze diverse per ciascun gruppo.

Le colonne delle tabelle presenti in piattaforma adattano automaticamente la loro larghezza in base allo spazio disponibile; per alcune tabelle è possibile disattivare questo comportamento automatico tramite un toggle situato in alto a destra, così da regolare la larghezza delle colonne in base al contenuto o alle proprie preferenze. Se la larghezza totale della tabella supera lo spazio orizzontale disponibile, è possibile navigare tra le colonne utilizzando lo scroll orizzontale.

Cambiare vista delle tabelle.

Filtri base

Cliccando sull'icona a imbuto, è possibile applicare filtri di base alle colonne, da soli o in combinazione, per cercare le *entry* mostrate nella tabella o restringere il campo di ricerca.

Video sui filtri base

Dettaglio di una entry

Facendo doppio click sulla riga di un'entry (o, in alternativa, cliccando sul toggle **[+]** dell'entry di interesse), gli utenti possono espandere la scheda di un item e avere accesso ad ulteriori risorse e links.

Video dimostrativo sull'apertura di un dettaglio di una entry

Filtri avanzati

Le tabelle permettono anche l'utilizzo di filtri avanzati, applicabili alla lista di *entries*. L'icona , infatti, permette sia di scegliere un attributo tramite l'opzione *Aggiungi filtro*, sia di utilizzare filtri già salvati in precedenza.

I filtri avanzati permettono la ricerca per uno o più attributi (che possono essere rappresentati da stringhe di testo, date o metadati), tramite l'uso di uno o più termini di ricerca o di operatori booleani per ogni attributo. I filtri avanzati, inoltre, possono essere modificati, copiati, salvati e applicati di default alla tabella.

Uso dei filtri avanzati

Menù contestuale

Oltre ai filtri, la tabella dei campioni mette a disposizione dell'utente un menù contestuale dal quale è possibile personalizzare le opzioni di visualizzazione e disabilitare o abilitare la selezione multipla delle *entry*. Ulteriori opzioni previste nel menù contestuale, come sistemi di importazione ed esportazione, possono essere disponibili solo per la tabella *Campioni* o differire tra le tabelle degli elementi principali.

Video su come utilizzare il menu contestuale

Nota Importante: le funzioni disponibili nel menù contestuale **richiedono la selezione dei campioni di interesse** nella pagina, esattamente come per l'aggiunta di campioni al carrello. La selezione multipla, attiva di default, può essere disattivata e attivata tramite l'apposita voce nel menù contestuale o tramite la scorciatoia da tastiera `Ctrl+Shift+m`. Qualunque tentativo di esportare un file, senza aver prima fornito indicazioni sugli oggetti che si desidera esportare, risulterà in un fallimento o nell'esportazione di un file vuoto.

Albero gerarchico

Nel menu contestuale delle tabelle che contengono informazioni sulla **gerarchia** o sull'elemento **padre** delle *entry* (ad esempio, nei metadati), è possibile trovare l'icona `Vedi albero gerarchico` l'attivazione della quale fornisce una visualizzazione ad albero interattiva degli elementi della tabella, come brevemente mostrato nel video che segue.

Video albero gerarchia

Nota Importante: se la tabella contiene più di 2000 elementi (righe) la funzione `Espandi vista` viene automaticamente disabilitata.

Esportazione del Codice Risultato (RIS_CD)

Le pagine dei Controlli qualità elencano le analisi effettuate ed importate identificandole con un apposito codice univoco, il **RIS_CD**.

Un **Codice Risultato (RIS_CD)**, (ad esempio, "210712-10003548-4TY_cgMLST-chewbbaca"), identifica un'analisi lanciata in una specifica data, insieme ai risultati corrispondenti. **Le prime 6 cifre del prefisso numerico esprimono la data** di esecuzione dell'analisi (in formato "YYMMDD"); la parte finale è costituita dal nome dell'analisi ("4TY_cgMLST-chewbbaca").

Nelle varie tabelle *Controlli qualità*, è possibile esportare i codici selezionati come CSV o copiarli. Questo consente di usare i codici RIS_CD per selezionare campioni per il processo di lancio analisi o per il download dei risultati.

Nota: La selezione tramite **Codice Risultato (RIS_CD)** permette di scaricare o lanciare analisi anche su risultati meno recenti, dando accesso quindi a **tutte le analisi importate** in piattaforma. Per lanciare analisi sui risultati di un'analisi più datata, è quindi necessario usare il **Codice Risultato RIS_CD**.

Il breve video che segue mostra il flusso di lavoro da seguire per selezionare gli input di un'analisi (o i file da scaricare) usando il RIS_CD.

Demo sull'esportazione di RIS_CD

1.3.3 Importazione / Esportazione dei dati tramite file

Il sistema offre la possibilità di importare e di esportare dati attraverso sorgenti disponibili nei formati più comuni:

- File di tipo `CSV`, `XLS` e `XLSX` (che possono essere anche esportati);
- database relazionali.

Il sistema è basato sulla configurazione di *template* specifici al fine di semplificare la complessità delle operazioni. Ciascun template fornirà autorizzazioni di accesso diversificate e includerà tutte le informazioni utili all'automazione delle attività: tipo di operazione, mappatura degli attributi, modalità di sincronizzazione (chiave unica e modalità di eliminazione), ecc.

Durante l'importazione, è possibile operare in modalità *merge* che consente di: - modificare le stringhe esistenti (identificate tramite una chiave unica); - inserire le nuove stringhe; - rimuovere quelle non più presenti (tramite cancellazione logica o aggiornamento dello stato dell'applicazione).

È presente una gestione completa degli errori, visualizzata nell'interfaccia utente in caso di utilizzo in modalità interattiva e inviata via e-mail. Una volta corretti gli errori, il file può essere ricaricato senza doverlo modificare, poiché le stringhe già importate verranno ignorate.

Template 1
Le funzionalità di import and export sono accessibili dal menu contestuale delle classi di CMDBuild, per le quali è stato configurato un template di importazione e/o esportazione.

Template 2

2. Campioni ufficiali

2.1 Gestione dei campioni ufficiali

Si comunica che, in accordo con i relativi Laboratori Nazionali di riferimento, a partire da Lunedì 3 Giugno 2024 è attiva una nuova modalità di fruizione dei genomi dei ceppi di *L. monocytogenes*,*Campylobacter* e, dal 2025, *Brucella*, caricati nella piattaforma del "Centro di Referenza Nazionale per Sequenze Genomiche di microrganismi patogeni: banca dati e analisi di bioinformatica (**GENPAT**) sono disponibili all'indirizzo https://genpat.izs.it.

Tale nuova modalità consentirà, agli utenti autorizzati, di visualizzare metadati minimi e risultati di caratterizzazione genomica di tutti i ceppi di *L. monocytogenes*, *Campylobacter* e *Brucella* isolati, nel territorio nazionale, nell'ambito del controllo ufficiale e delle altre attività ufficiali, come definite dal Regolamento (UE) 2017/625 e dal D. Lgs. 2 febbraio 2021, n.27.

Sarà inoltre possibile esportare i profili allelici e le sequenze assemblate (*de novo assembly*), per l'eventuale utilizzo nell'ambito delle attività e scopi del Servizio Sanitario Nazionale.

2.1.1 Regole per l'accesso ai dati

La piattaforma GENPAT consente agli utenti con profilo "IZS Listeria" di visualizzare metadati minimi e risultati di caratterizzazione genomica delle seguenti sequenze:

- Sequenze di ceppi di *L. monocytogenes* prelevati dal 01/01/2021, nel territorio nazionale, nell'ambito del controllo ufficiale e delle altre attività ufficiali, come definite dal Regolamento (UE) 2017/625 e dal D. Lgs. 2 febbraio 2021, n. 27;
- Valori di qualità delle sequenze conformi alla SOP accreditata `IZS TE B2.1.9 SOP032` .
- Valori di qualità delle chiamate alleliche conformi alla SOP accreditata `IZS TE B912.1 SOP001` .

La piattaforma consente inoltre agli utenti con profilo "IZS Campylobacter" di visualizzare metadati minimi e risultati di caratterizzazione genomica delle seguenti sequenze:

- Sequenze di ceppi di *Campylobacter* prelevati dal 01/01/2021, nel territorio nazionale, nell'ambito del controllo ufficiale e delle altre attività ufficiali, come definite dal Regolamento (UE) 2017/625 e dal D. Lgs. 2 febbraio 2021, n. 27.
- Valori di qualità delle sequenze conformi alla SOP accreditata `IZS TE B2.1.9 SOP032` .
- Valori di qualità delle chiamate alleliche conformi alla SOP accreditata `IZS TE B3.1.3 SOP071` .

La piattaforma consente inoltre agli utenti con profilo "IZS Brucella" di visualizzare metadati minimi e risultati di caratterizzazione genomica delle seguenti sequenze:

- Sequenze di ceppi di *Brucella* prelevati dal 01/01/2023, nel territorio nazionale, nell'ambito del controllo ufficiale e delle altre attività ufficiali, come definite dal Regolamento (UE) 2017/625 e dal D. Lgs. 2 febbraio 2021, n.27.
- Valori di qualità delle sequenze conformi alla SOP accreditata `IZS TE B2.1.9 SOP032` .
- Valori di qualità delle chiamate alleliche conformi alla SOP accreditata `IZS TE B3.1.3 SOP076` .

2.1.2 Qualità dei dati

Per ogni campione immesso in piattaforma, la pipeline NgsManager riporta nella tabella dei Campioni i dati di qualità (Sezione *Controllo qualità*) e vengono verificati i valori di qualità in accordo con le SOP sopra menzionate. Di seguito vengono riepilogati i valori principali.

NGS (`IZS TE B2.1.9 SOP032` **)**

Parametro	Valore accettabile
Numero di reads totali *Campylobacter sp.*	≥ 320.000
Numero di reads totali *L. monocytogenes*	≥ 580.000
Q-score	≥ 30

***L. monocytogenes* (** `IZS TE B912.1 SOP001` **)**

Parametro	Valore accettabile
Lunghezza del genoma *L. monocytogenes*	± 5%
Sequence Type e CC	100%
Numero di loci identificati cgMLST	≥ 95% (1660)

***Campylobacter* (** `IZS TE B3.1.3 SOP071` **)**

Parametro	Valore accettabile
Lunghezza del genoma *C. jejuni*	± 5%
Lunghezza del genoma *C. coli*	± 5%
Sequence Type e CC	100%
Numero di loci identificati cgMLST *C. jejuni*	≥ 95% (644)
Numero di loci identificati wgMLST *C. jejuni*	≥ 30% (838)
Numero di loci identificati cgMLST *C. coli*	≥ 95% (501)
Numero di loci identificati wgMLST *C. coli*	≥ 30% (743)

***Brucella* (** `IZS TE B3.1.3 SOP076` **)**

Parametro	Valore accettabile
Numero di loci identificati cgMLST *B.Melitensis*	≥ 95% (2455)
Numero di loci identificati wgMLST *Altre specie Brucella*	≥ 95% (1676)
Analisi SNP: coverage orizzontale	≥ 97%
Analisi SNP: coverage verticale	≥ 30

Inoltre, viene identificata la specie calcolata a partire dai dati genomici (campo `Specie Calcolata` della tabella *Campioni*). Tale valore deve essere concorde con il valore della specie dichiarata (campo `Specie`).

La pipeline automatica verifica i valori di qualità e la concordanza tra "specie" dichiarata e "specie calcolata"; successivamente, a seconda della correttezza dei valori, la pipeline imposta, nella colonna `Controllo` della tabella "Campioni", il risultato `OK` nel caso vengano soddisfatti tutti i valori, `Fallito` o `Attenzione` , con specifica indicazione dei valori non soddisfatti.

2.1.3 Profili IZS e LAB

Ad ogni utente associato ad una organizzazione (`ORG`) e abilitato alla gestione dei campioni ufficiali, vengono assegnati due profili della piattaforma GENPAT:

- **IZS**: per la gestione dei campioni ufficiali;
- **LAB**: per la gestione di campioni, sia ufficiali che non, appartenenti all'organizzazione.

Nella tabella seguente sono riepilogati

	Profilo IZS	**Profilo LAB**
Accesso ai campioni	Non ha associata alcuna organizzazione per i filtri sui campioni. Ciò consente di accedere a tutti i campioni ufficiali d'Italia (come definito precedentemente nella sezione "Regole per l'accesso ai dati"), prelevati a partire dal 2021 e che abbiano il campo `Controllo=OK` .	Ha associato `ORG` come organizzazione di appartenenza per i filtri sui campioni. Ciò consente di accedere a tutti i campioni inviati da `ORG` , indipendentemente dalla tipologia e dalla qualità dei dati.
Accesso ai metadati	Ha accesso limitato ai metadati dei campioni ufficiali di tutta Italia. Ha inoltre accesso limitato al livello di dettaglio dei metadati stessi. Ad esempio, è possibile conoscere il comune di prelievo ma non il punto esatto.	Ha accesso completo ai metadati (e al livello di dettaglio) dei campioni inviati da `ORG` .
Download	Consente di scaricare i profili allelici e l'assembly *de novo* dei campioni ufficiali italiani.	Consente di scaricare qualsiasi risultato di analisi disponibile su GENPAT sui campioni appartenenti a `ORG` .
Upload	Ha associato `ORG` come organizzazione di appartenenza per gli upload. Ciò consente di inviare sequenze di campioni appartenenti a `ORG` .	Ha associato `ORG` come organizzazione di appartenenza per gli upload.
Analisi	Consente di lanciare analisi di confronto fra i campioni ufficiali italiani.	Consente di lanciare qualsiasi analisi disponibile su GENPAT sui campioni appartenenti a `ORG` .

N.B.: Come profilo **IZS**, è possibile vedere, nella tabella *Campioni con anomalie*, i campioni inviati da `ORG` che hanno, nella colonna `Controllo` , un risultato diverso da `OK` . Tali campioni non possono essere utilizzati nel lancio di analisi. Si tiene a precisare che, nella tabella *Campioni con Anomalie*, non sono visibili i campioni anomali appartenenti ad altre organizzazioni.

2.1.4 Upload di campioni ufficiali

Dati di sequenza NGS (dati grezzi)

Relativamente ad isolati di *L. monocytogenes* e *Campylobacter* prelevati da campionamento ufficiale, la comunicazione di invio del ceppo, o dei dati NGS (dati grezzi) della sequenza, va effettuata tramite l'applicativo SEAP, analogamente a quanto fatto fino ad oggi.

Una procedura automatica assicura che i dati di sequenza e i matadati inviati al SEAP siano disponibili nella piattaforma GENPAT.

Il SEAP sarà integrato a breve all'interno della piattaforma GENPAT. Verrà data adeguata comunicazione non appena tale integrazione sarà completata.

Dati di assemblaggio

Inoltre, per i laboratori ufficiali che vogliano inviare i file di assemblati *de novo* direttamente nella piattaforma GENPAT, è disponibile una pagina di upload dedicata.

N.B.: Tale funzionalità è disponibile previa verifica di compatibilità della pipeline bioinformatica utilizzata. A tal proposito contattare il CRN GENPAT.

2.1.5 Analisi di bioinformatica

Analisi automatiche

Tutti i campioni inviati alla piattaforma GENPAT vengono processati automaticamente tramite la pipeline NgsManager. In particolare, per la produzione di assembly *de novo* e di profili allelici, tale pipeline è compatibile con la OneHealth analytical pipeline messa a disposizione da EFSA.

Analisi manuali

Le analisi disponibili sui campioni ufficiali sono:

- Grapetree: costruisce grafici ad albero usando i 2 algoritmi "Minimum Spanning Tree" (MSTree) e "Neighbor Joining" (NJ) a partire dai profili allelici dei campioni forniti, così come identificati nell'analisi cgMLST.
- ReporTree: analogamente a *GrapeTree* crea grafici di distanza tra microrganismi utilizzando i profili allelici dell'analisi cgMLST. L'albero viene generato con l'algoritmo "Minimum Spanning Tree" (MSTreeV2).
- ReporTree (SNP): questa analisi è disponibile solo per *Brucella* e genera grafici di distanza tra microrganismi a partire da un allineamento di SNP prodotto con il tool Snippy, derivato da mapping eseguiti sullo stesso riferimento genomico (GCF_000740415.1). L'albero viene costruito utilizzando l'algoritmo Minimum Spanning Tree (MSTreeV2).
- Surveillance *Listeria monocytogenes*: questa analisi è, al momento, disponibile solo per *Listeria monocytogenes*.

2.1.6 Download

Con il profilo IZS è possibile scaricare i profili allelici e l'assembly *de novo*. Per effettuare lo scarico dei files è obbligatorio compilare una nota indicando il motivo e l'utilizzo di tali dati.

2.1.7 Supporto

Per richieste di supporto utilizzare le funzioni di **Help desk** accessibili tramite la relativa voce presente nel menu contestuale del Wiki posto in alto a destra della piattaforma. Qui la pagina di dettaglio con video esplicativo.

2.2 Submission a EFSA

Profili autorizzati possono utilizzare la piattaforma per gestire la **selezione** e l'**approvazione** dei campioni da inviare ad **EFSA**.

Il processo si svolge in diverse fasi e coinvolge due gruppi principali:

- Chi invia ad EFSA (Laboratori di referenza)
- Chi approva (EFSA Country Officer)

2.2.1 Processo di invio ad EFSA (per i Laboratori di referenza)

1. Approvazione

Se hai i permessi necessari troverai la voce **Invia per approvazione** nel menu contestuale della pagina **Campioni**.

Tramite questo pulsante puoi inviare in approvazione al tuo **EFSA Country Officer** una selezione di campioni, il quale potrà approvare o rifiutare l'invio ad EFSA.

I campioni in stato di **approvazione**, **approvati per invio** e/o **non approvati** sono elencati nelle relative sezioni, raccolte sotto la voce **Submission EFSA** nel menu di navigazione laterale.

Solo i campioni **approvati per l'invio** potranno essere effettivamente trasmessi ad EFSA.

Video dimostrativo sull'invio per approvazione.

2. Invio ad EFSA

All'interno della pagina **2 - Approvati per invio**, sotto la voce **Submission EFSA** del menu di navigazione laterale, seleziona i campioni da inviare e clicca sul pulsante **Invia ad EFSA**.

Si aprirà una modale che richiede l'inserimento di un **token di autorizzazione** fornito da EFSA.

A seguito dell'invio, viene generata una scheda che consente di monitorare lo stato della submission e visualizzare eventuali report relativi all'invio.

Le submission completate sono raccolte all'interno della pagina **4 - Inviate ad EFSA**, sempre sotto la voce **Submission EFSA** nel menu di navigazione laterale.

Procedura di submission ad EFSA.

2.2.2 Processo di approvazione (EFSA Country Officer)

All'accesso in piattaforma come **EFSA Country Officer**, riceverai una notifica che segnala la presenza di campioni da approvare per l'invio ad EFSA.

Seguendo il link nella notifica, accederai alla sezione **1 - Da approvare**, sotto la voce **Submission EFSA** del menu di navigazione laterale.

In questa sezione troverai la lista dei campioni in attesa di approvazione. Selezionando i campioni e cliccando su **Modifica stato submission**, potrai **approvare** o **non approvare** la richiesta.

Di seguito un breve video su come modificare lo stato di una submission e visualizzare lo stato delle approvazioni e delle submission in piattaforma.

Approvazione e visualizzazione stato campioni.

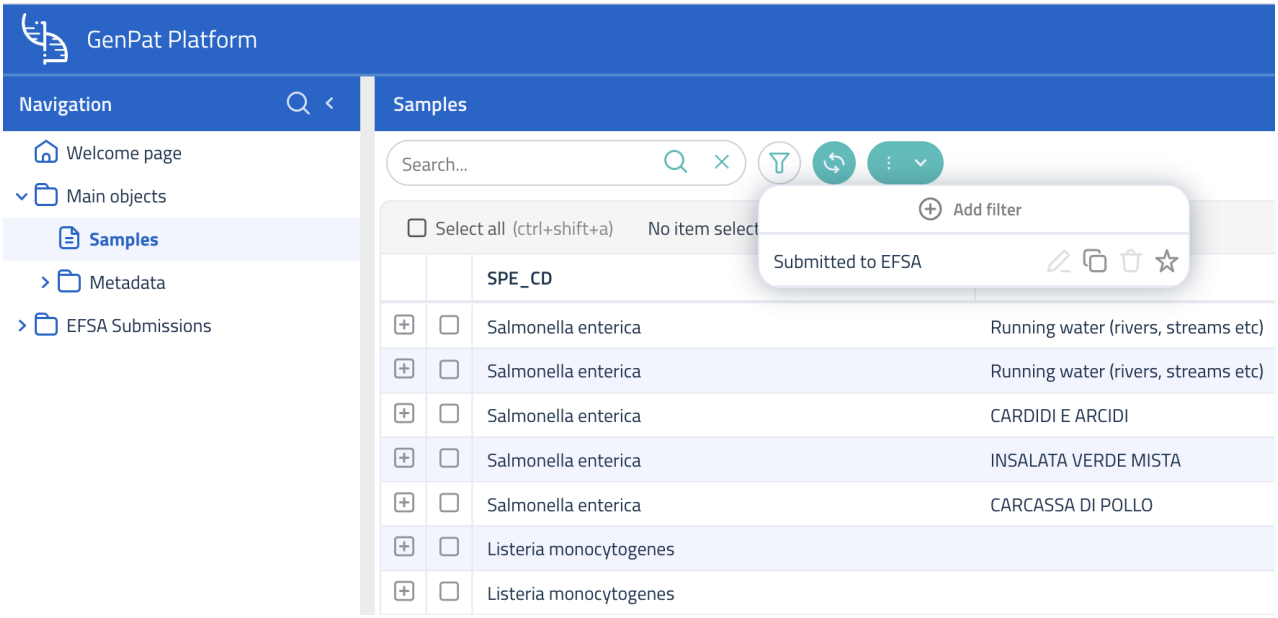
Dettaglio campioni

All'interno delle viste disponibili nel menu laterale sotto la voce **EFSA Submission** sono presenti informazioni riassuntive sui campioni, quali specie, stato del processo di invio, informazioni temporali.

Per maggiori dettagli sui campioni e sui metadati associati, puoi consultare la sezione **Campioni** disponibile sotto **Elementi principali** nel menu di navigazione sulla sinistra.

In questa pagina troverai un elenco completo di tutti i campioni (inviati ad EFSA, in approvazione, non approvati o approvati) con tutti i metadati collegati. Qui potrai applicare filtri base o avanzati per visualizzare dettagli specifici o eseguire esportazioni mirate. Per guide dettagliate su queste funzionalità, consulta la sezione Caratteristiche delle tabelle del wiki.

Nella la sezione **Campioni** è presente un filtro avanzato predefinito **Inviate ad EFSA** che filtra la tabella mostrando appunto tutti i campioni inviati ad EFSA. Questo filtro è accesibile cliccando sull'icona a forma di imbuto posizionata immediatamente sopra la tabella.



Visualizzazione filtro

2.2.3 Confronto tra campioni presenti su WGS Portal e campioni presenti in piattaforma

Questa sezione spiega come confrontare i campioni disponibili su **WGS Portal** (Interfaccia WEB del One-Health WGS System di EFSA) con quelli registrati in **piattaforma**, evidenziando le differenti tipologie di invio gestite dai due sistemi.

> **Nota:** applica sempre il filtro **Show privates** su WGS Portal prima di iniziare il confronto: senza questo filtro vedresti anche i campioni pubblici e lo scostamento risulterebbe ancora più marcato.

Differenze generali

Dopo aver abilitato **Show privates**, è normale riscontrare una discrepanza tra il numero di campioni mostrati su WGS Portal e quelli visibili sulla nostra piattaforma. Lo scostamento dipende dall'origine dei dati e dai flussi con cui i campioni sono stati caricati nel tempo.

Possibili discriminanti

Nel portale possono comparire campioni con codice `MIG-` e `EFSA-`:

- i campioni `MIG-` sono di record importati da EFSA nel sistema **WGS Portal** prima dell'attivazione del flusso di alimentazione. Questi non sono presenti in **piattaforma**;
- i campioni con prefisso `EFSA-` sono quelli inseriti nel **WGS Portal** dopo attivazione del flusso di alimentazione. Fra questi vi sono tipologie diverse di sequenze in base al tipo di alimentazione, come riportato in basso.

Attualmente riconosciamo quattro modalità con cui i campioni sono arrivati a EFSA:

1. Migrazione EFSA

- Inserimenti manuali effettuati da EFSA alla prima attivazione del WGS Portal.
- Identificativo tipico: `MIG-`.
- Disponibili solo su WGS Portal (campo *Code* = `MIG-***`).

2. Invii pre-piattaforma

- Submission manuali dei data provider prima dell'introduzione del flusso gestito tramite la nostra piattaforma.
- Identificativo tipico: `EFSA-`.
- Visibili su WGS Portal; non sempre presenti o allineati in GenPat.

3. EFSA Submission tramite piattaforma

- Flusso ufficiale e più recente, gestito tramite la nostra piattaforma.
- Identificativo tipico: `EFSA-`.
- La lista completa è disponibile in piattaforma all'interno della pagina **Inviate ad EFSA**; i corrispondenti campioni sono presenti anche su WGS Portal.

4. Inserimenti diretti EFSA

- EFSA carica direttamente i campioni sul WGS Portal, senza passare da GenPat.
- Identificativo tipico: `EFSA-`.
- Sul WGS Portal il campo **Submitter** è impostato a `EFSA`.

Tabella riassuntiva

Tipo di invio	Descrizione	Code	Origine dei dati	Dove trovarli
Migrazione EFSA	Record caricati da EFSA all'attivazione del WGS Portal	`MIG-`	EFSA → WGS Portal	WGS Portal (*Code* = *MIG-*)
Pre-piattaforma	Invii manuali dei data provider precedenti a GenPat	`EFSA-`	Data provider → EFSA	WGS Portal
Inserimenti diretti EFSA	Campioni caricati da EFSA sul WGS Portal	`EFSA-`	EFSA → WGS Portal	WGS Portal (*Submitter* = *EFSA*)
EFSA Submission tramite GenPat	Flusso gestito dalla piattaforma GenPat	`EFSA-`	Data provider → Piattaforma → EFSA	GenPat (**Inviate ad EFSA**) + WGS Portal

Importante: i campioni inviati dai *data provider* italiani verso EFSA hanno un identificativo che inizia con il prefisso `EFSA-`.

In sintesi

- Controlla sempre il campo **Submitter** su WGS Portal per capire chi ha caricato il campione.
- I campioni con `MIG-` provengono dalla migrazione iniziale di EFSA e non sono presenti in piattaforma.
- Per confrontare correttamente i dati, considera la data di introduzione dei diversi flussi e la modalità di invio associata a ciascun codice.

3. Elementi principali

3.1 Campioni

La pagina **Campioni** è una delle sezioni principali del sistema ed è accessibile dal menù di **Navigazione** laterale, sotto la voce **Elementi principali**. Contiene una tabella con i campioni disponibili in piattaforma e i relativi metadati, visibili in base al profilo di appartenenza e ad eventuali filtri attivi, come tag o carrello.

> **Suggerimento**
> Nella sezione dedicata alle funzionalità delle tabelle sono disponibili indicazioni utili su come interagire con questa pagina.

3.1.1 Schede Campioni

Facendo doppio clic su una riga della tabella (o, se presente, tramite click sull'icona [+]) si accede alla **scheda dettagliata** del campione: - nella parte superiore sono riportati i **metadati principali** e le informazioni derivate dalle analisi eseguite sul campione;
- nella parte inferiore sono elencate le **relazioni** del campione con altri database, eventuali alias e l'elenco dei **tag** associati.

Alcune sezioni includono un link di apertura rapida, indicato da un'icona accanto al titolo della sezione ( ).

Il video che segue mostra brevemente le operazioni che l'utente può eseguire sui campioni.

> **Nota:** Per motivi di privacy e sicurezza, i video sono stati realizzati nell'applicativo demo Cohesive. L'interfaccia potrebbe quindi presentare alcune differenze rispetto all'ambiente reale.

Caratteristiche della tabella Campioni

3.1.2 Funzioni della tabella dedicate ai campioni

Come per gli altri elementi principali della piattaforma visualizzati in formato tabellare (Alias, Metadati, Risultati analisi), anche nella tabella **Campioni** la barra multifunzione superiore consente di accedere a diverse opzioni come:

- strumenti di **filtro**,
- strumenti di **esportazione**,
- operazioni dirette sui campioni.

Il funzionamento generale è descritto nella sezione dedicata alla gestione delle tabelle; nello specifico, la tabella dei campioni presenta alcune peculiarità come l'utilizzo del **Carrello** o alcune voci specifiche disponibili, in base al profilo, all'interno del menù contestuale .

Carrello

Il carrello è una funzione disponibile esclusivamente per i campioni e rappresenta il passaggio preliminare per poter interagire con essi. Tramite il carrello è infatti possibile:

- lanciare analisi,
- lanciare download,
- aggiungere i campioni a un tag esistente oppure crearne di nuovi.

Per aggiungere uno o più **campioni selezionati** al carrello è sufficiente usare l'apposito pulsante . La freccia sul lato destro del pulsante apre un menù che consente di scegliere tra tre azioni:

- **Aggiungi al carrello**,
- **Sostituisci carrello**,
- **Rimuovi da carrello**.

Nota importante:
L'aggiunta di campioni al carrello richiede sempre la **selezione preventiva** dei campioni di interesse. La selezione multipla è attiva di default, ma può essere disattivata o riattivata tramite il menù contestuale o la scorciatoia da tastiera `Ctrl+Shift+M` .
Non è possibile aggiungere campioni al carrello senza averli selezionati.

La guida completa al carrello e alle sue funzioni è disponibile nella pagina dedicata.

3.1.3 Campioni pubblici e campioni privati

All'interno della tabella Campioni sono disponibili due tipologie di dati: *Campioni pubblici* e *Campioni privati*.

- I **campioni pubblici** corrispondono a sequenze depositate nei database di NCBI (ad es. *Assembly*, *Nucleotide*, *SRA* ecc.) e possono essere importati nella piattaforma tramite la procedura di Upload da NCBI.
- I **campioni privati** invece sono sequenze non presenti in database pubblici, caricate nella piattaforma o attraverso procedure automatizzate successive al sequenziamento, oppure mediante le modalità di Upload manuale.

Per praticità sono disponibili due filtri preimpostati che consentono di visualizzare **solo i campioni pubblici** oppure **solo i campioni privati**. È inoltre possibile salvare uno dei due filtri come predefinito, così da trovarlo già attivo ad ogni accesso, in base alle tue preferenze. Di seguito troverai un video dimostrativo sull'utilizzo dei filtri.

Utilizzo filtri campioni pubblici e privati
Puoi trovare ulteriori informazioni sull'utilizzo dei filtri di base ed dei filtri avanzati nella sezione di questo wiki dedicata alle funzionalità delle tabelle.

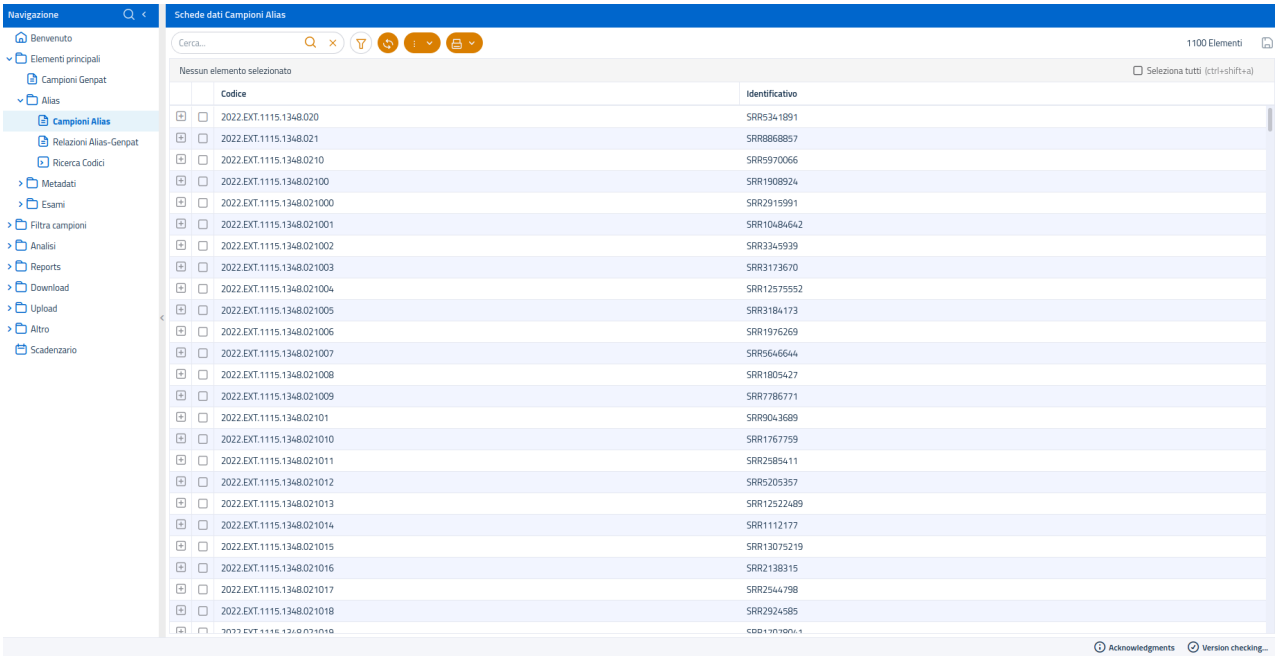
3.2 Alias

Nella sezione **Elementi principali** del menù di **Navigazione** laterale, è presente l'insieme delle funzionalità **Alias**. Tali funzioni riguardano le relazioni e le corrispondenze tra codici che si riferiscono agli stessi campioni, come i codici dei database SEAP e SILAB o i codici originali di campioni esterni caricati in piattaforma (ad esempio, da database ENA o NCBI).

3.2.1 Interfaccia Alias dei campioni

Nella pagina **Alias dei campioni** vengono riportati, in tabella, i codici dei campioni presenti in piattaforma, insieme ai corrispondenti codici originali prima dell'import (colonna **Identificativo**).

Analogamente alla pagina **Campioni**, anche per gli Alias le *entries* sono organizzate tramite un sistema a tabelle; pertanto, l'utente ha a disposizione, per gli Alias, tutte le funzionalità e le modalità di interazione che può utilizzare con i campioni.

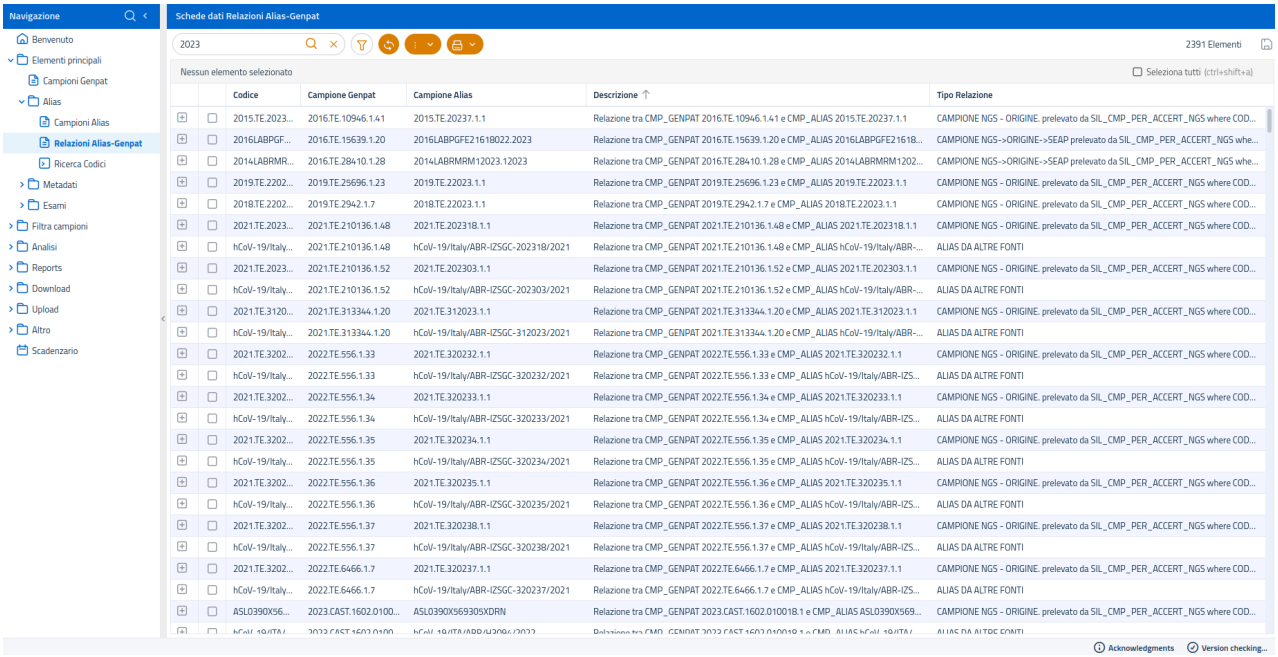


Per quanto riguarda la tabella degli Alias, le colonne che è possibile visualizzare sono di numero inferiore rispetto a quelle consultabili dalla tabella *Campioni*, in quanto esse comprendono solo i codici del campione (Codice Genpat e Identificativo originale) e i metadati ad esso associati.

3.2.2 Elenco degli alias per campione

La pagina riporta, all'interno della tabella, i codici delle *entries* in Genpat e le rispettive relazioni con i codici Alias nel database, accompagnate da un campo di descrizione della relazione. La tabella dell'**Elenco degli alias per campione** è concepita in modo da permettere agli utenti di associare il codice Genpat al codice originale (e *vice versa*), di risalire all'origine del campione e alla sua importazione in Genpat dal database esterno.

Anche in questo caso, tutte le funzioni delle tabelle sono disponibili all'utente.


Navigazione
Benvenuto
Elementi principali
Campioni Genpat
Alias
Campioni Alias
Relazioni Alias-Genpat
Ricerca Codici
Metadati
Esami
Filtra campioni
Analisi
Reports
Download
Upload
Altro
Scadenzario
Schede dati Relazioni Alias-Genpat
2023
2391 Elementi
Nessun elemento selezionato
Seleziona tutti (ctrl+shift+a)
Codice
Campione Genpat
Campione Alias
Descrizione
Tipo Relazione
2015.TE.2023...
2016.TE.10946.1.41
2015.TE.20237.1.1
Relazione tra CMP_GENPAT 2016.TE.10946.1.41 e CMP_ALIAS 2015.TE.20237.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
2016LABPGF...
2016.TE.15639.1.20
2016LABPGFE21618022.2023
Relazione tra CMP_GENPAT 2016.TE.15639.1.20 e CMP_ALIAS 2016LABPGFE21618...
CAMPIONE NGS->ORIGINE->SEAP prelevato da SIL_CMP_PER_ACCERT_NGS whe...
2014LABRMR...
2016.TE.28410.1.28
2014LABRMRM12023.12023
Relazione tra CMP_GENPAT 2016.TE.28410.1.28 e CMP_ALIAS 2014LABRMRM1202...
CAMPIONE NGS->ORIGINE->SEAP prelevato da SIL_CMP_PER_ACCERT_NGS whe...
2019.TE.2202...
2019.TE.25696.1.23
2019.TE.22023.1.1
Relazione tra CMP_GENPAT 2019.TE.25696.1.23 e CMP_ALIAS 2019.TE.22023.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
2018.TE.2202...
2019.TE.2942.1.7
2018.TE.22023.1.1
Relazione tra CMP_GENPAT 2019.TE.2942.1.7 e CMP_ALIAS 2018.TE.22023.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
2021.TE.2023...
2021.TE.210136.1.48
2021.TE.202318.1.1
Relazione tra CMP_GENPAT 2021.TE.210136.1.48 e CMP_ALIAS 2021.TE.202318.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
hCoV-19/Italy...
2021.TE.210136.1.48
hCoV-19/Italy/ABR-IZSGC-202318/2021
Relazione tra CMP_GENPAT 2021.TE.210136.1.48 e CMP_ALIAS hCoV-19/Italy/ABR-...
ALIAS DA ALTRE FONTI
2021.TE.2023...
2021.TE.210136.1.52
2021.TE.202303.1.1
Relazione tra CMP_GENPAT 2021.TE.210136.1.52 e CMP_ALIAS 2021.TE.202303.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
hCoV-19/Italy...
2021.TE.210136.1.52
hCoV-19/Italy/ABR-IZSGC-202303/2021
Relazione tra CMP_GENPAT 2021.TE.210136.1.52 e CMP_ALIAS hCoV-19/Italy/ABR-...
ALIAS DA ALTRE FONTI
2021.TE.3120...
2021.TE.313344.1.20
2021.TE.312023.1.1
Relazione tra CMP_GENPAT 2021.TE.313344.1.20 e CMP_ALIAS 2021.TE.312023.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
hCoV-19/Italy...
2021.TE.313344.1.20
hCoV-19/Italy/ABR-IZSGC-312023/2021
Relazione tra CMP_GENPAT 2021.TE.313344.1.20 e CMP_ALIAS hCoV-19/Italy/ABR-...
ALIAS DA ALTRE FONTI
2021.TE.3202...
2022.TE.556.1.33
2021.TE.320232.1.1
Relazione tra CMP_GENPAT 2022.TE.556.1.33 e CMP_ALIAS 2021.TE.320232.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
hCoV-19/Italy...
2022.TE.556.1.33
hCoV-19/Italy/ABR-IZSGC-320232/2021
Relazione tra CMP_GENPAT 2022.TE.556.1.33 e CMP_ALIAS hCoV-19/Italy/ABR-IZS...
ALIAS DA ALTRE FONTI
2021.TE.3202...
2022.TE.556.1.34
2021.TE.320233.1.1
Relazione tra CMP_GENPAT 2022.TE.556.1.34 e CMP_ALIAS 2021.TE.320233.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
hCoV-19/Italy...
2022.TE.556.1.34
hCoV-19/Italy/ABR-IZSGC-320233/2021
Relazione tra CMP_GENPAT 2022.TE.556.1.34 e CMP_ALIAS hCoV-19/Italy/ABR-IZS...
ALIAS DA ALTRE FONTI
2021.TE.3202...
2022.TE.556.1.35
2021.TE.320234.1.1
Relazione tra CMP_GENPAT 2022.TE.556.1.35 e CMP_ALIAS 2021.TE.320234.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
hCoV-19/Italy...
2022.TE.556.1.35
hCoV-19/Italy/ABR-IZSGC-320234/2021
Relazione tra CMP_GENPAT 2022.TE.556.1.35 e CMP_ALIAS hCoV-19/Italy/ABR-IZS...
ALIAS DA ALTRE FONTI
2021.TE.3202...
2022.TE.556.1.36
2021.TE.320235.1.1
Relazione tra CMP_GENPAT 2022.TE.556.1.36 e CMP_ALIAS 2021.TE.320235.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
hCoV-19/Italy...
2022.TE.556.1.36
hCoV-19/Italy/ABR-IZSGC-320235/2021
Relazione tra CMP_GENPAT 2022.TE.556.1.36 e CMP_ALIAS hCoV-19/Italy/ABR-IZS...
ALIAS DA ALTRE FONTI
2021.TE.3202...
2022.TE.556.1.37
2021.TE.320238.1.1
Relazione tra CMP_GENPAT 2022.TE.556.1.37 e CMP_ALIAS 2021.TE.320238.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
hCoV-19/Italy...
2022.TE.556.1.37
hCoV-19/Italy/ABR-IZSGC-320238/2021
Relazione tra CMP_GENPAT 2022.TE.556.1.37 e CMP_ALIAS hCoV-19/Italy/ABR-IZS...
ALIAS DA ALTRE FONTI
2021.TE.3202...
2022.TE.6466.1.7
2021.TE.320237.1.1
Relazione tra CMP_GENPAT 2022.TE.6466.1.7 e CMP_ALIAS 2021.TE.320237.1.1
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
hCoV-19/Italy...
2022.TE.6466.1.7
hCoV-19/Italy/ABR-IZSGC-320237/2021
Relazione tra CMP_GENPAT 2022.TE.6466.1.7 e CMP_ALIAS hCoV-19/Italy/ABR-IZS...
ALIAS DA ALTRE FONTI
ASL0390X56...
2023.CAST.1602.0100...
ASL0390X569305XDRN
Relazione tra CMP_GENPAT 2023.CAST.1602.010018.1 e CMP_ALIAS ASL0390X569...
CAMPIONE NGS - ORIGINE. prelevato da SIL_CMP_PER_ACCERT_NGS where COD...
Acknowledgments
Version checking...

3.3 Metadati

Nella sezione "Elementi principali" del menù di navigazione laterale, è presente l'insieme delle funzionalità relative ai **Metadati** (link pagina Wikipedia).

I metadati in piattaforma sono suddivisi nelle categorie:

- **Specie** : elenco delle specie degli organismi, insieme ai corrispondenti codici identificativi. La categoria è a sua volta suddivisa in:
  - **Specie Patogeno**
  - **Specie Host**
- **Materiali** : elenco dei materiali o delle matrici alimentari da cui è stato originariamente isolato il campione, insieme ai corrispondenti codici identificativi.
- **Punto Prelievo** : elenco dei luoghi in cui è stato originariamente prelevato il campione (nomi di enti, aziende, ospedali e/o coordinate di latitudine e longitudine), insieme ai corrispondenti codici identificativi.

I metadati sono fondamentali per l'upload di un nuovo campione in piattaforma.

I metadati possiedono una gerarchia, indicata nelle pagine "Albero Specie" e "Albero Materiali".

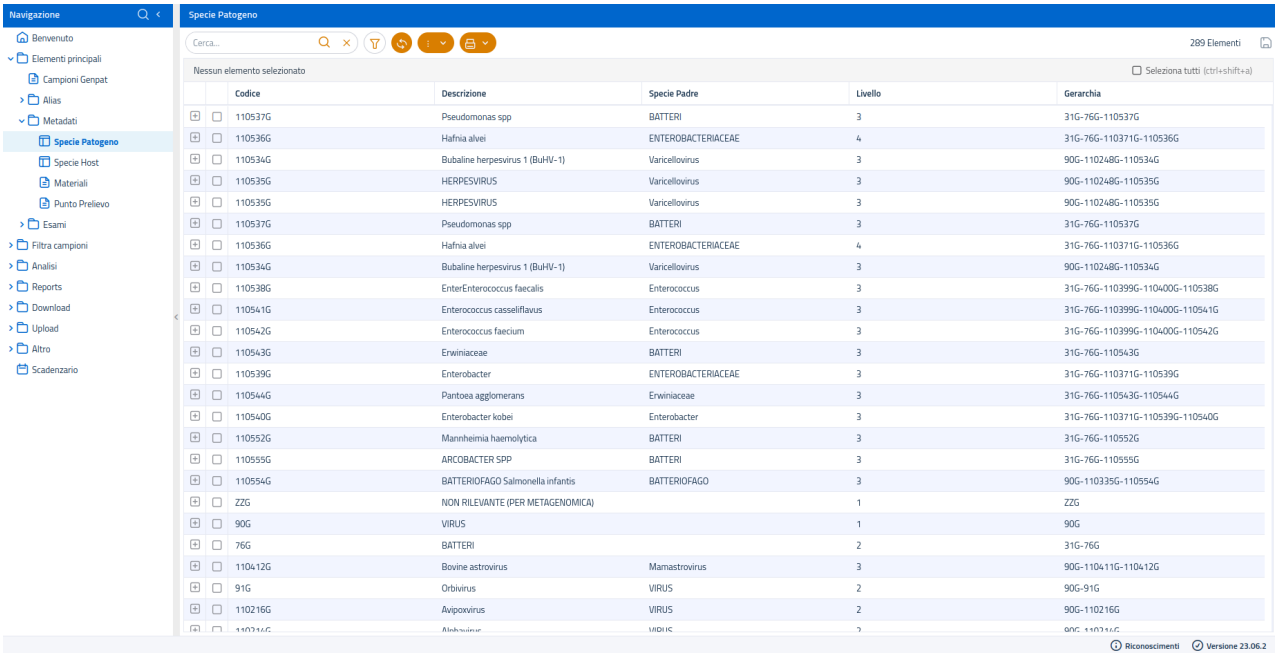
3.3.1 Codifiche secondo standard EFSA

Per garantire uniformità di codifiche agli utenti esterni, nella piattaforma i metadati dei campioni aderiscono alle specifiche EFSA FOODEX2.

Per una lista delle modifiche apportate ai database dei metadati si invita a consultare l'apposita pagina delle novità e versionamento.

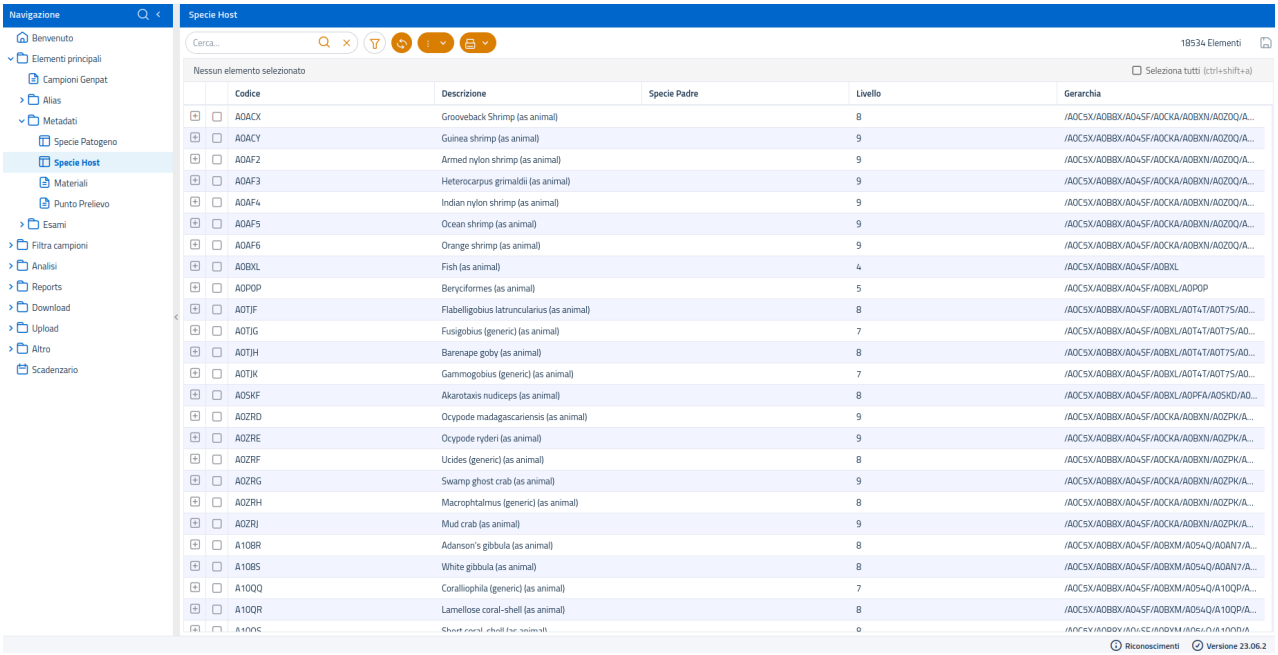
3.3.2 Specie Patogeno

La tabella della Specie Patogeno elenca le specie di patogeni e i corrispondenti codici con cui esse vengono identificate nella piattaforma. Le operazioni disponibili sono le stesse della tabella dei campioni.



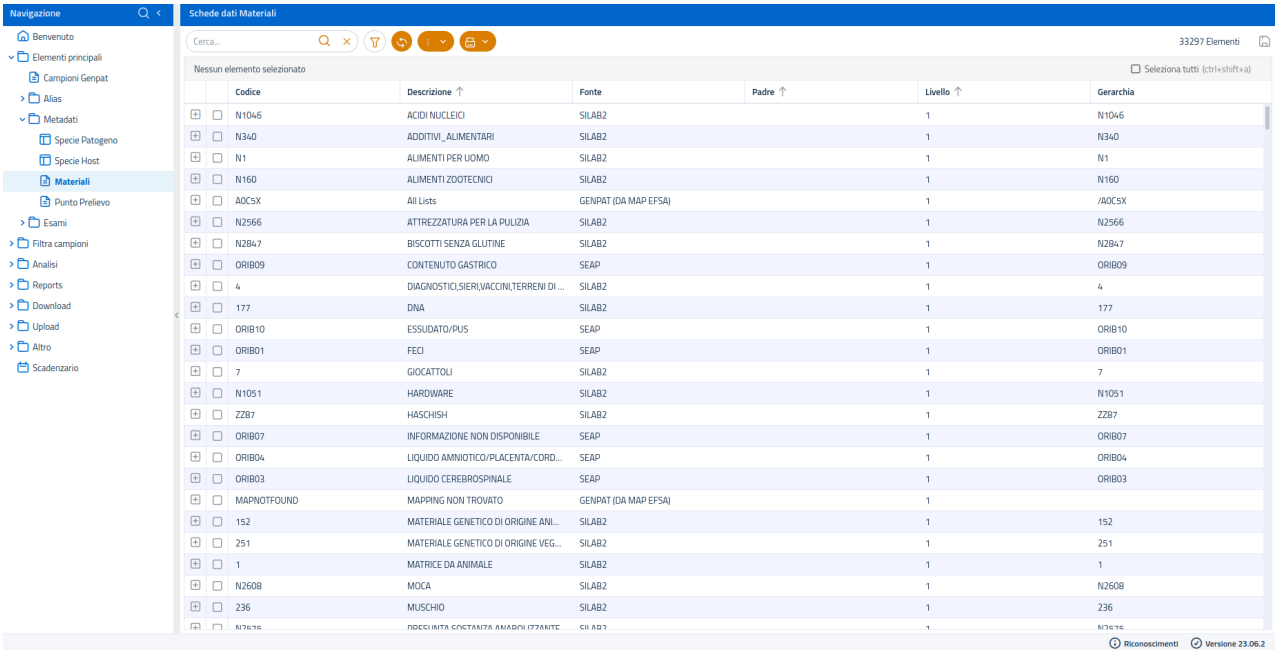
3.3.3 Specie Host

La tabella della Specie Host elenca le specie degli organismi ospiti ed i codici corrispondenti. La lista di organismi ospiti fa riferimento agli organismi i cui genomi reference siano stati aggiunti alla piattaforma. Le operazioni disponibili sono le stesse della tabella dei campioni.



3.3.4 Materiali

La tabella dei Materiali elenca matriali, superfici, matrici alimentari e strutture biologiche da cui sono stati isolati i campioni, insieme ai codici corrispondenti. Le operazioni disponibili sono le stesse della tabella dei campioni.



3.3.5 Punto Prelievo

La tabella del Punto Prelievo elenca i codici delle località in cui è avvenuto il primo prelievo o recupero del materiale da cui è stato isolato un campione. Il punto prelievo comprende provincia, comune, eventuale indirizzo e struttura o ente. Le operazioni di base disponibili sulla tabella sono le stesse di quella dei campioni ma per alcuni profili è possibile aggiungere punti prelivo da sorgenti esterne in autonomia.

Per profili autorizzati è possibile aggiungere punti prelivo da sorgenti esterne in autonomia, una volta selezionato, entrerà a far parte dei punti prelievo della piattaforma.

Aggiungi punto prelievo

Le sorgenti esterne alla piattaforma attualment disponibili sono:

- Aziende registrate in Banca Dati Nazionale (BDN)
- Stabilimenti Italiani riconosciuti o registrati

Per effettuare la ricerca in entrambe le sorgenti esterne, è necessario disporre di un codice (**codice azienda** per BDN e **Approval Number** per le altre strutture). Le funzioni di aggiunta del punto prelievo sono disponibili sia sulla pagina **Punto Prelievo** che in fase di preparazione e caricamento dei metadati. Di seguito dei video dimostrativi.

Pagina Punto Prelievo sampling-point.mp4

Pagina Preparazione Upload upload-add-sampling-point.mp4

3.4 Risultati analisi

Questa pagina è raggiungibile da **Elementi principali** del menù di navigazione laterale della piattaforma. In essa sono elencati, in visualizzazione tabellare, tutti i risultati delle analisi bioinformatiche della piattaforma già eseguite sui singoli campioni e che sono disponibili, come input, per l'eventuale lancio di nuove analisi. Le analisi di bioinformatica sono contrassegnate da *fonte BIOINFONAS*.

Nota: poter verificare la presenza di risultati di analisi, su un campione in piattaforma, è fondamentale per sapere se sono presenti tutti i file necessari per il lancio di analisi a valle; tali analisi, infatti, possono, a loro volta, richiedere come input i risultati ottenuti da analisi precedenti. Di seguito sono riportati due links alla documentazione del lancio analisi.

- Modalità di lancio analisi
- Introduzione ai tools

Le possibilità di interazione con la tabella e le funzioni disponibili sono le stesse della tabella dei campioni.

3.4.1 Segnala errore

Nel caso di un campione sottoposto più volte ad una medesima analisi, il sistema adopererà automaticamente i risultati importati più recenti come input per analisi successive che lo richiedano.

Nel caso i risultati di un'analisi non siano attendibili a causa di errori e per impedirne l'utilizzo come input per analisi successive, è disponibile, all'interno del menu contestuale della pagina, la voce **Segnala errore** che consente di marcare come errati gli elementi selezionati attraverso la tabella.

Importante! Quando si segnala un'analisi come errata, il sistema marcherà automaticamente come errori tutte le analisi presenti in piattaforma che hanno utilizzato i suoi risultati come input. **Esempio.** Se si segnala come errato un assembly ottenuto da 2AS_denovo__shovill, allora il sistema invaliderà automaticamente tutti gli eventuali esami di typing (mlst, cgMLST, ecc.) calcolati a partire da quell'assembly.

Il video che segue mostra brevemente come **segnalare un errore**:

Segnalare un errore

4. Filtra campioni

4.1 Motivi

4.1.1 Tag

Cos'è un tag

I tag sono una funzione della piattaforma che permette la suddivisione dei campioni in raccolte. La creazione di un tag è utile nel caso di campioni legati a progetti o focolai specifici. Se dei campioni appartengono ad un tag, fintanto che esso resterà attivo, solo i campioni ad esso associati potranno essere visualizzati nelle tabelle della piattaforma.

Per esempio, se sono stati analizzati dei campioni legati ad un ipotetico focolaio XXX, è possibile creare un tag "focolaio_XXX", in modo da associare tra loro i campioni che ne fanno parte. Questo permette di avere rapidamente a disposizione i campioni ogni volta che è necessario lanciare un'analisi, poichè si evita il processo di ricerca e selezione di ogni singolo campione prima dell'analisi.

Un tag corrispondente ad una raccolta di campioni può quindi essere attivato per visualizzare i campioni o i risultati ad essi correlati, ma anche per lanciare una nuova analisi o pipeline su tutti i campioni appartenenti al tag, specificando unicamente il tag desiderato, anzichè tutti i campioni che ne fanno parte, come input dell'analisi.

Il breve video seguente dimostra l'attivazione di un tag. Per informazioni approfondite sui tag, si faccia riferimento alla pagina Gestione Tag.

Nota: Per questioni di privacy e sicurezza dei dati, il video è stato realizzato nell'applicativo demo Cohesive, pertanto l'interfaccia può presentare alcune marginali differenze.

Motivi-Genpat.mp4

4.1.2 Gestione tag

In questa guida vengono illustrate, tramite una serie di videotutorial, le operazioni disponibili per i tag.

Nota: Per questioni di privacy e sicurezza dei dati, i video in questa pagina sono stati realizzati nell'applicativo demo Cohesive, pertanto l'interfaccia può presentare alcune marginali differenze.

Nella pagina **Gestione tag** si possono trovare informazioni riguardo il tag attivo (se ve ne è uno); da questa pagina è anche possibile disattivare il tag corrente o attivare un tag esistente.

Il video seguente mostra brevemente i contenuti della pagina: nella sezione **Gestione tag** sarà possibile trovare 4 elementi espandibili ("cards"), ogni *card* è dedicata ad una delle pricipali operazioni disponibili per i tag: **Attiva tag**, **Crea tag**, **Modifica tag** e **Rimuovi tag**.

Manage_tags.mp4

Attiva tag

La prima *card* dà accesso all'attivazione dei tag: cliccando nel campo di ricerca o sul pulsante "Seleziona tag" (a destra), sarà possibile scegliere e selezionare un tag tra quelli disponibili, elencati in una finestra pop-up. Una volta selezionato il tag, questo verrà attivato cliccando sul pulsante "Attiva tag".

Activate_tag.mp4

Più usati, Recenti, Preferiti

Le *cards* **Più usati**, **Recenti** e **Preferiti** permettono di selezionare rapidamente tag da attivare in base alle tue preferenze. Le *cards* **Più usati** e **Recenti** vengono riempite in modo automatico rispettivamente con una lista di **5 tag** da te più utilizzati e con una lista degli ultimi **5 tag** da te attivati.

La lista dei **Preferiti** invece dipende da tue dirette indicazioni e non ha limitazioni sul numero di tag selezionabili come preferiti. Ovviamente ti suggerisco di non inserirne troppi, sarebbe come non averne :). Nel seguente video una breve dimostrazione di come poter aggiungere e togliere tag dalla lista dei preferiti.

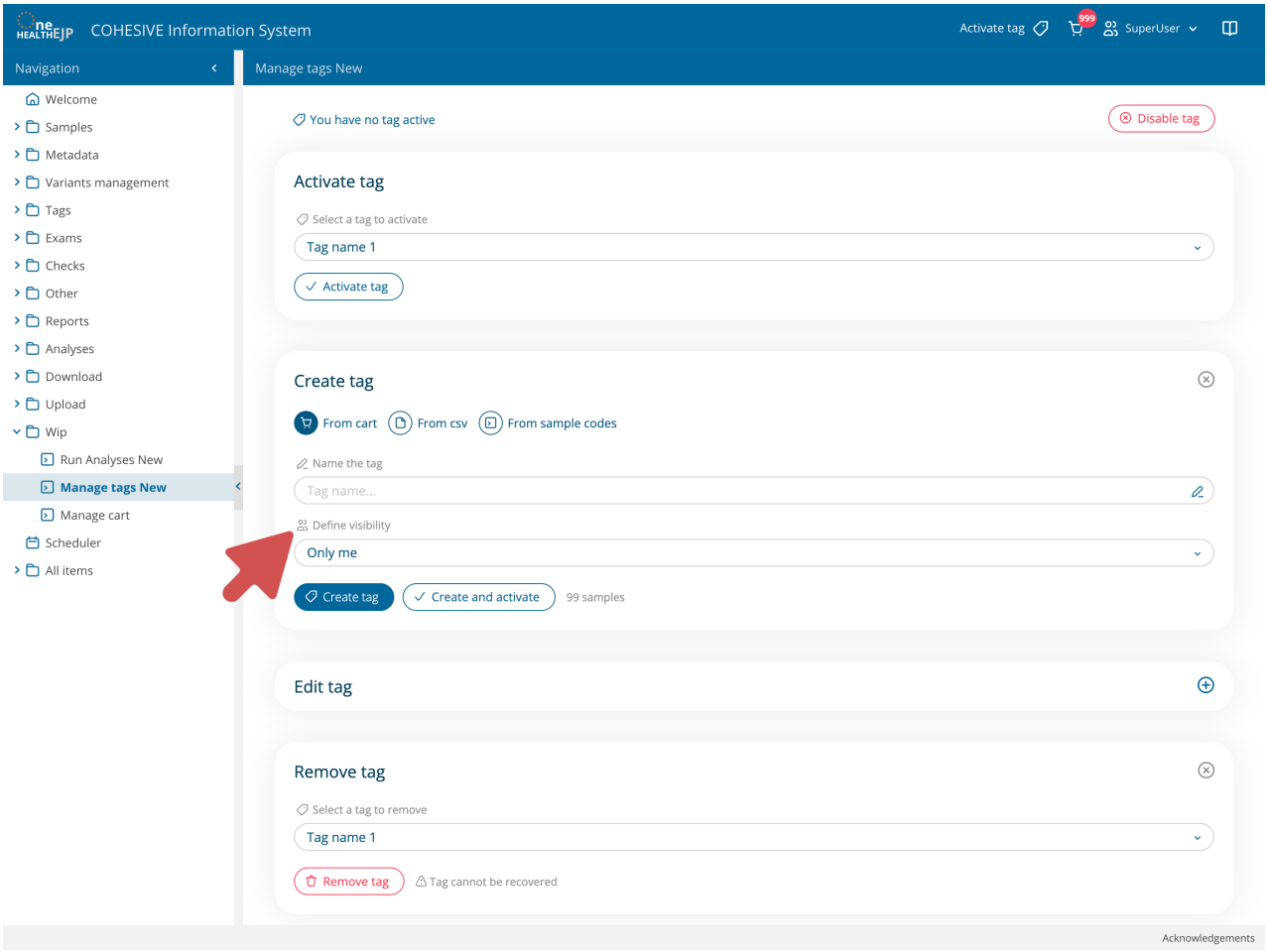
starred-tags.mp4

Crea tag

La *card* **Crea tag** offre modi diversi 3 tra cui scegliere per la creazione di un nuovo tag: è possibile usare una raccolta di campioni dal carrello, un file CSV o da codici campione (in quest'ultimo caso i codici dei campioni vengono scritti o incollati, in formato di lista separata da virgole, direttamente nel campo "Codici campioni"). La selezione di una delle opzioni provocherà l'aggiornamento della *card*, che mostrerà i campi specifici per l'opzione scelta.

Create_tag.mp4

Per tutte le opzioni e i metodi di creazione di un tag, sarà presente l'opzione **Visibilità**. Tale funzionalità permette di specificare se si desideri salvare il tag unicamente per la visualizzazione e l'uso personale, o se lo si voglia rendere disponibile anche per il gruppo di appartenenza dell'utente.



Modifica tag

La *card* **Modifica tag** permette di modificare un tag pre-esistente aggiungeno o rimuovendo campioni: esattamente come per **Crea tag**, anche qui è possibile scegliere tra carrello, CSV o lista di codici. Sarà anche possibile modificare la visibilità del tag selezionato e rimuovere da esso tutti i contenuti, senza però eliminare il tag.

Edit_tag.mp4

Rimuovi tag

L'ultima *card* (**Rimuovi tag**) ha lo scopo di permettere l'eliminazione di un tag a scelta. Come per la sezione **Attiva tag**, cliccando sul campo di ricerca o sul pulsante "Seleziona tag" (a destra) è possibile selezionare il tag desiderato; per eliminarlo, sarà necessario cliccare sul pulsante rosso "Rimuovi tag".

Remove_Tag.mp4

4.2 Carrello

4.2.1 Introduzione

Il carrello è una funzione estremamente flessibile, integrata nella piattaforma. Il funzionamento è simile al carrello tipico dei siti e-commerce ma possiede alcune funzionalità specifiche.

Il carrello ha lo scopo di conservare **temporaneamente** una selezione di campioni, i quali possono essere accumulati o rimossi al fine di essere utilizzati per diversi scopi: filtrare la piattaforma, lanciare analisi, fare operazioni di download, visualizzare report, creare tag.

Per informazioni dettagliate sulla funzione carrello, si può fare riferimento alla pagina sulla gestione del carrello. Anche per le operazioni di analisi, download e generazione di report rimandiamo alle relative aree della documentazione. Nella pagina corrente è invece descritto come aggiungere campioni al carrello e come attivarlo al fine di filtrare la piattaforma.

Aggiungere campioni al carrello

Il video qui in basso mostra brevemente come aggiungere **entries** al carrello.

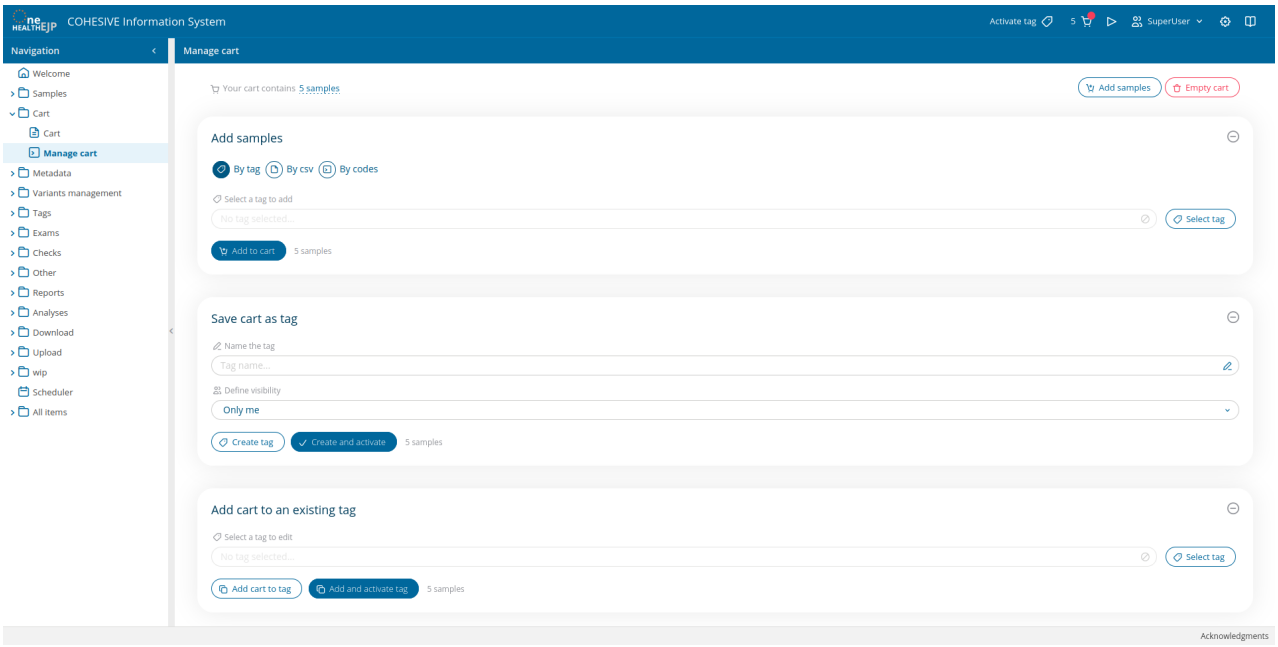
> **Nota:** Per questioni di privacy e sicurezza dei dati, il video è stato realizzato nell'applicativo demo Cohesive, pertanto l'interfaccia può presentare alcune differenze marginali.

Aggiunta di campioni al carrello

4.2.2 Gestione carrello

La funzione carrello è simile a quella di un sito di e-commerce; si può accedere al carrello dal menù di navigazione a sinistra (**Filtra campioni**), dove sono disponibili sia l'accesso al carrello che le funzioni di gestione del carrello. E' possibile accedere al carrello anche dall'apposita icona della barra delle funzioni, che funge da scorciatoia per la pagina **Gestisci carrello**.

La pagina **Gestisci carrello** presenta, in alto a destra, due pulsanti per azioni rapide: **Aggiungi campioni** (per aggiungere campioni manualmente dalla tabella dei campioni) e **Svuota carrello**.



Le tre funzioni principali del carrello sono raggruppate in tre *schede* espandibili: - **Aggiungi campioni al carrello**, - **Salva carrello come tag**, - **Aggiungi carrello ad un tag esistente**.

Aggiungi campioni al carrello

Da questa scheda è possibile modificare il contenuto del carrello, aggiungendovi campioni in tre modalità diverse: da un *tag* salvato, tramite un file *csv*, o da *codici* (scrivendo o incollando dei codici campione nel campo "*Codici campioni*"). La selezione di una di queste modalità provocherà l'aggiornamento della scheda, che mostrerà le opzioni e i campi disponibili per la modalità selezionata.

Il video che segue mostra brevemente i passaggi di tale fase.

> **Nota:** Per questioni di privacy e sicurezza dei dati, il video è stato realizzato nell'applicativo demo Cohesive; pertanto, l'interfaccia può presentare alcune marginali differenze.

Aggiungi campioni al carrello

Salva carrello come tag

Questa scheda permette all'utente di creare un nuovo tag direttamente dai campioni contenuti nel carrello. Per ulteriori informazioni sui tag, si faccia riferimento alla sezione dedicata; le nuove funzionalità disponibili per i tag sono invece descritte nella pagina Gestione tag.

Il breve video seguente mostra come salvare un carrello come tag.

Nota: Per questioni di privacy e sicurezza dei dati, il video è stato realizzato nell'applicativo demo Cohesive, pertanto l'interfaccia può presentare alcune marginali differenze.

Carrello come tag

Aggiungi carrello ad un tag esistente

L'ultima delle schede espandibili permette all'utente di aggiungere, ad un tag esistente, i campioni presenti nel carrello. Per ulteriori informazioni sui tag, si prega consultare la sezione dedicata o la pagina Gestione tag per una trattazione sulle loro funzionalità.

Il video seguente dimostra come aggiungere i campioni nel carrello al tag desiderato.

Nota: Per questioni di privacy e sicurezza dei dati, il video è stato realizzato nell'applicativo demo Cohesive, pertanto l'interfaccia può presentare alcune marginali differenze.

Aggiunta del carrello ad un tag esistente

4.2.3 Attivazione del carrello

Il carrello è funzionale allo stoccaggio temporaneo di campioni; esso, inoltre, attraverso la sua "attivazione", può essere abilitato alla funzione di filtro.

Nella pagina "**Gestisci carrello**" è presente il tasto di attivazione (in alto a destra); una volta premuto, i campioni presenti nel carrello saranno usati come termini per il filtraggio della piattaforma.

La funzione "**Attiva carrello**" consente, infatti, di filtrare la piattaforma tramite una selezione di campioni, permettendo di effettuare operazioni su di essi (come lanciare analisi, richiedere download o visualizzare report), senza la necessità di attivare un tag.

L'attivazione del carrello (in alternativa all'attivazione del tag), è necessaria per il lancio di un'analisi da carrello. Il carrello sarà automaticamente attivato se selezionato come fonte della raccolta di campioni per l'analisi.

> L'attivazione del carrello ha la precedenza su qualsiasi tag attivo e la sua disattivazione riporta la piattaforma allo stato precedente (per esempio, **se era attivo un tag, esso viene riattivato**).

L'azione di filtro operata dal carrello ha 2 differenze fondamentali rispetto al filtro di un tag:

- Il carrello è **transiente**: esso non viene salvato e può essere modificato in maniera semplice ed immediata aggiungendo o rimuovendo campioni.
- Nel caso di un carrello attivo, **la piattaforma** viene filtrata in base ai campioni presenti nel carrello; ciò significa che verranno filtrati, oltre alle liste di campioni, anche gli elementi presenti in **Visualizza analisi**, **Visualizza download** e **Visualizza importazioni**, rendendo più facile e rapido consultare le schede delle operazioni e delle analisi eseguite su specifiche *entries*, senza la necessità di cercarle o filtrarle a partire dall'intero elenco.

Il video che segue mostra come attivare o disattivare il carrello e spiega l'impatto di tali azioni sulla piattaforma.

> **Nota:** Per questioni di privacy e sicurezza dei dati, il video è stato realizzato nell'applicativo demo Cohesive; pertanto l'interfaccia può presentare alcune differenze marginali.

Attivazione del carrello in Cohesive

5. Lancia analisi

5.1 Introduzione

La pagina per il lancio di analisi della piattaforma può essere raggiunta dal menù di navigazione laterale (**Analisi > Lancia analisi**), tramite la scorciatoia sulla barra delle funzioni in alto a destra (tasto) o, direttamente, dalla **Pagina di benvenuto**.

5.1.1 Lancia analisi

Le funzionalità principali del sistema includono un flusso di lavoro semplificato per eseguire analisi sui campioni desiderati, l'interfaccia del quale è condivisa con la procedura di download.

Il procedimento per il *lancio di analisi* è guidato e suddiviso in 3 fasi: la fase **1** (**Campioni**) richiede di specificare una fonte da cui reperire la lista di campioni da analizzare, o a cui sono associati i file necessari per l'analisi; la fase **2** (**Tools**) porta alla schermata di selezione delle analisi; la fase **3** (**Input**) ha lo scopo di permettere la scelta della fonte dei file che verranno usati come input.

Le guide in dettaglio sulle 3 fasi del sistema di lancio analisi sono consultabili ai seguenti links:

1. Selezione dei campioni (**Campioni**);
2. Scelta del tipo di analisi (**Tools**);
3. Selezione dell'input (**Inputs**).

Il video che segue mostra tutte le fasi del procedimento di lancio analisi.

> **Nota:** Per questioni di privacy e sicurezza dei dati, il video presente nella pagina è stato realizzato nell'applicativo demo Cohesive, pertanto l'interfaccia può presentare alcune marginali differenze.

Fasi della procedura di lancio analisi

> **Nota 1:** il lancio di un'analisi da **carrello** ne causerà automaticamente l'attivazione. Per maggiori informazioni sull'attivazione del carrello, si consulti l'apposita sezione della guida attiva carrello.
>
> **Nota 2:** per il lancio di un'analisi è possibile anche utilizzare uno o più **RIS_CD**, tramite l'apposito pulsante in alto a destra. L'utilizzo del RIS_CD è utile per selezionare in maniera univoca i risultati di una specifica analisi, dando così facile accesso anche a risultati di analisi meno recenti.

Ulteriori informazioni sull'uso del RIS_CD per il Lancio Analisi e per le funzioni di Download sono disponibili nella sezione della guida "Esportazione del Codice Risultato (RIS_CD)".

5.1.2 Controllo analisi

Una volta terminata l'analisi, è possibile consultarne la scheda in "**Controllo analisi**", come mostrato nel video precedente. Nella pagina *Controllo analisi* sono elencate le *run* lanciate dall'utente, assieme ad informazioni utili come la tipologia di analisi, il tool utilizzato e lo stato del processo. Da qui sarà anche possibile ripetere l'analisi o importare l'analisi, se essa è terminata con successo. Per informazioni al riguardo, si consulti la pagina dedicata.

5.2 Fasi del sistema di lancio analisi

Il sistema per il lancio di analisi prevede una suddivisione logica in 3 fasi:

1. **Campioni** (selezione dei campioni);
2. **Tools** (scelta del tipo di analisi e degli strumenti di bionformatica);
3. **Inputs** (selezione dell'input);

Una guida rapida al lancio di analisi è disponibile nella pagina Wiki di introduzione a "Lancia analisi", in cui è presente anche un video dimostrativo dell'intero procedimento.

> **Nota:** Per questioni di privacy e sicurezza dei dati, alcuni dei video presenti nella pagina sono stati realizzati nell'applicativo demo Cohesive, pertanto l'interfaccia può presentare alcune marginali differenze.

In questa sezione del Wiki, vengono esposte le 3 fasi: *Campioni*, *Tools* e *Inputs*.

5.2.1 Fase 1: Campioni

Affinchè si possa procedere con la scelta dell'analisi, è prima necessario fornire al sistema una raccolta di codici. Tale fase è necessaria per preparare il lancio di un'analisi su file associati ad ognuno dei campioni specificati, (ad esempio, le *read grezze* dei campioni, nel caso del trimming oppure gli assembly fasta, nel caso di cgMLST con chewBBACA). L'analisi selezionata verrà eseguita per ogni codice specificato.

Le scelte a disposizione dell'utente per la selezione dei codici sono il **carrello** o un **tag**. Un tag è una raccolta statica, definita e salvata dall'utente, contenente codici di campioni. Il carrello consente una raccolta transiente di codici, definita dall'utente e che può essere abilitata, come funzione di filtro, tramite attivazione del carrello. Nel caso l'utente scelga di fornire i codici per il lancio analisi tramite carrello, esso viene automaticamente attivato, acquisendo priorità su qualsiasi eventuale tag attivo. La disattivazione del carrello riporta la piattaforma allo stato antecedente la sua attivazione.

Quando non vi sono campioni nel carrello o un tag attivo, è possibile cliccare sul link preposto alla selezione del tag (pulsante *Attiva tag*) o al riempimento del carrello (pulsante *Gestisci carrello*) dalle apposite pagine.

E' disponibile anche una funzionalità per il riempimento rapido del carrello tramite l'uso di una lista di codici o di un file CSV.

Il breve video seguente mostra l'esecuzione della fase di selezione dei codici tramite tag.

Selezione dei codici tramite tag

5.2.2 Fase 2: Tools

La fase denominata *Tools* mette a disposizione dell'utente un'interfaccia per la scelta dell'analisi (o della pipeline) da eseguire sui campioni precedentemente selezionati e, successivamente, dello stumento bioinformatico con il quale si desidera affettuare l'analisi. Le tipologie di analisi disponibili in piattaforma sono elencate come riquadri che indicano nome e tipologia di appartenenza. Esse, tuttavia, possono anche essere filtrate per categoria, visualizzate come schema ad albero o vagliate tramite l'apposito campo di ricerca.

A seguito della selezione dell'analisi, all'utente viene proposta una schermata di conferma che offre alcune informazioni utili sull'analisi scelta e sul programma bioinformatico selezionato. In questa pagina, l'utente potrà scegliere il tool (anche detto *metodo*) bioinformatico dal menù a tendina della sezione *Seleziona metodo*, se in piattaforma sono presenti due o più tools bioinformatici che possono eseguire la stessa analisi.

Il breve video che segue mostra la selezione di analisi e tool.

Selezione di analisi e tool

5.2.3 Fase 3: Input

L'ultima fase necessaria per terminare il lancio di un'analisi è quella di definizione degli input. Con la prima fase (*Campioni*) sono stati forniti i codici dei campioni di interesse ma rimane da specificare quali file usare, tra quelli associati ad ognuno di tali campioni, per l'analisi selezionata; ad esempio, un'analisi come il "trimming" usa, come input, le *read grezze* dei campioni mentre il clustering con GrapeTree necessita dei file tabulari da cgMLST. Nel primo caso verranno selezionati i file FASTQ delle *read* dei campioni inseriti nella fase 1 laddove, nel secondo caso, verranno selezionati i corrispondenti files dei profili allelici prodotti da chewBBACA.

> Questa fase richiede, dunque, che i file necessari per il lancio di un'analisi debbano essere in precedenza prodotti ed importati, (ciò significa che, ad esempio, non sarebbe possibile lanciare GrapeTree su campioni sui quali non sia stato eseguito prima chewBBACA).

I campi dell'interfaccia, di questa ultima fase, sono già precompilati con l'input standard per il tool bioinformatico scelto; l'intervento dell'utente, pertanto, sarà necessario solo nel caso in cui vi sia più di un input idoneo disponibile. Dalla stessa interfaccia, sarà possibile specificare alcune opzioni, se disponibili e, per alcune analisi, accedere alla modalità di lancio avanzata.

Il breve video che segue mostra la fase di selezione dell'input durante il lancio di un'analisi.

Selezione dell'input

Input: modalità avanzata

I campioni nella piattaforma possono essere identificati da due stringhe: - `.TE.` (campioni sequenziati); - `.EXT.` (campioni importati manualmente o da database esterno).

Alcuni tool bioinformatici nella piattaforma devono ricevere una specifica istruzione affinchè possano utilizzare, come input, i file appartenenti ad entrambe le tipologie.

Per permettere di fornire tale istruzione, la terza ed ultima fase del sistema di lancio analisi presenta un interruttore per il passaggio dalla "**Modalità di base**" di selezione dell'input alla "**Modalità avanzata**", la quale consente la scelta tramite caselle di spunta.

Il breve video che segue mostra l'applicazione della *Modalità avanzata* su uno dei tool per i quali è disponibile: la pipeline CFSAN.

Modalità avanzata di selezione degli input

Selezione di reference multipli

Alcune analisi, (ad esempio il mapping eseguito con **Ivar** o le pipeline che lo comprendono, come Mapping dei virus segmentati), prevedono la possibilità di selezionare più di un reference. La selezione avviene aprendo la finestra pop-up dei genomi reference, attraverso la quale sarà possibile scegliere un qualunque numero di file, spuntando le caselle corrispondenti. La ripetizione di tale processo non sovrascriverà i reference immessi in precedenza: essi verranno semplicemente aggiunti alla lista.

La lista dei reference selezionati può essere azzerata (tasto **Pulisci**); inoltre, i reference in essa presenti possono essere rimossi singolarmente attraverso l'icona .

La lista può inoltre essere ordinata, facendo uso degli appositi tasti a forma di freccia.

> **Nota:** l'ordine dei reference nella lista influisce sul risultato finale del mapping multi-reference. Ordinare per primi i genomi reference principali.

Il video seguente mostra come interagire con le opzioni per la selezione di reference multipli.

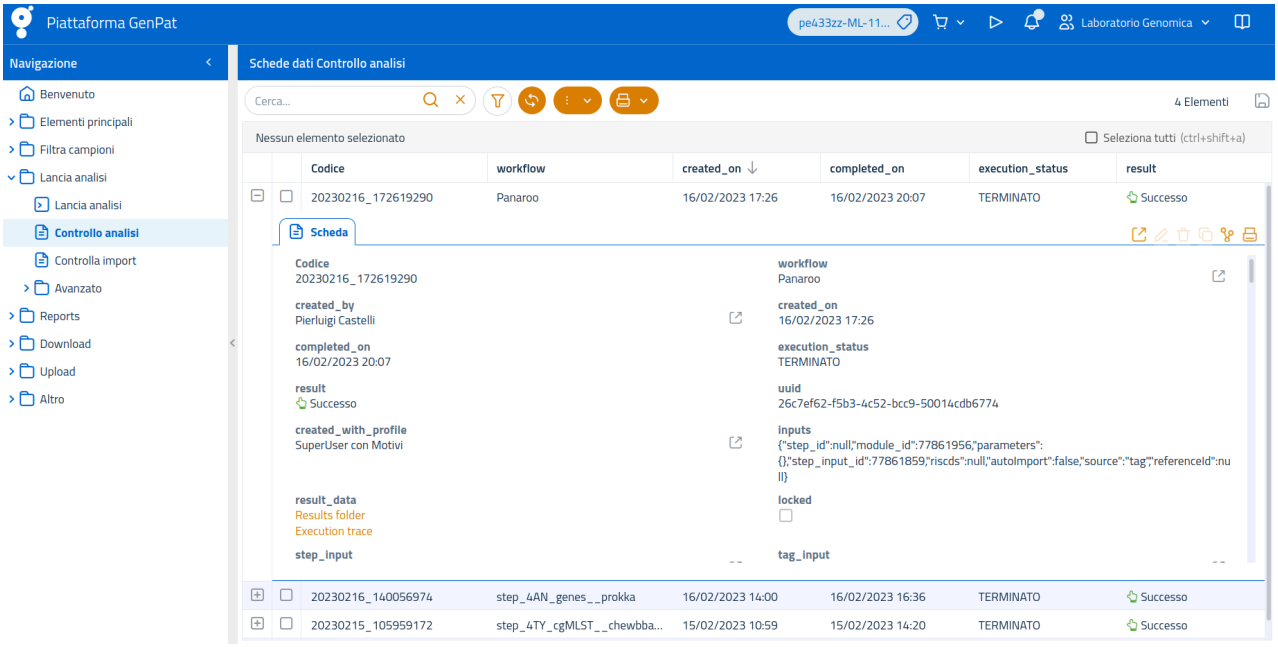
Opzioni di selezione di reference multipli

5.3 Controllo analisi

Le analisi lanciate sono elencate nella pagina **Pipelines lanciate**, raggiungibile dalla sezione **Analisi** del menù di navigazione laterale. Ogni analisi è associata ad una scheda espandibile tramite il bottone "+" situato alla sua sinistra.

5.3.1 Scheda analisi

La scheda dell'analisi elenca informazioni riguardanti la *run* effettuata. Se l'analisi è terminata con successo, si può accedere ad informazioni tecniche sul processo, agli input usati e alle funzioni di "Rilancia l'analisi" e "Importazione" des résultats. Il link alla cartella dei risultati, inoltre, è disponibile direttamente nella scheda (**Results folder**, sezione *Dati risultato*).



Cartella dei risultati (Results folder)

I risultati di un'analisi sono visibili tramite un sistema che permette di navigare e cercare i file generati, senza lasciare la piattaforma. Sono disponibili strumenti di download che consentono la selezione multipla e la ricerca oltre che la possibilità di scaricare sia un singolo file generato, sia una cartella con più file. Il video seguente mostra, in sintesi, le diverse operazioni rese disponibili da queste funzioni.

Cartella dei risultati

Nota La dimensione massima del download di file multipli, tramite le opzioni `Tutto`, `Ricerca` e `Selezionati`, al momento è **20GB**. L'indicatore di dimensione, all'interno del bottone, si colora di rosso quando viene superato il limite consentito. Il download della singola risorsa, invece, non ha limiti.

5.3.2 Importazione delle analisi

Una volta terminata un'analisi con successo, i file prodotti vengono salvati in una cartella temporanea, dove rimarranno per 45 giorni. Durante questo lasso di tempo, gli utenti potranno accedere, come già accennato, alla cartella dei risultati dalla sezione *Dati risultato* della scheda dell'analisi; in alternativa, si possono scaricare i risultati tramite la funzione di download della piattaforma, che permette il download manuale e fornisce il comando `wget` da usare per il download da terminale.

Terminati i 45 giorni, la cartella temporanea verrà rimossa dal server. **Affinchè i risultati dell'analisi vengano salvati permanentemente, questi devono essere importati**.

> Nella maggior parte dei casi, **il lancio di un'analisi in piattaforma che, come input, necessiti di file prodotti da un'altra analisi, richiede che tali file siano stati precedentemente importati**.

Per importare dei risultati, è sufficiente cliccare sul tasto **Importazione**, situato in basso a sinistra nella scheda dell'analisi, come dimostrato brevemente nel video che segue. Il sistema invierà una notifica all'utente quando l'import verrà completato.

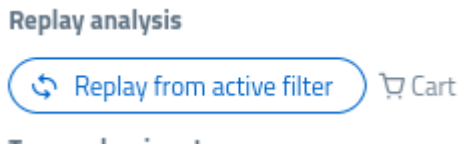
Importazione dell'analisi

> Si noti che non tutte le analisi possono essere importate.

La lista ed i progressi degli *import* sono disponibili alla pagina **Visualizza importazioni** della sezione **Analisi**, nel menù di navigazione laterale.

5.3.3 Rilanciare un'analisi

In fondo alla scheda di un'analisi terminata, è possibile accedere all'opzione **Rilancia l'analisi**. Tale funzione permette di ripetere analisi già eseguite con successo, mantenendo gli stessi parametri e le stesse opzioni precedenti. Il nuovo set di campioni può essere fornito con il carrello attivo o con un tag attivo.



Il video che segue mostra come rilanciare un'analisi.

> **Nota:** Per questioni di privacy e sicurezza dei dati, il video è stato realizzato nell'applicativo demo Cohesive, pertanto l'interfaccia può presentare alcune marginali differenze.

Rilanciare un'analisi

5.3.4 Conserva analisi

All'interno della scheda analisi, è presente la voce **Conserva analisi** la quale permette di scegliere se conservare indefinitamente un'analisi non importabile, impedendo che i relativi dati vengano rimossi dal sistema una volta scaduto il tempo limite di 45 giorni.

Conserva analisi

5.4 Descrizione tools

5.4.1 Introduzione

Le analisi, su isolati batterici o virali disponibili in piattaforma, sono organizzate in categorie. I nomi di tali insiemi seguono una nomenclatura atta a descrivere la tipologia di analisi ed il suo livello di esecuzione.

Sistema di nomenclatura

Prefissi

La categoria di un'analisi ed il suo livello di esecuzione vengono riassunti da un breve codice, usato come prefisso del nome dell'accertamento. Di seguito viene riportata una tabella sintetica di tali prefissi:

1PP analisi di pre-processamento
2AS *tools* per *l'assembly*
2MG analisi di metagenomica
3TX classificazione tassonomica (*taxa*)
4TY tipizzazione *in silico*
4AN annotazione dei genomi (genome annotation)

Il primo carattere del prefisso è un numero indicante l'usuale livello di esecuzione: il **pre-processamento** viene normalmente eseguito prima di qualunque altra analisi e gli viene quindi attribuito livello pari ad **1**; le **analisi tassonomiche** richiedono, come input, i files di *assembly* e sono quindi collocate al livello **3**, dopo pre-processamento ed *assembly*.

Oltre alle analisi lanciabili dagli utenti è presente un'ulteriore classe, denominata **0SQ**, che identifica le analisi per la valutazione della qualità delle sequenze (***Sequence Quality***), lanciate in automatico sulle nuove *reads* aggiunte in piattaforma.

Nomi delle analisi

Il nome dell'analisi segue il codice usato come prefisso. Esso descrive il tipo di manipolazione del dato che verrà effettuata dai *tools* bioinformatici disponibili per tale scopo.

L'analisi *trimming*, ad esempio, è classificata come analisi di pre-processamento; pertanto, il nome completo è **1PP_trimming**. Analogamente, per lanciare un *de novo* assembly sarà necessario utilizzare l'analisi denominata **2AS_denovo** che manterrà lo stesso nome, a prescindere dal fatto che sarà poi possibile scegliere tra più tools bioinformatici.

Nelle sezioni sottostanti sono elencate, in apposite tabelle, tutte le analisi appartenenti ad ogni gruppo presente nel lancia analisi della piattaforma, insieme ad una loro breve descrizione e ai links verso le rispettive pagine.

Suffissi

Molte delle analisi disponibili in piattaforma possono essere eseguite con più softwares alternativi, ognuno con le proprie caratteristiche e differenze rispetto agli altri. A tali programmi si fa riferimento come a "*tools*" bioinformatici o "metodi".

Il nome di un'analisi viene quindi completato usando come suffisso 2 caratteri underscore, ("_"), seguiti dal nome del *tool* scelto; ad esempio, "2AS_denovo__spades" e "2AS_denovo__unicycler" eseguono entrambi il *de novo assembly*; tuttavia, mentre il primo usa il software "spades", il secondo utilizza "unicycler".

Le analisi che possono essere eseguite da più *softwares* permetteranno la scelta del metodo da utilizzare tramite un menù a tendina nella fase 2 (*Tools*) del procedimento di lancio analisi.

I *tools* disponibili per ogni analisi sono elencati nelle rispettive pagine Wiki dedicate alle singole analisi.

Nota: nella maggior parte dei casi, le *long reads* prodotte da sistemi Nanopore, Iontorrent e apparati per *long reads* Illumina devono essere processate in maniera diversa rispetto alle *short reads*. Alcuni *tools* bionformatici disponibili sono, quindi, specifici per effettuare una data analisi **esclusivamente** su *long reads*; possono anche essere presenti intere analisi solo per il processamento di tali *files*.

La pagina "*Tools* bioinformatici per le long reads" permette di consultare le liste di tali analisi.

Analisi su singolo campione

Prefisso	Nome dell'analisi	Descrizione	Tools
1PP	trimming	rimozione dei residui di bassa qualità dalle "raw reads"	trimmomatic fastp chopper
	hostdepl	deplezione delle sequenze dell'ospite: le "reads" vengono mappate contro il genoma dell'organismo ospite selezionato, per rimuovere le sequenze dell'ospite	Bowtie minimap2
	filtering	le reads vengono mappate contro il genoma di un organismo di interesse e vengono tenute solo quelle per cui c'è match	Bowtie minimap2
	downsampling	riduzione del numero di sequenze nelle regioni del genoma con eccessivo coverage verticale non informativo	BBnorm
	generated	con uno script in-house un file FASTA viene convertito in formato FASTQ, usando valori fittizi e predefiniti per la qualità del base-calling. Questa analisi viene usata esclusivamente per semplificare l'uso di alcuni tools bioinformatici che non prevedono solo l'uso di files FASTQ.	fasta2fastq
2AS	denovo	assembly "de novo": costruzione degli scaffold del genoma a partire dall'insieme di contigs	SPAdes unicycler Shovill flye PlasmidSPAdes
	mapping	mappatura delle sequenze contro un genoma di riferimento	Bowtie ivar Snippy Medaka
	hybrid	assembly ibrido di short e long reads	unicycler
2MG	denovo	de novo assembly per metagenomica: il software metaSPAdes costruisce il grafico di *de Bruijn* di tutte le reads usando SPAdes, poi trasformato in grafico di assembly, ricostruendo i percorsi che corrispondono a frammenti genomici all'interno di un metagenoma	metaSPAdes
3TX	class	classificazione tassonomica degli organismi di appartenenza delle reads e controllo delle contaminazioni	kraken kraken2 ConFindr Centrifuge
	species	identificazione della specie (batterica o virale) più vicina / identificazione del miglior reference virale	kmerfinder blast vdabricate mash
4TY	MLST	Multi-Locus Sequence Typing "in silico": utilizza schemi di 7 geni conservati per l'assegnazione di Sequence Type e Clonal Complex	mlst
	cgMLST	analisi filogenetica del core genome Multi-Locus Sequence Typing in silico: chiamata allelica su schemi specie-specifici di alleli relativi al core genome dell'organismo	chewBBACA mentalist blastMLST
	flaA	accertamento specifico per Campylobacter. Determinazione della variante del locus flaA "in silico" (MLST per flaA)	flaA
	lineage	assegnazione del lineage per SARS-CoV2	Pangolin

		assegnazione del lineage per il West Nile Virus	westnile
	wgMLST	MLST sull'intero genoma (whole genome)	chewBBACA
	plasmid	tipizzazione e ricostruzione delle sequenze plasmidiche a partire dagli assembly da Whole Genome Sequencing (WGS)	MOB-suite
	genes	genome annotation - annotazione funzionale del genoma tramite ricerca degli ORF (Open Reading Frame) nel genoma dell'organismo e identificazione delle possibili proteine codificate	Prokka
4AN	AMR	predizione della presenza di geni di antibiotico-resistenza	abricate blast staramr ResFinder filtering

Analisi Multi Sample

Tipo di analisi	Nome analisi / Tool	Descrizione
	Grapetree	costruisce alberi MST (Minimum Spanning Tree) e NJ (Neighbor Joining) in formato nwk, a partire dai profili allelici
	Reportree	a partire dai profili allelici, costruisce alberi MST (Minimum Spanning Tree) in formato nwk e genera un file geoJSON a partire dai metadati
Gene-by-gene based clustering	Augur	esegue la pipeline Augur di Nextrain per l'analisi filogenetica, producendo un albero con algoritmo Maximum Likelihood (ML) a partire dai FASTA provenienti dal mapping o dal "de novo" assembly
	Surveillance	lancia manualmente ReporTree con la stessa modalità usata per il processo automatico di sorveglianza
	Panaroo	calcola una matrice binaria di presenza/assenza di geni accessori nei genomi dei campioni, a partire dai file gff prodotti da Prokka (genome annotation)
Pangenome extraction	Snippy-core	esegue Snippy per identificare le mutazioni (SNPs e indels) tra le read e un genoma aploide di riferimento, seguito da Snippy-core per costruire il file core.vcf a partire dai vcf prodotti da Snippy. Il file core.vcf contiene le mutazioni "core" tra quelle elencate nei singoli files vcf di Snippy
	CFSAN	identificazione di SNP con analisi filogenetica reference-based
SNP-based clustering	kSNP3	identificazione di SNP senza reference con analisi filogenetica. Produce un albero Maximum Likelihood
	VCF2MST	costruisce rapidamente un albero MST a partire da un file VCF senza necessità di effettuare inferenze filogenomiche

Pipeline

Oltre al lancio delle singole analisi, in piattaforma sono disponibili anche delle pipeline automatiche, costituite da analisi già concatenate tra loro. Le pipelines sono concepite per rendere più semplice e rapida l'esecuzione di un flusso di lavoro di uso frequente.

Nella tabella sottostante sono elencate, per ogni pipeline, le singole analisi che la costituiscono (colonna "Analisi"; per informazioni sui singoli software usati per le analisi nelle pipeline, fare riferimento alle corrispondenti pagine Wiki).

Nome Pipeline	Descrizione	Analisi
Emergenza Covid	assembly rapido di campioni di SARS-CoV2 e assegnazione lineage	2AS_mapping + 4TY_lineage
Deplezione & *de novo*	deplezione delle reads dell'ospite dalle reads trimmate e successivo *de novo* assembly	1PP_hostdepl + 2AS_denovo
Draft del Genoma	mapping e *genome annotation*. Il mapping viene eseguito sia con Bowtie, sia con Snippy.	2AS_mapping + 4AN_genes
Ricerca di enterotossine di *S. aureus*	assembly *de novo* e successivo BLAST per identificare la presenza del gene codificante per enterotossina nel genoma	2AS_denovo + 4AN_AMR
NgsManager	macro-pipeline che esegue, in base alla tipologia di campione, i moduli per campioni *SARS-CoV2*, *Batteri* o *Virus*	pipeline "Processamento *Raw Reads*", "Emergenza Covid", "WGS sui Batteri", "Typing sui Batteri", "Draft del Genoma"
Processamento *Raw Reads*	controllo qualità delle reads, trimming e classificazione virus/batteri	0SQ_rawreads + 1PP_trimming + 3TX_class
Filtraggio e *de novo*	rimuove dalle *raw reads* le letture provenienti da organismi di non interesse e, consecutivamente, avvia il *de novo* assembly	1PP_filtering + 2AS_denovo
Tipizzazione sui Batteri	calcolo della specie; calcolo del coverage orizzontale e verticale; annotazione geni; identificazione geni di virulenza e antibiotico resistenza; tipizzazione	2AS_mapping + 3TX_species + 4AN_genes + 4AN_AMR + 4TY_wgMLST + 4TY_cgMLST + 4TY_MLST + 4TY_flaA
WNV - lineage calculation and mapping	calcolo del lineage per campioni di *West Nile Virus* e mapping contro il reference del lineage calcolato	4TY_lineage + 2AS_mapping
Mapping Virus Segmentati	esegue il mapping dei segmenti di genoma dei virus segmentati usando più di un reference	2AS_mapping
Plasmidi (AMR)	pipeline per la tipizzazione dei plasmidi con MOB-Suite	4TY_plasmid + 4AN_AMR
nf-core/ampliseq	pipeline dalla community nf-core per il sequenziamento ed il *denoising* degli ampliconi. Supporta coppie di read Illumina, single-end Illumina, PacBio o IonTorrent	nf-core/ampliseq

Pipeline per il controllo qualità (QC):

Nome Pipeline	Descrizione
QC FastQC	pipeline di controllo qualità per eseguire singolarmente il software FastQC sulle raw reads o sulle reads trimmate
QC Nanoplot	pipeline di controllo qualità per eseguire singolarmente il software Nanoplot sulle raw reads da apparati Nanopore
QC Quast	pipeline di quality check per eseguire singolarmente il software Quast sui file dell'assembly o dell'assembly ibrido

5.4.2 *Tools* bioinformatici per tecnologie non-Illumina

Di seguito sono riportati i tool bioinformatici specifici per tecnologie di sequenziamento diverse da Illumina *paired-end*. Ognuno dei *tools* sotto riportati è disponibile nell'analisi di riferimento ed è da usare **esclusivamente** con *reads* prodotte da sequenziatore della tecnologia specificata. Analogamente, le *reads* prodotte da tali dispositivi potranno essere processate **solo** con gli appositi programmi e non con quelli utilizzati per *short reads*.

Analisi per *reads* da apparati Oxford Nanopore Technologies (ONT)

Analisi	Tool	Note
1PP_filtering__minimap2	minimap2	solo per *reads* provenienti da apparati Nanopore
1PP_hostdepl__minimap2	minimap2	solo per *reads* provenienti da apparati Nanopore
1PP_trimming__chopper	chopper	solo per *reads* provenienti da apparati Nanopore
2AS_denovo__flye	Flye	solo per *reads* provenienti da apparati Nanopore
2AS_hybrid__unicycler	Unicycler	usato per l'assembly ibrido di *short reads* Illumina e *reads* Nanopore
2AS_mapping__medaka	Medaka	solo per *reads* provenienti da apparati Nanopore
3TX_class__centrifuge	Centrifuge	solo per *reads* provenienti da apparati Nanopore

Analisi per *long reads* provenienti da apparati Thermofisher IonTorrent

Analisi	Tool	Note
1PP_trimming__fastp	fastp	solo per *reads* provenienti da apparati IonTorrent
1PP_trimming__trimmomatic	trimmomatic	per *reads* IonTorrent e per *short reads paired-end*
2AS_denovo__plasmidspades	plasmidSPAdes	per *reads* da apparati IonTorrent (assemblaggio di sequenze plasmidiche)
2AS_denovo__shovill	shovill	per *reads* IonTorrent e *short reads paired-end*
2AS_denovo__spades	SPAdes	per *reads* IonTorrent e coppie di *short reads*
2AS_mapping__bowtie	bowtie	per *reads* IonTorrent e *short reads paired-end*
2AS_mapping__ivar	iVar	per *reads* IonTorrent e *short reads paired-end*
2AS_mapping__snippy	snippy	per *reads* IonTorrent e *short reads paired-end*
3TX_class__confindr	ConFindr	per *reads* IonTorrent e *short reads paired-end*
3TX_class__kraken	kraken	per *reads* IonTorrent e *short reads paired-end*
3TX_class__kraken2	kraken2	per *reads* IonTorrent e *short reads paired-end*
3TX_species__mash	Mash	per *reads* IonTorrent e *short reads paired-end*
4AN_AMR__resfinder	resfinder	per *reads* IonTorrent e *short reads paired-end*

5.4.3 Single sample

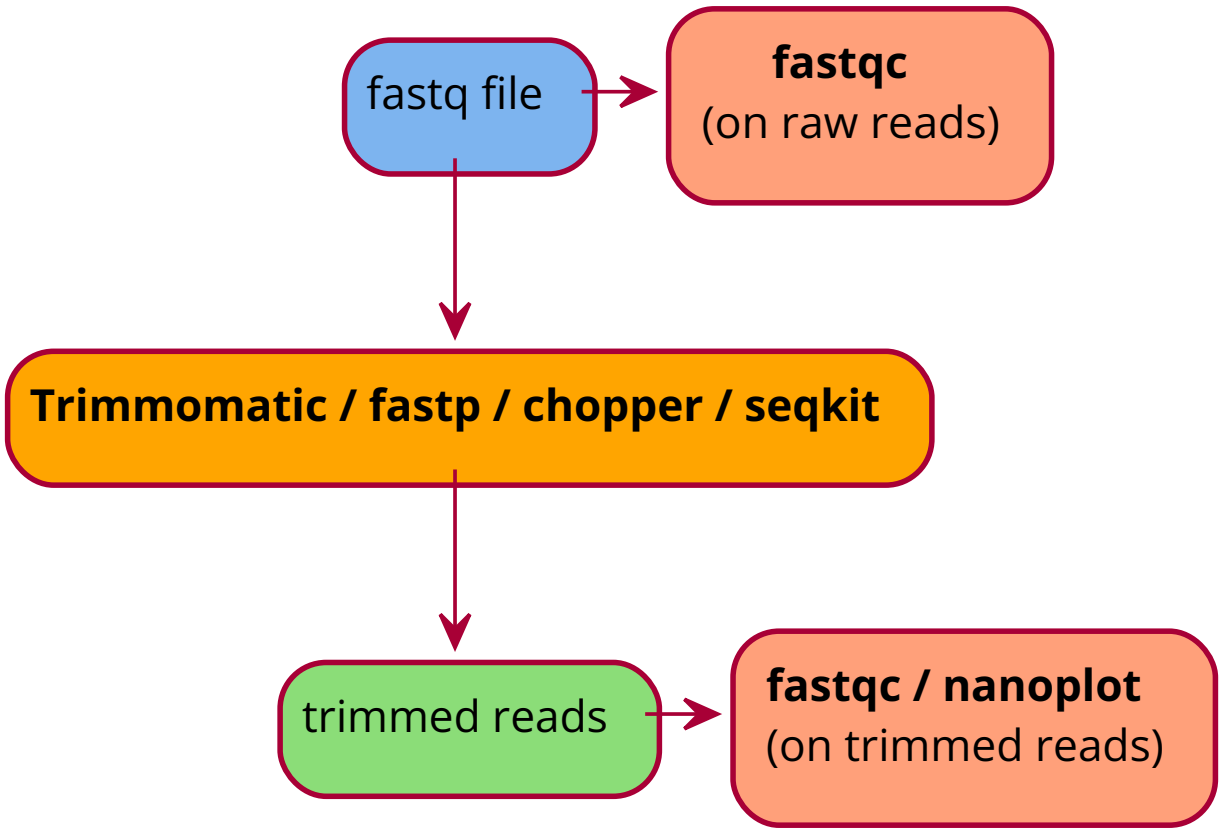
1 Pre-processing

1PP_trimming

Introduzione

Il *trimming* ha lo scopo di rimuovere i residui di bassa qualità dalle read ottenute dal sequenziatore.

I tool disponibili per il trimming in GenPat sono "trimmomatic" e "fastp". L'analisi **1PP_trimming** è costituita dall'esecuzione di uno dei tools per trimming e di **fastQC**. Quest'ultimo elabora le statistiche di qualità delle *reads* e viene eseguito sia sulle *raw reads*, prima del trimming, sia sulle nuove *reads* trimmate.



Tool
Input
Result
Quality Check tool

Lancia 1PP_trimming

Una volta selezionata l'analisi **1PP_trimming** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico da usare tra quelli disponibili per l'analisi. I tools utilizzabili, in questo caso, sono:

- **trimmomatic** - tool di trimming delle read per dati NGS Illumina

  Pagina GitHub di trimmomatic: https://github.com/usadellab/Trimmomatic

- **fastp** - pre-processamento all-in-one per i file fastQ

  Pagina GitHub di fastp: https://github.com/OpenGene/fastp

- **chopper** - filtraggio e trimming dei file fastQ provenienti da sequenziamento a *long reads*, come PacBio o ONT

  Pagina GitHub di chopper: https://github.com/wdecoster/chopper

Nota 1: Per il trimming di *long reads* usare esclusivamente **chopper**; per il trimming di *short reads* usare **trimmomatic** o **fastp**.

Nota 2: La guida ai parametri di Chopper è disponibile nella sua pagina GitHub ufficiale.

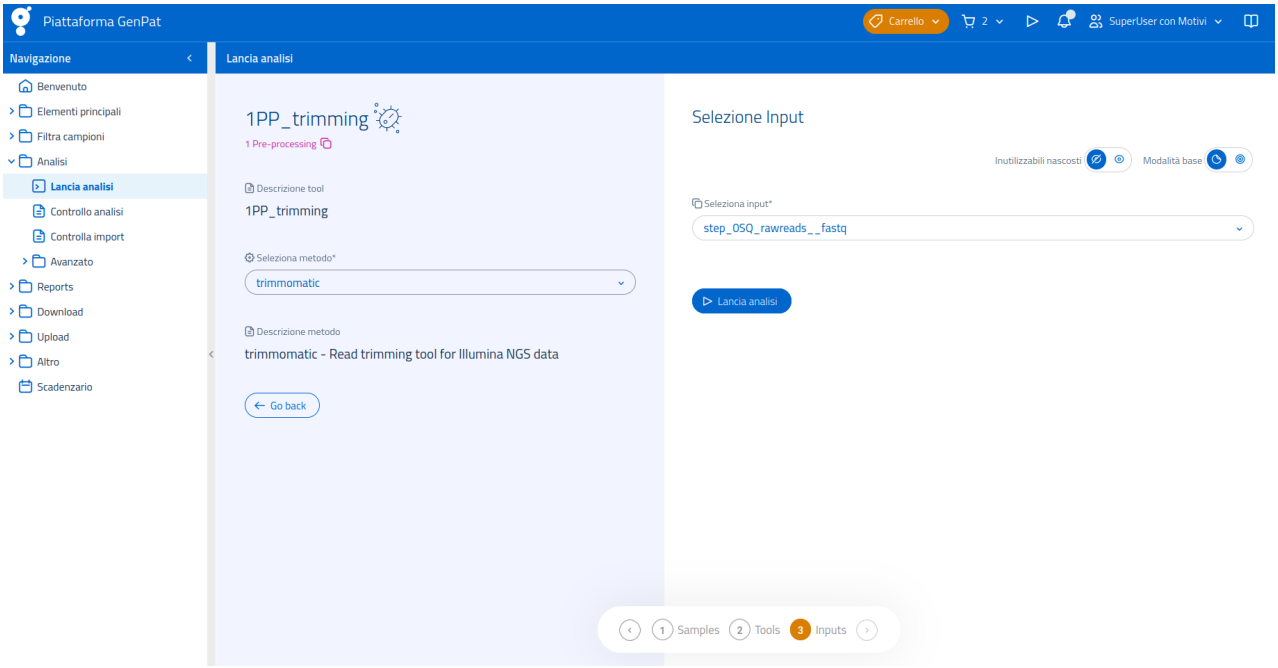
- **seqkit** - un kit di tool multipiattaforma e ultrarapido per la manipolazione di file fastA/fastQ

  pagina GitHub di seqkit: https://bioinf.shenwei.me/seqkit/usage/#grep

Nota 3: Questo strumento elimina, dalle reads, tutte le letture in cui è presente la sequenza dell'adattatore indicata in "Custom adapter sequence", utilizzando la funzione "grep" del tool.

Nota 4: Solo per illumina reads. Per questa analisi è possibile utilizzare come input sia l'analisi *0SQ_rawreads* che *1PP_trimming*.

Nell'ultima sezione viene invece specificato l'input: poichè il trimming viene eseguito sui fastQ ottenuti dal sequenziatore, le scelte disponibili saranno "step_0SQ_rawreads__fastq" per i fastq delle reads interne (codice 20XX.TE.XXXX.X.X) e "step_0SQ_rawreads__external", per i fastq importati (codice 20XX.EXT.XXXX.X.X). La schermata di selezione input di *1PP_trimming* mette a disposizione dell'utente la modalità di selezione input avanzata, in quanto l'analisi è eseguibile sia su file di origine interna, sia su file importati.



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 3 sotto-cartelle:

- **meta**: (*metadati*) in cui vengono salvati i file di log e di configurazione del processo eseguito;
- **qc**: (*quality check*) in cui si trovano i file del controllo qualità. In questa cartella sono presenti le cartelle *meta* e *result*. Per questa analisi, il controllo qualità viene eseguito con **fastqc**. Il calcolo del QC non viene eseguito in caso di lancio analisi con Seqkit;
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

Le tabelle in basso presentano le liste di file presenti nelle cartelle, in base al tool usato, insieme ad alcune informazioni utili. Si noti che nel nome del file verrà riportato il nome del programma che è stato scelto per l'analisi.

Trimmomatic

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_R1_trimmomatic.fastq	read 1 (R1) trimmata	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_R2_trimmomatic.fastq	read 2 (R2) trimmata	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_unpaired_trimmomatic.fastq	unpaired reads trimmate	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_R1_trimmomatic_fastqc.html	qualità delle read R1	cartella qc > result
DSXXXXXXXX-DTXXXXXX_ID_R1_trimmomatic_fastqc.zip	qualità R1 (file zip)	cartella qc > result
DSXXXXXXXX-DTXXXXXX_ID_R2_trimmomatic_fastqc.html	qualità delle read R2	cartella qc > result
DSXXXXXXXX-DTXXXXXX_ID_R2_trimmomatic_fastqc.zip	qualità R2 (file zip)	cartella qc > result
DSXXXXXXXX-DTXXXXXX_ID_unpaired_trimmomatic_fastqc.html	qualità delle unpaired read	cartella qc > result
DSXXXXXXXX-DTXXXXXX_ID_unpaired_trimmomatic_fastqc.zip	qualità delle unpaired read (file zip)	cartella qc > result

fastp

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_fastp_R1.fastq.gz	read 1 (R1) trimmata	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_fastp_R2.fastq.gz	read 2 (R2) trimmata	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_fastp_unpaired.fastq.gz	unpaired read trimmate	cartella "result"

chopper

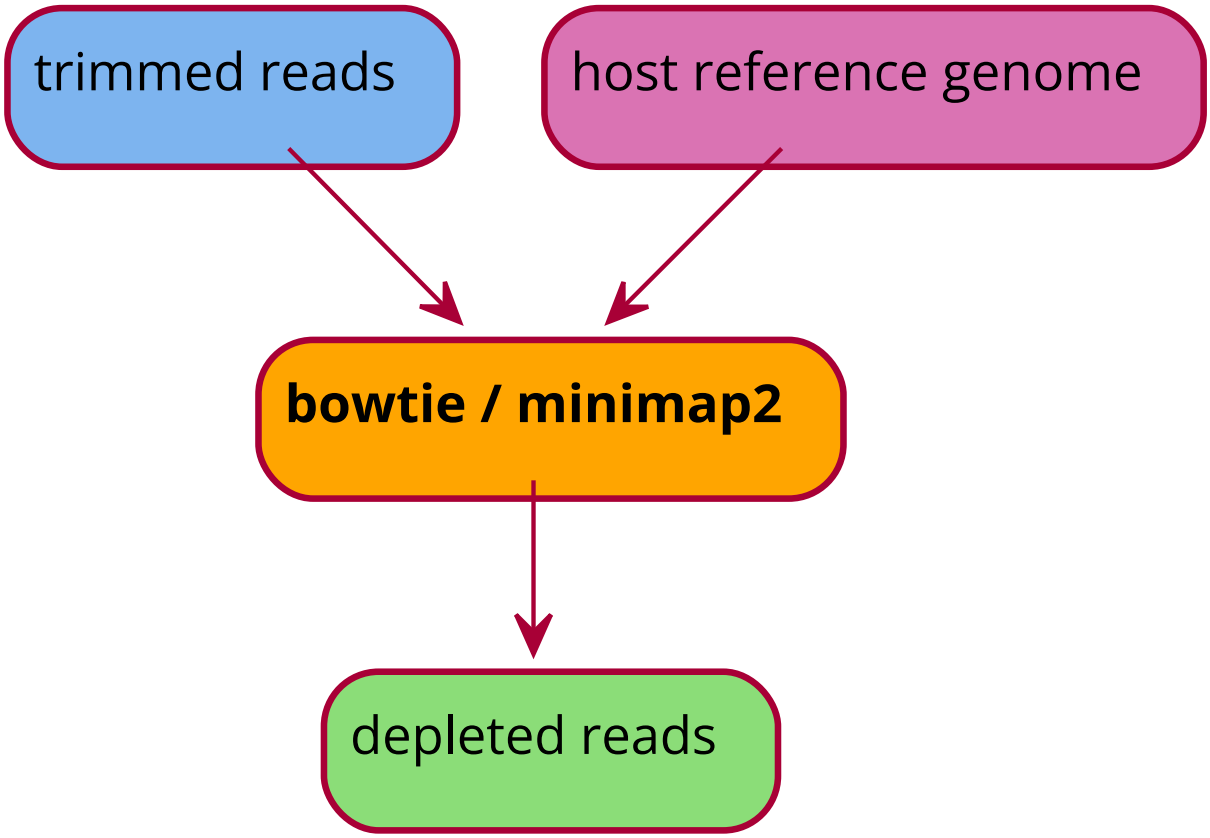
File	Descrizione	Posizione
filtered_reads.fastq.gz	long read trimmate	cartella "result"

1PP_hostdepl

Introduzione

L'analisi *1PP_hostdepl* mappa le read contro il genoma dell'organismo ospite selezionato, permettendo di rimuovere le read contaminanti dell'ospite dai file FASTQ (*deplezione* delle sequenze dell'ospite).

I file di output sono dei nuovi file FASTQ di read ripulite.



Tool
Input
Result
Reference

Lancia analisi 1PP_hostdepl

Una volta selezionata l'analisi **1PP_hostdepl** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico ("metodo") da usare tra quelli disponibili per l'analisi. I tool utilizzabili per questa analisi sono:

- Per tutte le read:
- **Bowtie** - un allineatore di *short reads* ultrarapido ed efficiente in termini di memoria.

Pagina GitHub di Bowtie: https://github.com/BenLangmead/bowtie

- Solo per *long reads* da apparati con tecnologia nanopore:
- **minimap2** - un programma versatile di allineamento che allinea sequenze di DNA o mRNA contro un ampio database di riferimento.

Pagina GitHub di minimap2: https://github.com/lh3/minimap2

Nella sezione dedicata alla selezione dell'input, sarà necessario indicare sia le read di input che un reference, come parametro addizionale. La deplezione dell'ospite avviene, infatti, mappando le read dei file FASTQ contro un genoma di riferimento, quello dell'ospite, le cui sequenze contaminanti devono essere riconosciute ed eliminate dalle read. Nella definizione del parametro *Host*, sarà pertanto possibile selezionare uno qualunque dei genomi di riferimento per specie caricate in GenPat.

Nota: Nel caso in cui l'ospite desiderato non sia presente tra quelli disponibili, si invita ad inoltrare una richiesta tramite e-mail al reparto bioinformatica (bioinformatica@izs.it), per l'aggiunta del genoma di riferimento in GenPat.

La schermata di selezione input di questa analisi mette a disposizione dell'utente la modalità di selezione input avanzata, in modo da poter selezionare contemporaneamente file FASTQ precedentemente processati in modo diverso.

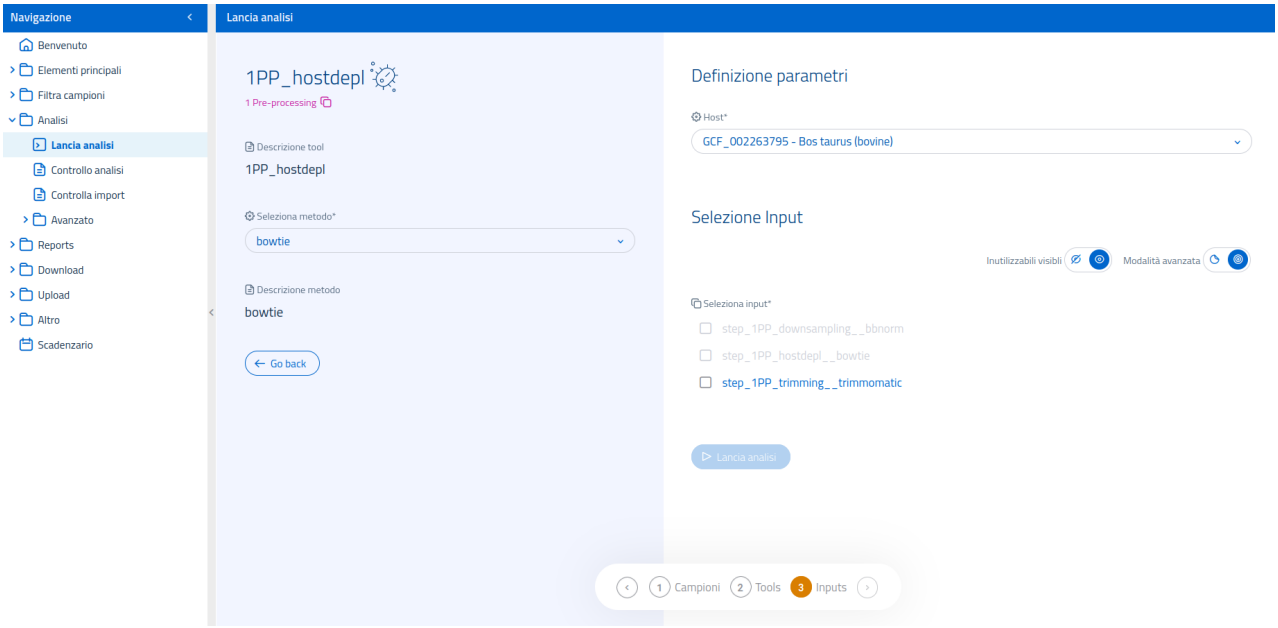
Input per Bowtie:

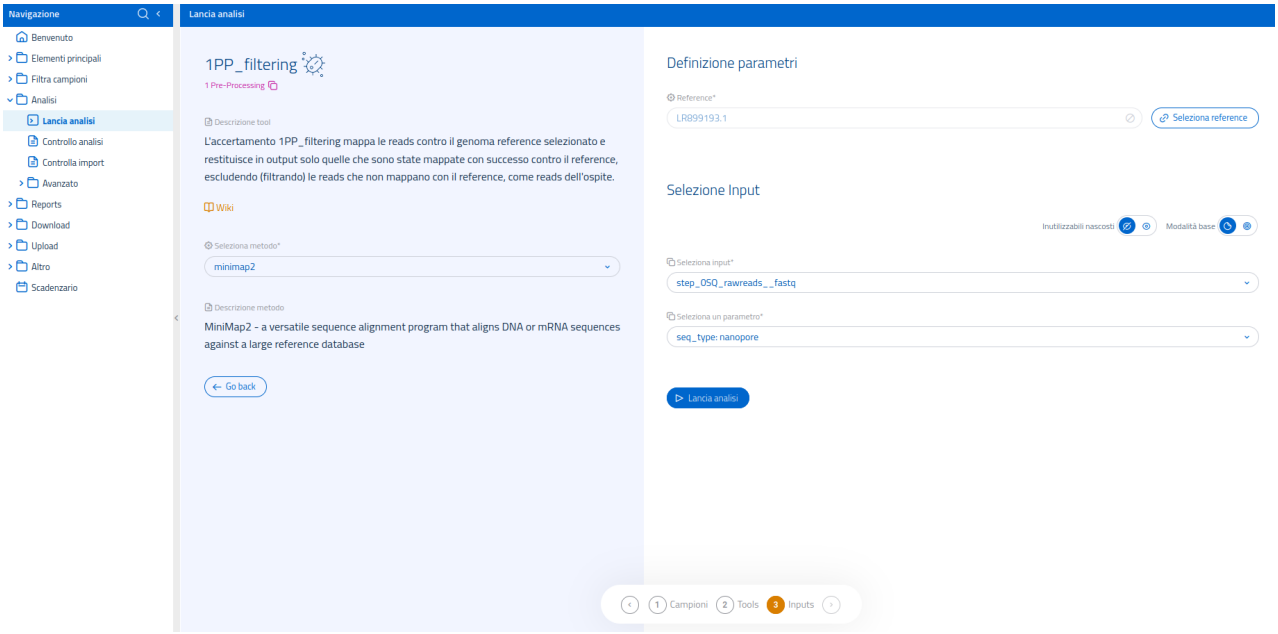
- step_1PP_trimming
- step_1PP_hostdepl
- step_1PP_downsampling

Input per minimap2:

- step_0SQ_rawreads

Se il software selezionato è **minimap2**, gli input potranno provenire eclusivamente da **step_0SQ_rawreads** di **single** *long reads* prodotte da dispositivi basati su tecnologia nanopore. Per ulteriori informazioni sul software **minimap2**, si invita a consultare la pagina GitHub ufficiale di minimap2.





Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 2 sotto-cartelle:

- **meta**: (*metadati*) in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

Le tabelle sottostanti presentano la lista di files prodotti dai tool disponibili per 1PP_hostdepl, insieme ad alcune informazioni utili.

Bowtie:

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_R1_vdhost_HOSTID.fastq	read 1 (R1) depleta	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_R2_vdhost_HOSTID.fastq	read 2 (R2) depleta	cartella "result"

minimap2:

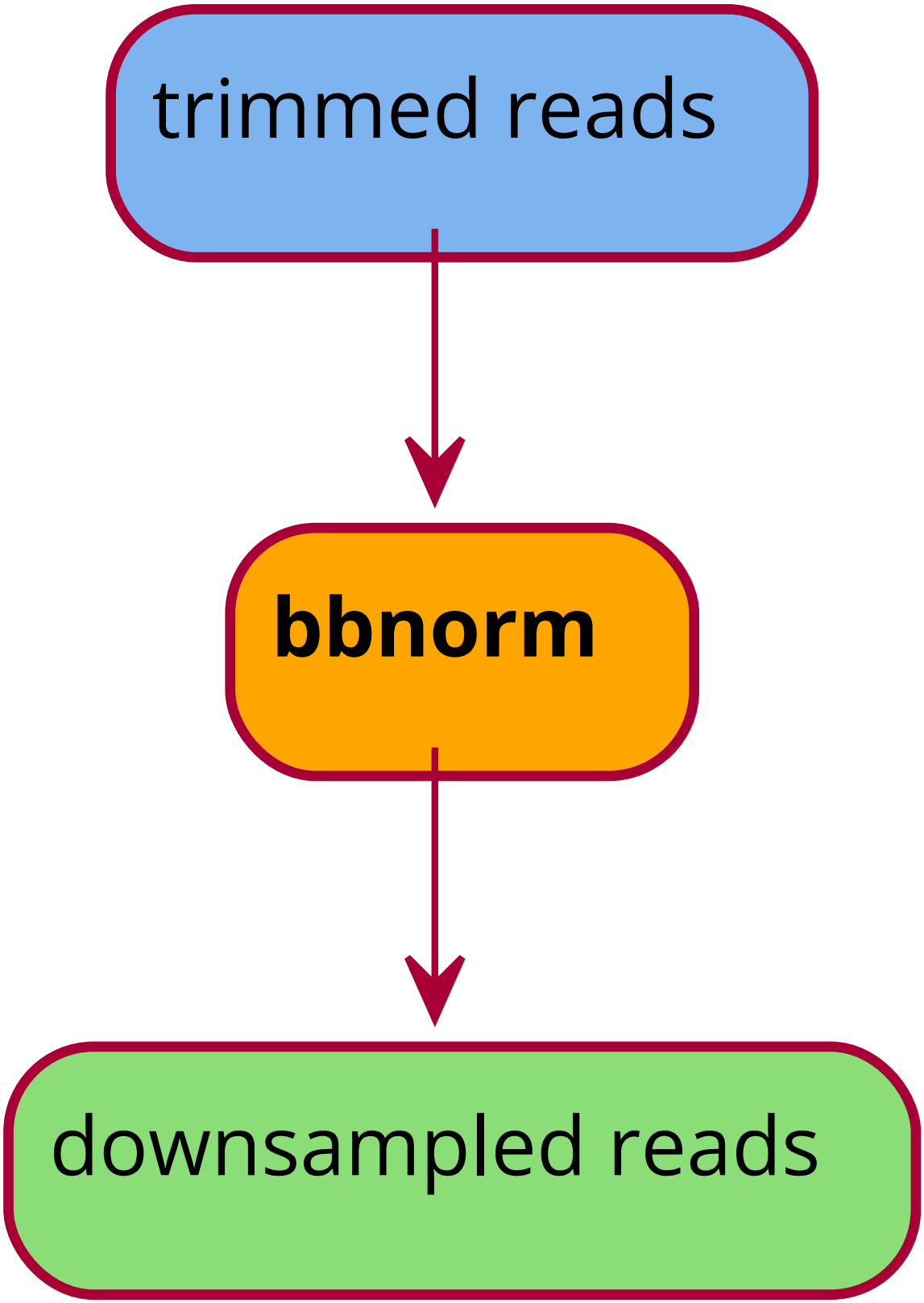
File	Descrizione	Posizione
DSXXXXXX-DTXXXXXX_ID_minimap2_HOSTID.fastq.gz	fastq depleto della *single read* da nanopore device (file compresso)	cartella "result"

1PP_downsampling

Introduzione

Il *downsampling* è definibile come un processo di riduzione di *read depth* (*coverage* verticale), in specifiche posizioni o all'interno di determinate regioni del genoma, allo scopo di diminuire il numero di read non informative ed aumentare la qualità dell'assembly e delle analisi a valle. BBnorm normalizza la *read depth* usando un *hashing* di k-mers.

I protocolli di sequenziamento possono spesso provocare l'accumulo di read per una sezione del genoma. L'eccesso di informazione, in questi casi, rallenta l'esecuzione delle analisi a valle, costrette a processare un volume maggiore di dati, pur non essendo tutti questi informativi. In questi casi, il downsampling permette di evitare l'allungamento dei tempi di calcolo, scartando coppie di read in eccesso, fino al raggiungimento di un valore limite specificato, ovvero il *coverage* verticale desiderato.



Tool

Lancia Analisi 1PP_downsampling

Una volta selezionata l'analisi **1PP_downsampling** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico ("metodo") da usare tra quelli disponibili per l'analisi. Il tool utilizzabile, in questo caso, è:

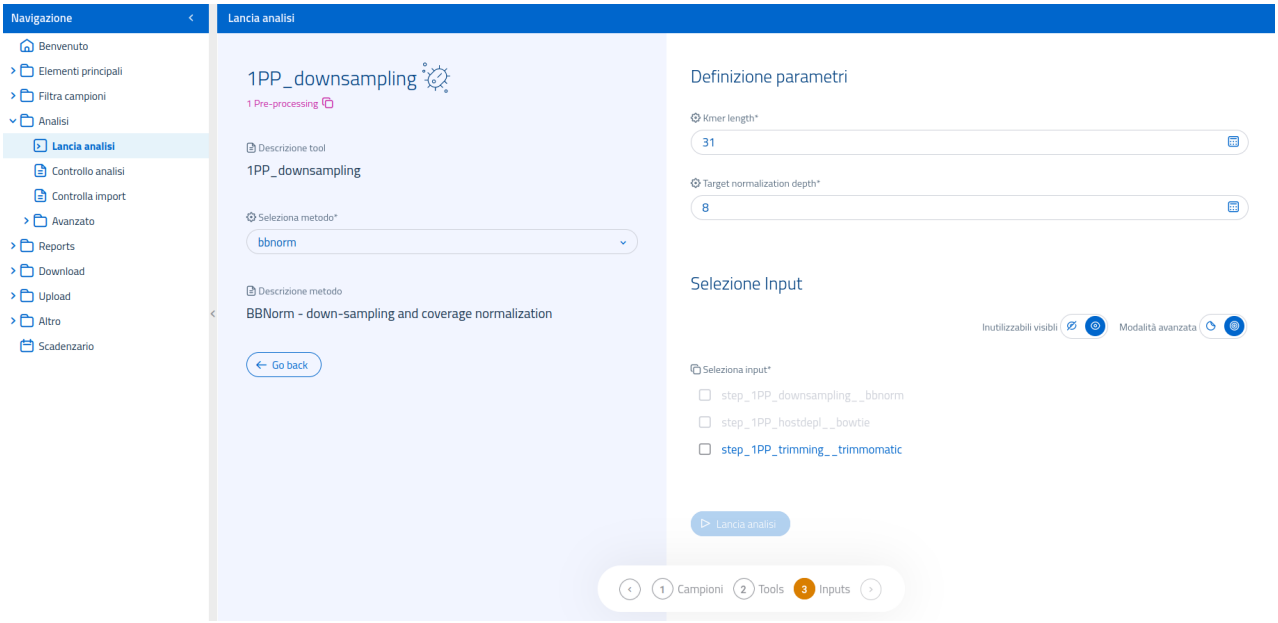
- **BBnorm** - strumento di correzione degli errori e normalizzazione basato su K-mer.

> **Manuale di BBnorm:** https://manpages.debian.org/unstable/bbmap/bbnorm.sh.1.en.html

Nella sezione dedicata alla selezione dei parametri, sarà necessario indicare sia le *read* di input sia alcuni parametri addizionali, atti a determinare il coverage verticale finale che si desidera avere per i genomi dei campioni.

La sezione di selezione degli input mette a disposizione la modalità di selezione input avanzata. E' infatti possibile che il downsampling venga lanciato a seguito di analisi eseguite a monte e diverse tra loro:

- step_1PP_trimming
- step_1PP_hostdepl
- step_1PP_downsampling



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Parametri

> Una panoramica dei parametri di BBnorm è disponibile al seguente link: https://manpages.debian.org/unstable/bbmap/bbnorm.sh.1.en.html.

I 2 parametri "kmer length" e "Target" dipendono da:

- specie del campione (e quindi lunghezza del genoma);
- per "Target", dalla lunghezza dei kmer da usare, come specificata nel parametro "kmer length".

Kmer length ha un valore di default di 31. Il parametro supporta una qualunque lunghezza per i kmer ma viene consigliata una lunghezza inferiore a 32 per una maggiore efficienza del tool.

Target indica la normalizzazione dei **k-mer** usati da BBnorm per eseguire il *downsampling*.

> **Attenzione:** il parametro "Target normalization value" **non** corrisponde al coverage verticale finale che si otterrà con l'esecuzione del tool; piuttosto, secondo la guida ai parametri di BBnorm, esso controlla la profondità verticale dei k-mers che vengono usati dal tool per normalizzare le *read*.

Per via del funzionamento del software è pertanto necessario determinare sperimentalmente il valore del secondo parametro, in relazione al primo ("kmer length") e alla specie.

Di seguito sono riportati due esempi per *Listeria monocytogenes*:

- specie *L. monocytogenes*, kmer length = 30, target = 31 -> vertical coverage = 40X
- specie *L. monocytogenes*, kmer length = 30, target = 8 -> vertical coverage = 10X

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 2 sotto-cartelle:

- **meta**: (*metadati*) in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

La tabella in basso presenta la lista di files presenti nelle cartelle, insieme ad alcune informazioni utili.

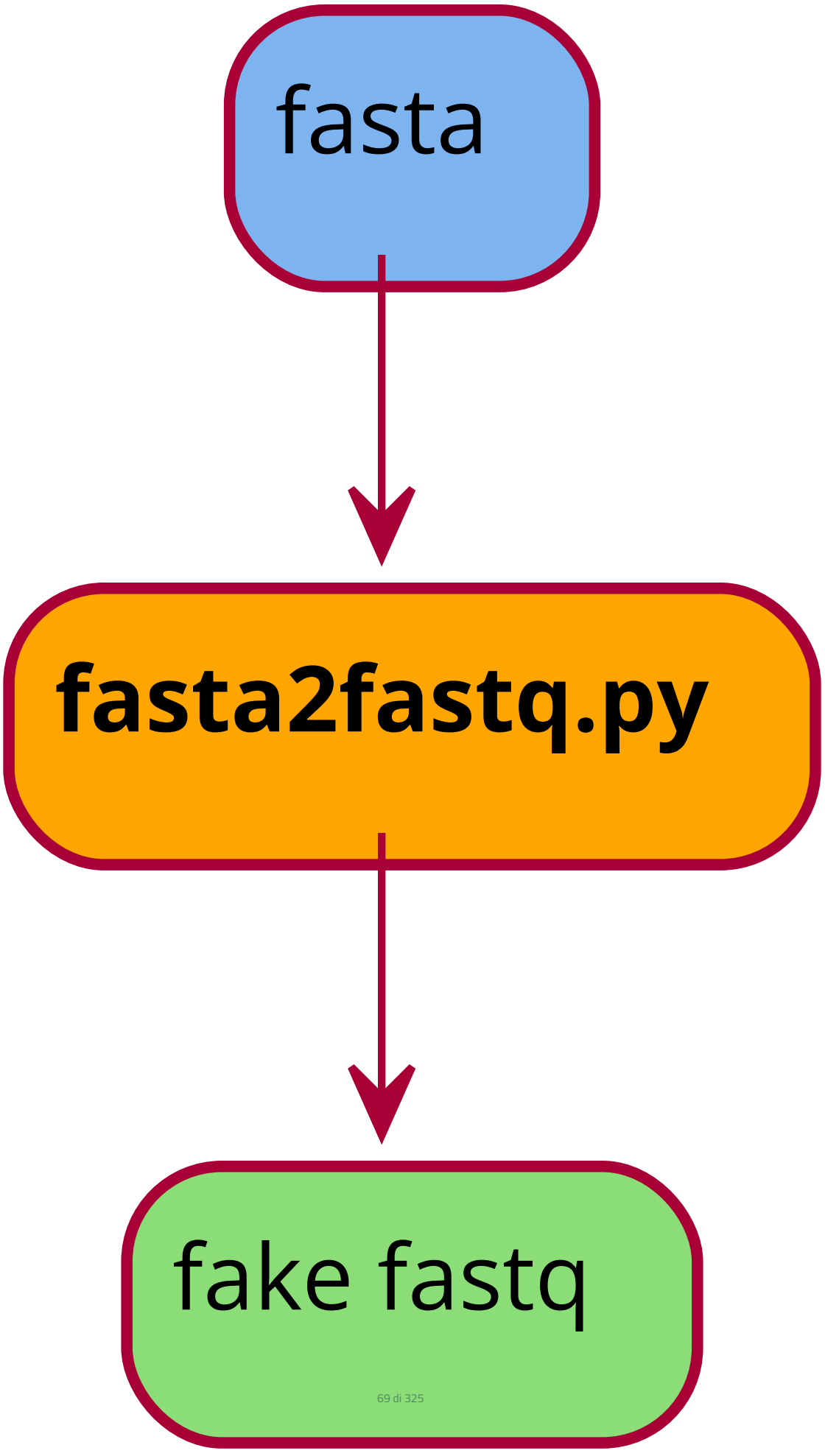
File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_bbnorm_kXX_tX_R1.fastq.gz	read 1 (R1) downsampled	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_bbnorm_kXX_tX_R2.fastq.gz	read 2 (R2) downsampled	cartella "result"

1PP_generated

Introduzione

L'analisi *1PP_generated* costruisce un file FASTQ a partire da un file FASTA. Per ottenere tale risultato viene utilizzato uno script realizzato appositamente per GenPat, (**fasta2fastq.py**), che genera le righe con statistiche di qualità fittizie e modifica gli *headers* perchè siano conformi al formato fastq. Il software costruisce i file per entrambi i filamenti antiparalleli di DNA (R1 e R2).

Attenzione: Poichè le statistiche di qualità generate sono fittizie, lo scopo di questa analisi è unicamente quello di permettere l'uso di sequenze disponibili solo come file FASTA, per analisi i cui programmi accettano, normalmente, solo il formato fastq come input. I file di output, pertanto, non devono essere utilizzati, in sostituzione delle sequenze FASTQ effettive, nei workflow per i quali la valutazione della qualità delle *reads* ha importanza.


fasta
fasta2fastq.py
fake fastq

Lancia analisi 1PP_generated

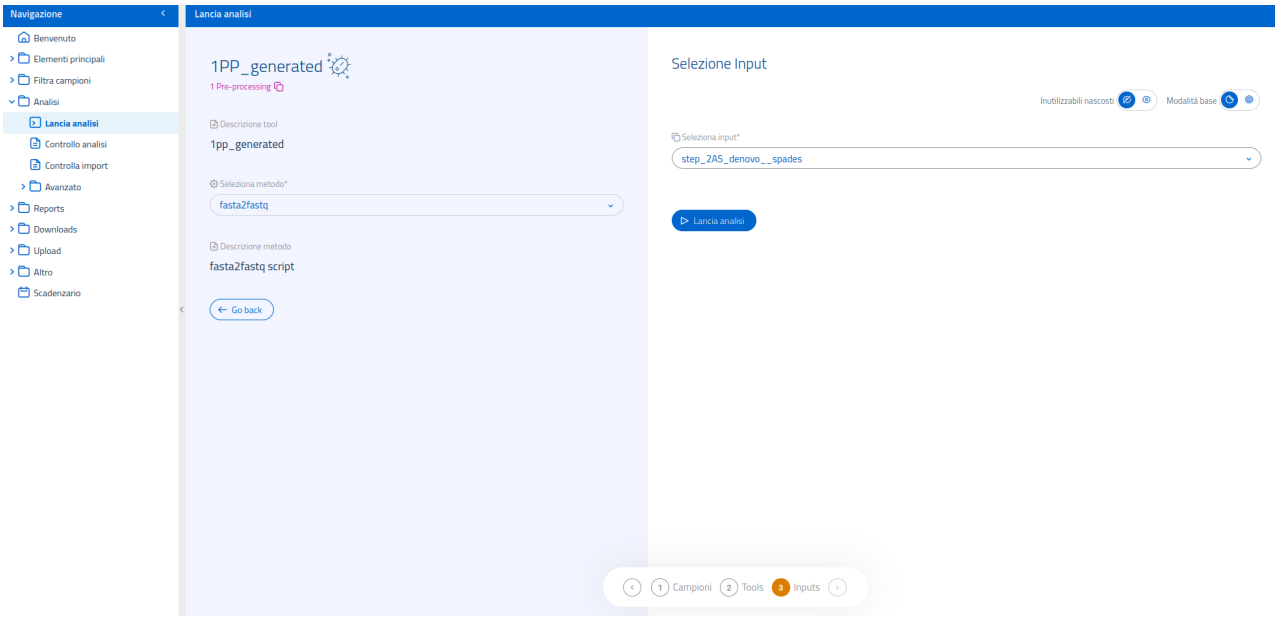
Una volta selezionata l'analisi **1PP_generated** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico ("metodo") da usare tra quelli disponibili per l'analisi. Il solo tool utilizzabile per questa analisi è **fasta2fastq**.

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per 1PP_generated sono:

- step_2AS_mapping
- step_2AS_denovo
- step_2AS_import

Sarà necessario specificare le sequenze di input - in genere provenienti da *de novo* assembly o da mapping - e il genoma reference usato per il mapping, nel caso venga selezionato quest'ultimo come input.



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 2 sotto-cartelle:

- **meta**: (*metadati*) in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

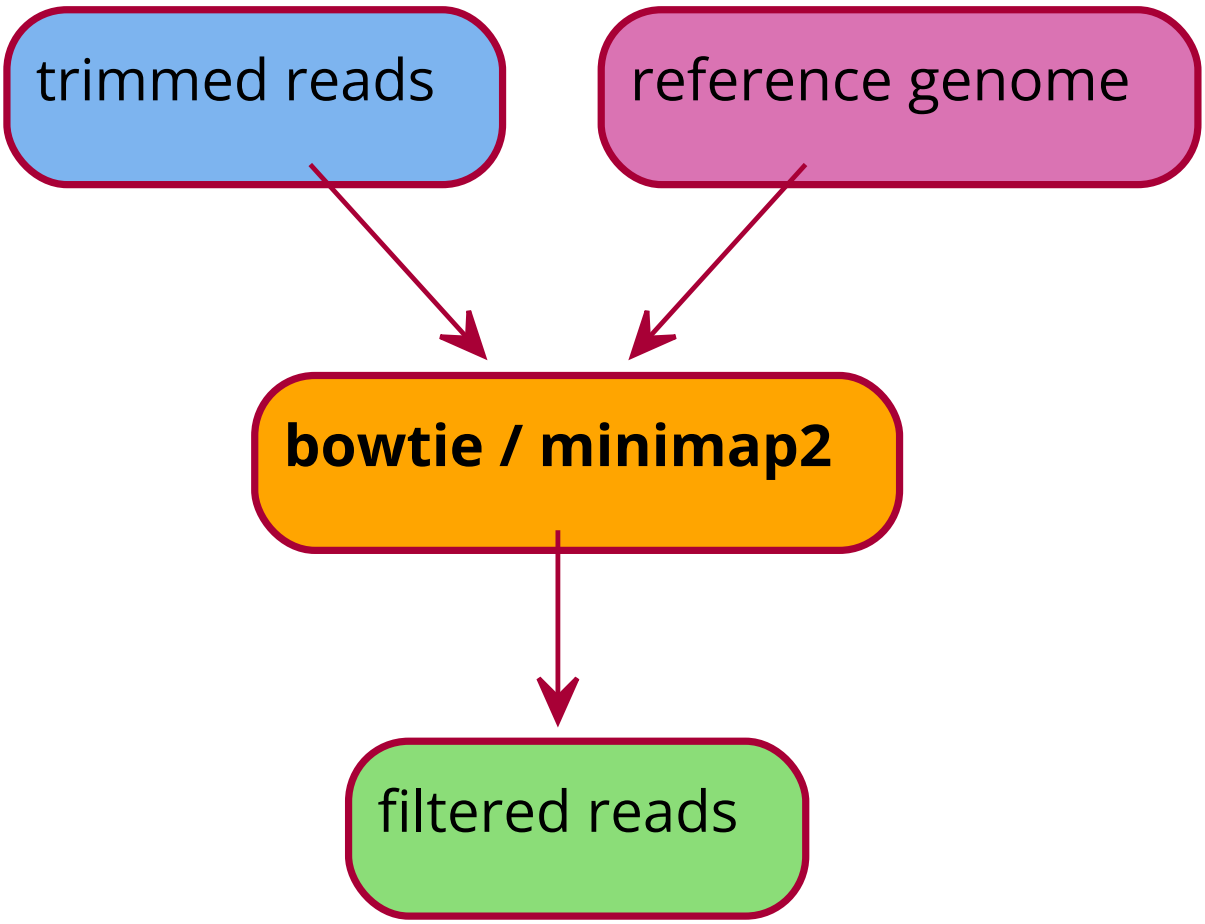
La tabella sottostante elenca i file prodotti da fasta2fastq.py.

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_fasta2fastq_R1.fastq.gz	fastq fittizio del filamento + (file compresso: .gz)	cartella "results"
DSXXXXXXXX-DTXXXXXX_ID_fasta2fastq_R2.fastq.gz	fastq fittizio del filamento - (file compresso: .gz)	cartella "results"

1PP_filtering

Introduzione

L'analisi *1PP_filtering* mappa le *read* contro il genoma reference selezionato e restituisce in output solo quelle che sono state mappate con successo contro il reference, escludendo ("filtrando") le altre.



Tool
Input
Result
Reference

Lancia analisi 1PP_filtering

Una volta selezionata l'analisi **1PP_filtering** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico ("metodo") da usare tra quelli disponibili per l'analisi. I tool utilizzabili per questa analisi sono:

- Per tutte le read:
- **Bowtie** - un allineatore di *short reads* ultrarapido ed efficiente in termini di memoria.

Pagina GitHub di Bowtie: https://github.com/BenLangmead/bowtie

- Solo per *long reads* da apparati con tecnologia nanopore:
- **minimap2** - un programma versatile di allineamento che allinea sequenze di DNA o mRNA contro un ampio database di riferimento.

Pagina GitHub di minimap2: https://github.com/lh3/minimap2

Nella sezione dedicata alla selezione dell'input sarà necessario indicare sia le read di input che un genoma reference tra quelli disponibili, che verrà usato per il mapping.

La schermata di selezione input di questa analisi mette a disposizione dell'utente la modalità di selezione input avanzata, in modo da poter selezionare contemporaneamente file FASTQ precedentemente processati in modo diverso.

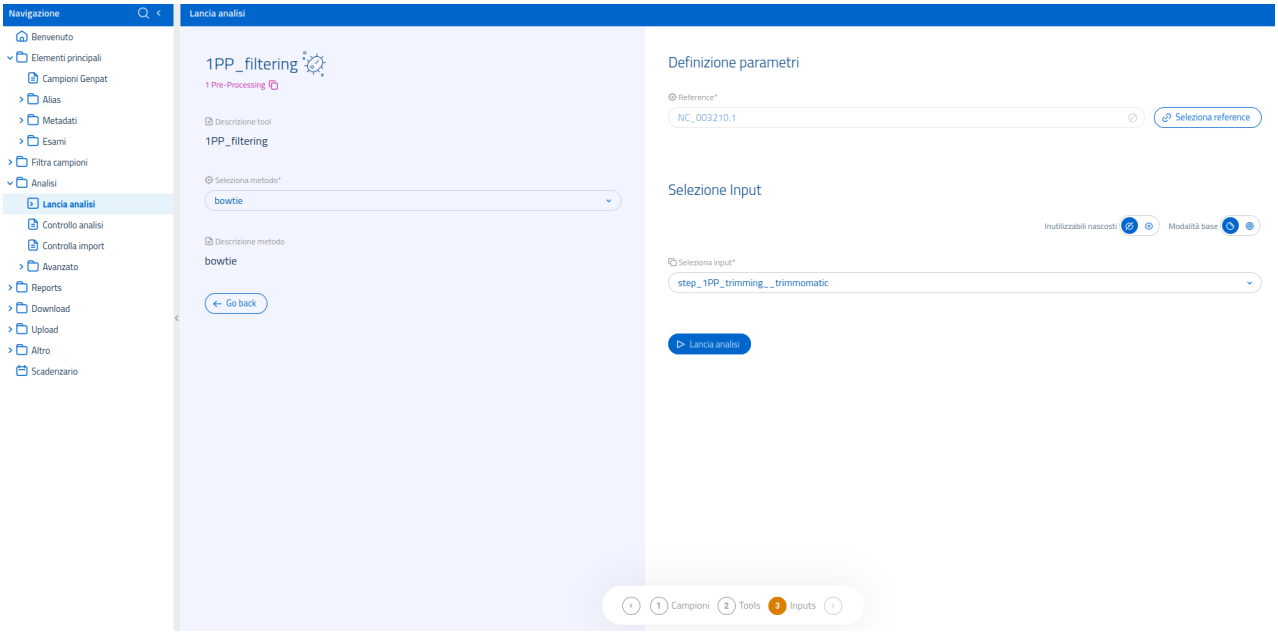
Input per Bowtie:

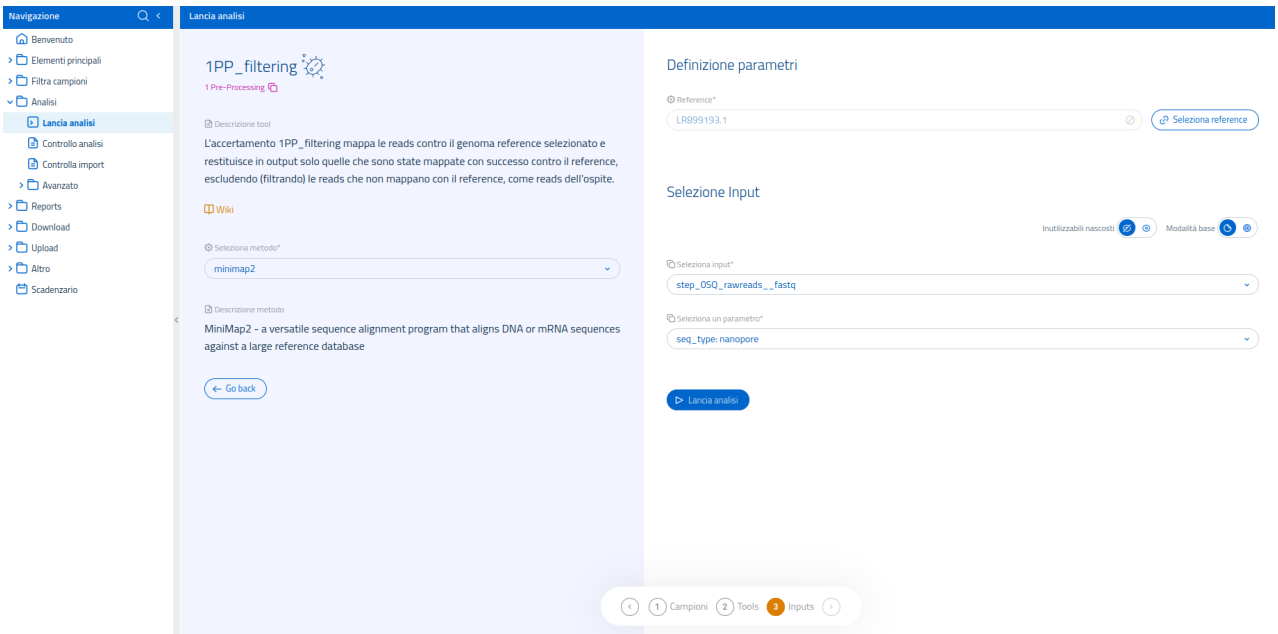
- step_1PP_trimming
- step_1PP_hostdepl
- step_1PP_downsampling
- step_1PP_filtering

Input per minimap2:

- step_0SQ_rawreads

Se il software selezionato è **minimap2**, gli input potranno provenire eclusivamente da **step_0SQ_rawreads** di **single** *long reads* prodotte da dispositivi basati su tecnologia nanopore. Per ulteriori informazioni sul software **minimap2**, si invita a consultare la pagina GitHub ufficiale di minimap2.





Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 2 sotto-cartelle:

- **meta**: (*metadati*) in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

Le tabelle sottostanti presentano la lista di files prodotti dai tools disponibili per 1PP_filtering, insieme ad alcune informazioni utili.

Bowtie:

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_R1_bowtie_REFID.fastq.gz	read 1 (R1) filtrata (file compresso)	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_R2_bowtie_REFID.fastq.gz	read 2 (R2) filtrata (file compresso)	cartella "result"

minimap2:

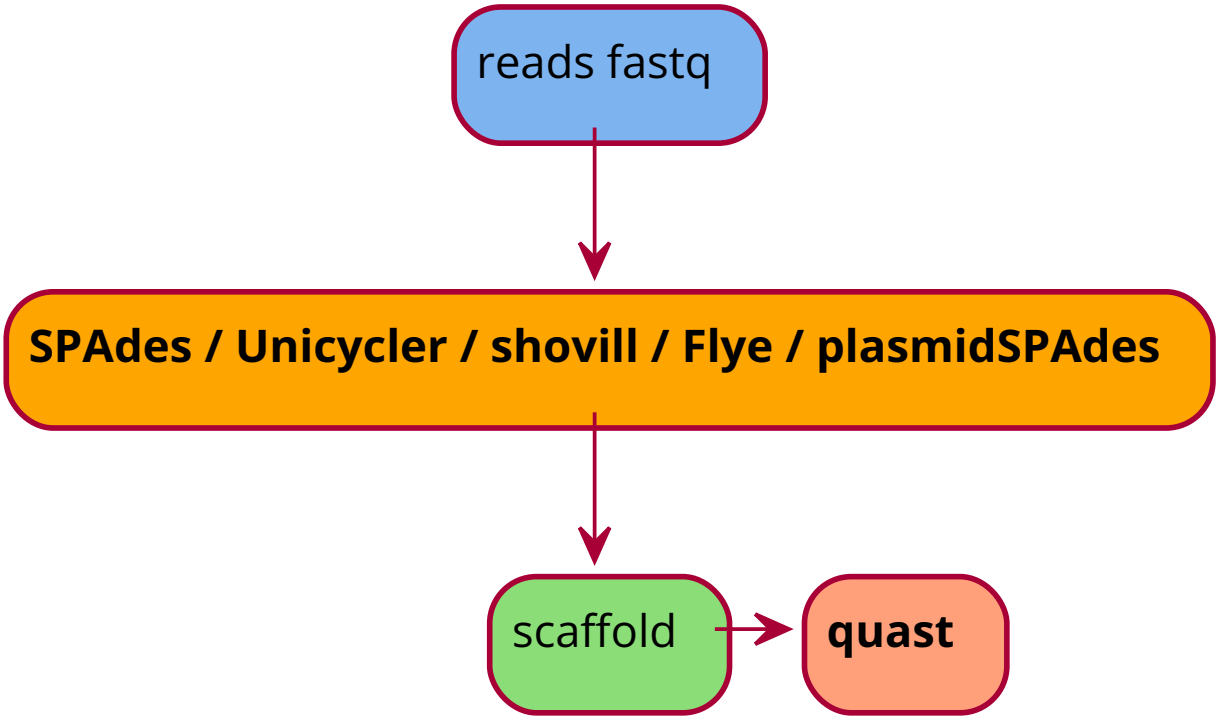
File	Descrizione	Posizione
DSXXXXXX-DTXXXXXX_ID_minimap2_REFID.fastq.gz	fastq filtrato della single read nanopore (file compresso)	cartella "result"

2 Assembly

2AS_denovo

Introduzione

2AS_denovo esegue l'assembly *de novo* delle sequenze, producendo gli scaffolds del genoma.



Tool
Input
Result
Quality Check tool

Lancia analisi 2AS_denovo

Una volta selezionata l'analisi **2AS_denovo** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico ("metodo") da usare tra quelli disponibili per l'analisi. I metodi utilizzabili per questo accertamento sono:

- **SPAdes** - St. Petersburg genome assembler.

  Pagina GitHub di SPAdes: https://github.com/ablab/spades

- **Unicycler** - pipeline di assemblaggio ibrido per genomi batterici.

  Pagina GitHub di Unicycler: https://github.com/rrwick/Unicycler

- **shovill** - pipeline per l'assemblaggio di genomi di isolati batterici provenienti da read paired-end Illumina.

  Pagina GitHub di shovill: https://github.com/tseemann/shovill

- **Flye** - assemblatore *de novo* per read di sequenziamento a singola molecola.

  Pagina GitHub di Flye: https://github.com/fenderglass/Flye

- **plasmidSPAdes** - St. Petersburg genome assembler per l'assemblaggio di plasmidi batterici.

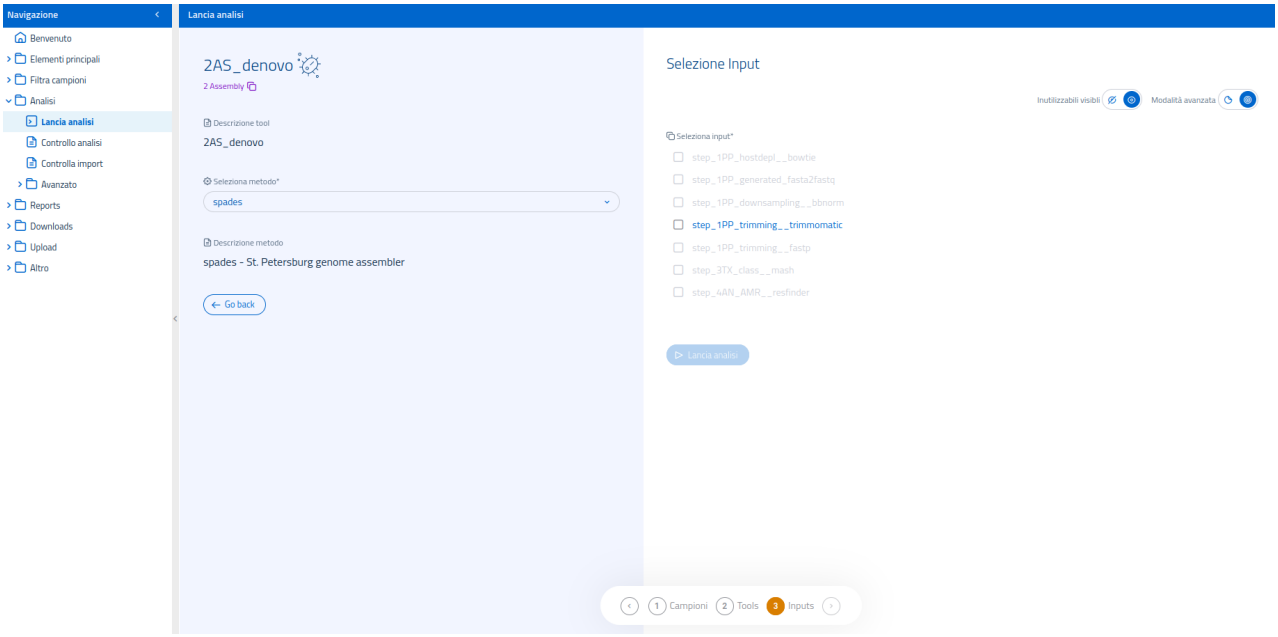
  Pagina GitHub di plasmidSPAdes: https://github.com/ablab/spades

Nota: il tool **Flye** è dedicato esclusivamente all'assemblaggio di *long reads* generate da apparati Oxford Nanopore; il tool **plasmidSPAdes** è un'iterazione di **SPAdes** concepita per l'uso con plasmidi.

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input possono provenire da:

- step_1PP_trimming
- step_1PP_hostdepl
- step_1PP_downsampling
- step_1PP_generated
- step_3TX_class
- step_4AN_AMR



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 3 sotto-cartelle:

- **meta**: (*metadati*) in cui vengono salvati i file di log e di configurazione del processo eseguito; **qc**: (*quality check*) in cui si trovano i file del controllo qualità. In questa cartella sono presenti le cartelle *meta* e *result*. Per questa analisi, il controllo qualità viene eseguito con **quast**;
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

Le tabelle in basso presentano la lista di files presenti nelle cartelle, insieme ad alcune informazioni utili.

SPAdes

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_spades_contigs.fasta	file dei contigs del *de novo* assembly	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_spades_scaffolds.fasta	file degli scaffolds	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_spades_scaffolds_L200.fasta	file contenente gli scaffolds con lunghezza >200bp	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_quast.csv	file delle statistiche di qualità dell'assembly	cartella qc > result

Unicycler

File	Descrizione	Posizione
001_best_spades_graph.gfa	*de novo assembly* in formato .gfa - fase di pulizia 1	cartella "result"
002_overlaps_removed.gfa	*de novo* assembly in formato .gfa - fase di pulizia 2	cartella "result"
003_bridges_applied.gfa	*de novo* assembly in formato .gfa - fase di pulizia 3	cartella "result"
004_final_clean.gfa	*de novo* assembly in formato .gfa - fase di pulizia 4	cartella "result"
005_polished.gfa	*de novo* assembly in formato .gfa - fase di pulizia 5	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_unicycler_assembly.fasta	file dei contigs del *de novo* assembly	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_unicycler_scaffolds_L200.fasta	file contenente gli scaffolds con lunghezza >200bp	cartella "result"
assembly.gfa	*de novo* assembly in formato .gfa	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_quast.csv	file delle statistiche di qualità dell'assembly	cartella qc > result

shovill

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_shovill.fasta	file dei contigs del *de novo* assembly	cartella result
DSXXXXXXXX-DTXXXXXX_ID_quast.csv	file delle statistiche di qualità dell'assembly	cartella qc > result

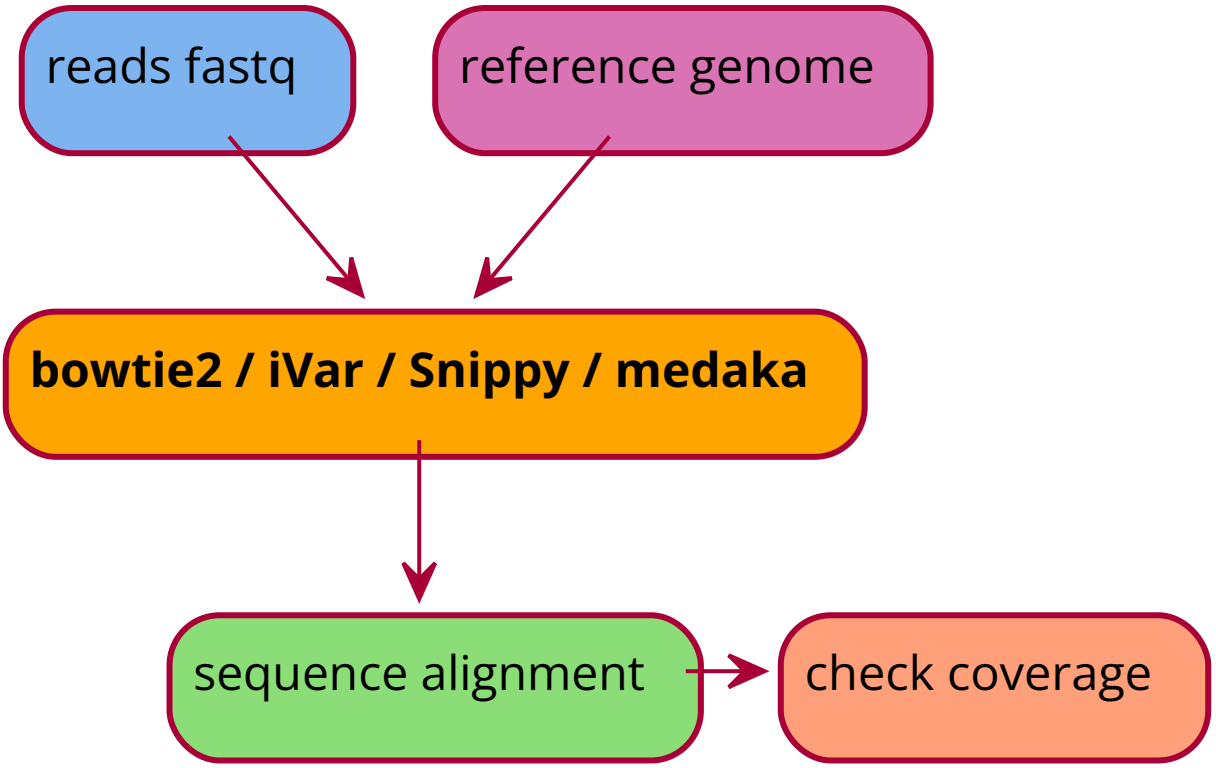
Flye

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_flye.fasta	file dei contigs del *de novo* assembly	cartella result
DSXXXXXXXX-DTXXXXXX_ID_quast.csv	file delle statistiche di qualità dell'assembly	cartella qc > result

2AS_mapping

Introduzione

L'analisi *2AS_mapping* esegue l'**assembly con reference** (o **mapping**) delle sequenze.



Tool
Input
Result
Reference
Quality check

Lancia analisi 2AS_mapping

Una volta selezionata l'analisi **2AS_mapping** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il tool bioinformatico ("metodo") da usare tra quelli disponibili per l'analisi. I tool utilizzabili per questa analisi sono:

- **bowtie2** - uno strumento ultrarapido ed efficiente in termini di memoria per l'allineamento delle read di sequenziamento a lunghe sequenze reference.

  Pagina GitHub di bowtie2: https://github.com/BenLangmead/bowtie2

- **iVar** - pacchetto informatico per il sequenziamento virale basato su ampliconi.

  Pagina GitHub di iVar: https://github.com/andersen-lab/ivar

- **Snippy** - chiamata rapida delle varianti aploidi e allineamento del genoma *core*.

  Pagina GitHub di Snippy: https://github.com/tseemann/snippy

Nota: Nel caso si utilizzi il metodo *snippy*, è indispensabile che il campione scelto abbia l'analisi 4AN_genes associata (Prokka).

- **medaka** - tool per creare sequenze "consensus" e chiamate di varianti, a partire da dati di sequenziamento Nanopore.

  Pagina GitHub di medaka: https://github.com/nanoporetech/medaka

Nota: medaka è un *tool open source* sviluppato dalla Oxford Nanopore Technologies Ltd., da usare esclusivamente su dati di sequenziamento provenienti da apparati Nanopore.

Nella sezione dedicata alla *definizione dei parametri*, sarà necessario selezionare un reference per il mapping, dalla tabella pop-up accessibile cliccando sul tasto **Seleziona** reference. Il pop-up elenca sequenze di due tipologie, entrambe utilizzabili come reference:

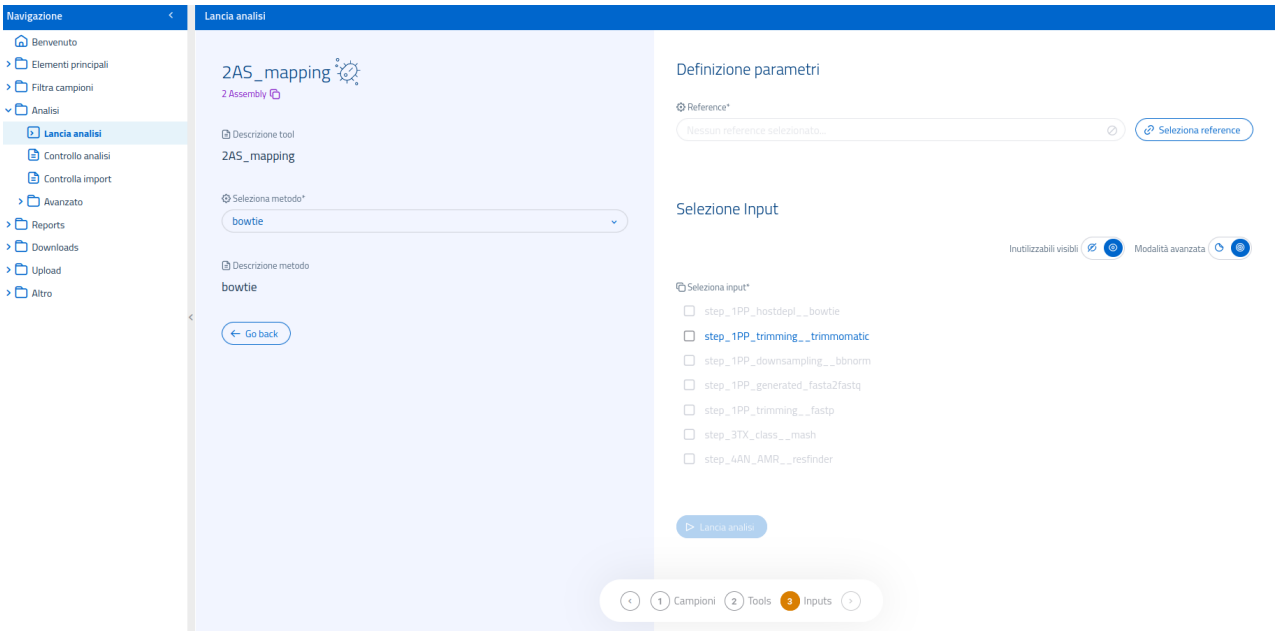
1. **reference fasta files**: file importati da database esterni - come ad es. NCBI - (elencati nella tabella delle reference di GenPat: https://genpat.izs.it/cmdbuild/ui/#classes/T_REFERENCE/cards);
2. la consensus o il *de novo* assembly di **un campione presente in GenPat** (database interni).

Con il metodo "ivar", è possibile selezionare più di un reference (si faccia riferimento alla sezione "Reference multipli", più in basso).

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input sono gli stessi utilizzabili anche per 2AS_denovo:

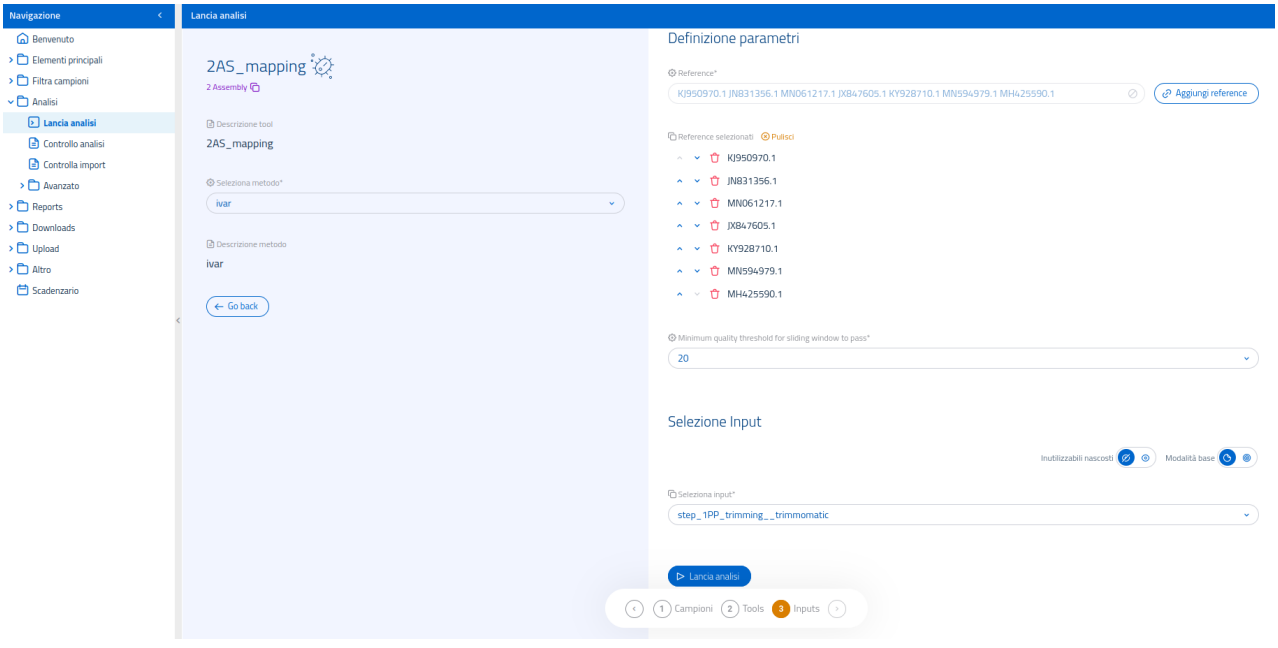
- step_1PP_trimming
- step_1PP_hostdepl
- step_1PP_downsampling
- step_1PP_generated
- step_3TX_class
- step_4AN_AMR



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Reference multipli

Se *2AS_mapping* viene lanciato con **Ivar**, è possibile selezionare più di un reference, come mostrato nell'immagine riportata di seguito. La guida all'uso del sistema di selezione di reference multipli è consultabile nell'apposita sezione della Wiki sul lancio di analisi.



Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 2 sotto-cartelle:

- **meta**: (*metadati*) in cui vengono salvati i file di log e di configurazione del processo eseguito;
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

Le tabelle sottostanti elencano i files prodotti dai tool disponibili per 2AS_mapping.

Bowtie2

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_bowtie_REFID.fasta	file della consensus	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_bowtie_REFID.sorted.bam	file di allineamento in formato .bam (Binary Alignment Map)	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_bowtie_REFID.sorted.bam.bai	file .bai (indice del file BAM)	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_bowtie_REFID.var.flt.vcf	file delle varianti identificate in formato .vcf	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_bowtie_REFID_coverage_plot.png	grafico della distribuzione del coverage	cartella "result"

iVar e Snippy

Si noti che l'uso di **ivar** comprende l'esecuzione dei tool *snippy*, *samtools* e, infine, *ivar*, mentre **2AS_mapping__snippy** esegue esclusivamente il software da cui prende il nome.

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_ivar_REFID.fasta	consensus ottenuta con ivar (**Nota: 2AS_mapping__snippy non produce questo file**)	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.aligned.fa	versione del reference con `-` nelle posizioni con `sequencing depth = 0` (numero di read che allineano in quella posizione = 0) e `N` per `depth` compresa tra 0 ed il numero minimo di *reads* considerate nel coverage dei siti (**non sono elencate varianti nel file**)	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.bam	file di allineamento in formato .bam (Binary Alignment Map)	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.bam.bai	file .bai (indice del file BAM)	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.bed	file .bed (Browser Extensible Data)	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.consensus.fa	genoma reference con rappresentazione di tutte le varianti	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.consensus.subs.fa	genoma reference con rappresentazione delle sole varianti di sostituzione	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.txt	riassunto della *run* di snippy	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.tab	tabella delle varianti in formato .tsv	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.csv	tabella delle varianti in formato .csv	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.filt.vcf	varianti filtrate da Freebayes	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.raw.vcf	varianti non filtrate da Freebayes	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.subs.vcf	tabella delle sole varianti di sostituzione in formato .vcf	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.gff	varianti in formato GFF3	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.html	versione .html della tabella .tab delle varianti	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.vcf	output file di snippy (varianti identificate), in formato .vcf (header di riepilogo + tabella)	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.vcf.gz	output file .vcf compresso di snippy	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID.vcf.gz.csi	indice bcftools del file .vcf.gz	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_vdsnippy_REFID_coverage_plot.png	grafico della distribuzione del coverage	cartella "result"

Per ulteriori informazioni sui files prodotti da *snippy*, i loro formati ed il loro contenuto, si invita a fare riferimento al manuale di snippy.

medaka

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_medaka_REFID.fasta	file fasta dell'allineamento	cartella "result"

Visualizzazione dei dati

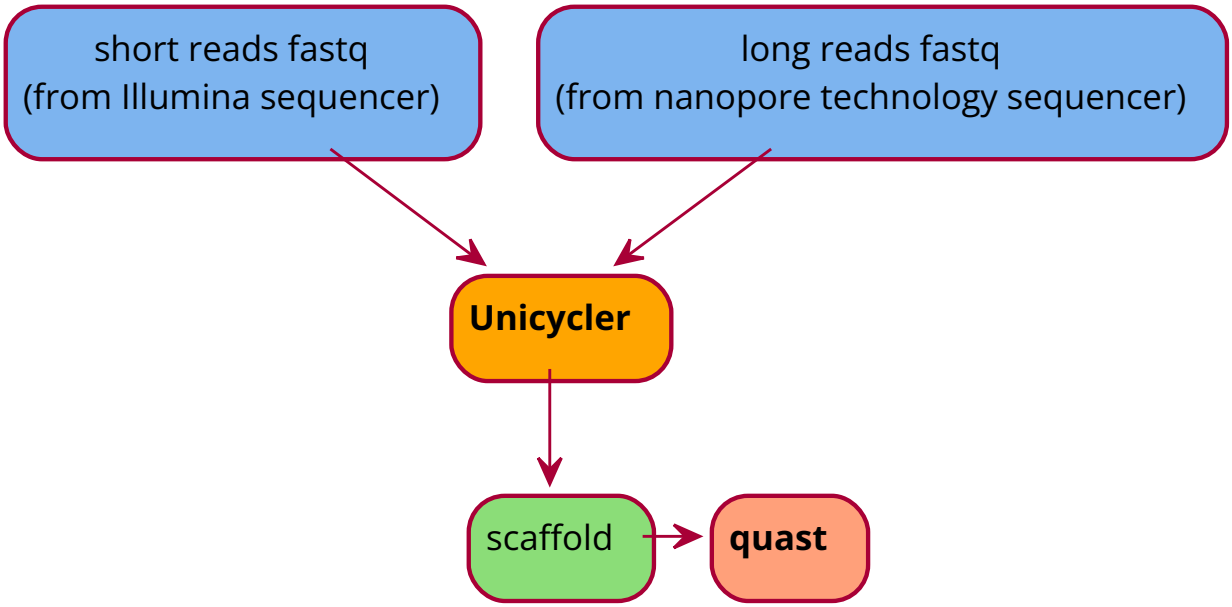
Per valutare la qualità delle sequenze ottenute, l'allineamento delle reads sul reference può essere visualizzato usando software specifici (ad esempio Tablet, BioEdit e uGene) per leggere i file .bam e .bam.bai:

- Tablet (GNU/Linux, macOS, Windows): https://ics.hutton.ac.uk/tablet/download-tablet/;
- BioEdit (Windows): https://thalljiscience.github.io/;
- uGene (GNU/Linux, macOS, Windows): http://ugene.net/ugene/.

2AS_hybrid

Introduzione

2AS_hybrid effettua l'assemblaggio ibrido di *short* e *long reads*. L'assembly ibrido permette di coprire le regioni ripetute del DNA e di collocarle nelle posizioni corrette, grazie alle tecnologie di sequenziamento di terza generazione che producono *long reads*; contemporaneamente, l'assembly ibrido aumenta l'accuratezza, grazie all'utilizzo delle tecnologie di sequenziamento di seconda generazione, che producono *short reads*.



Tool
Input
Result
Quality Check tool

Lancia analisi 2AS_hybrid

Una volta selezionata l'analisi **2AS_hybrid** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico ("metodo") da usare tra quelli disponibili per l'analisi. L'unico metodo disponibile per l'assembly ibrido è:

- **Unicycler** - pipeline di assemblaggio ibrido per genomi batterici.

Pagina GitHub di Unicycler: https://github.com/rrwick/Unicycler

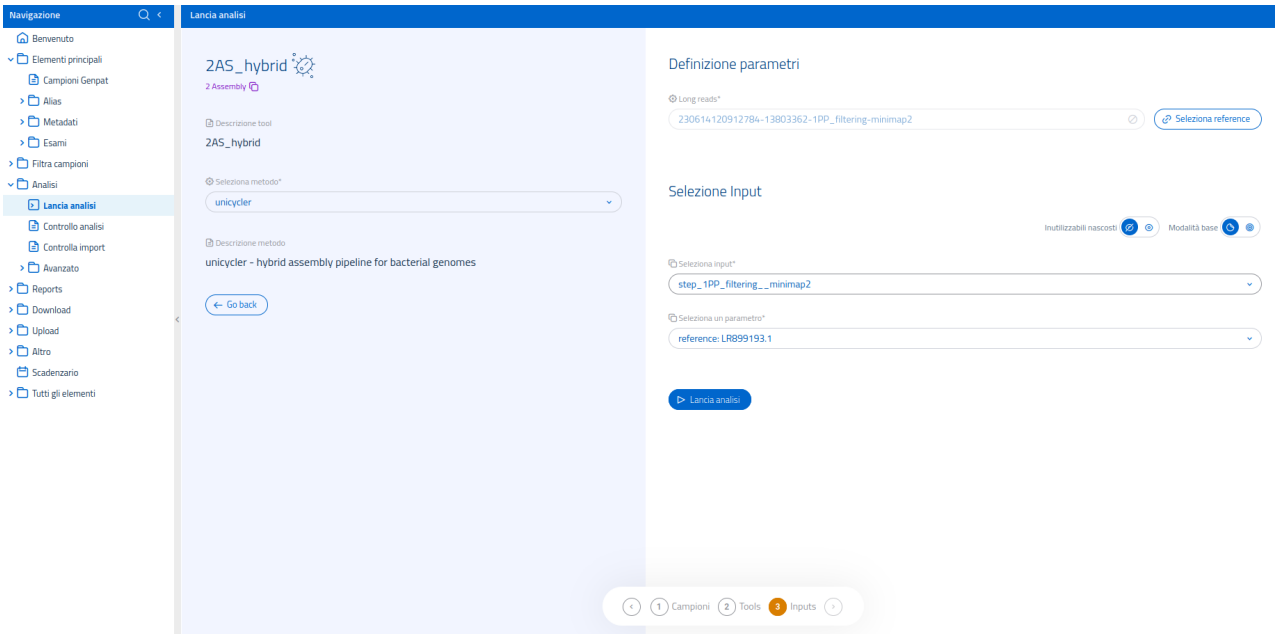
2AS_hybrid richiede 2 input (si faccia riferimento all'immagine in basso):

1. *Long reads* da sequenziatori con tecnologia nanopore (il campo "Long reads" è obbligatorio);
2. *Short reads* da sequenziatori Illumina (il campo "Seleziona input" è obbligatorio). I possibili input per le *short reads* possono provenire dalle analisi di pre-processing delle *reads* Illumina, come:

   step_1PP_trimming
   step_1PP_filtering
   step_1PP_hostdepl
   step_1PP_downsampling.

> **Nota:** l'esecuzione dell'assembly ibrido fallirà se, come secondo input, vengono utilizzate *long reads* al posto delle *short reads* Illumina.

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 3 sotto-cartelle:

- **meta**: (*metadati*) in cui vengono salvati i file di log e di configurazione del processo eseguito; **qc**: (*quality check*) in cui si trovano i file del controllo qualità. In questa cartella sono presenti le cartelle *meta* e *result*. Per questa analisi, il controllo qualità viene eseguito con **quast**;
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

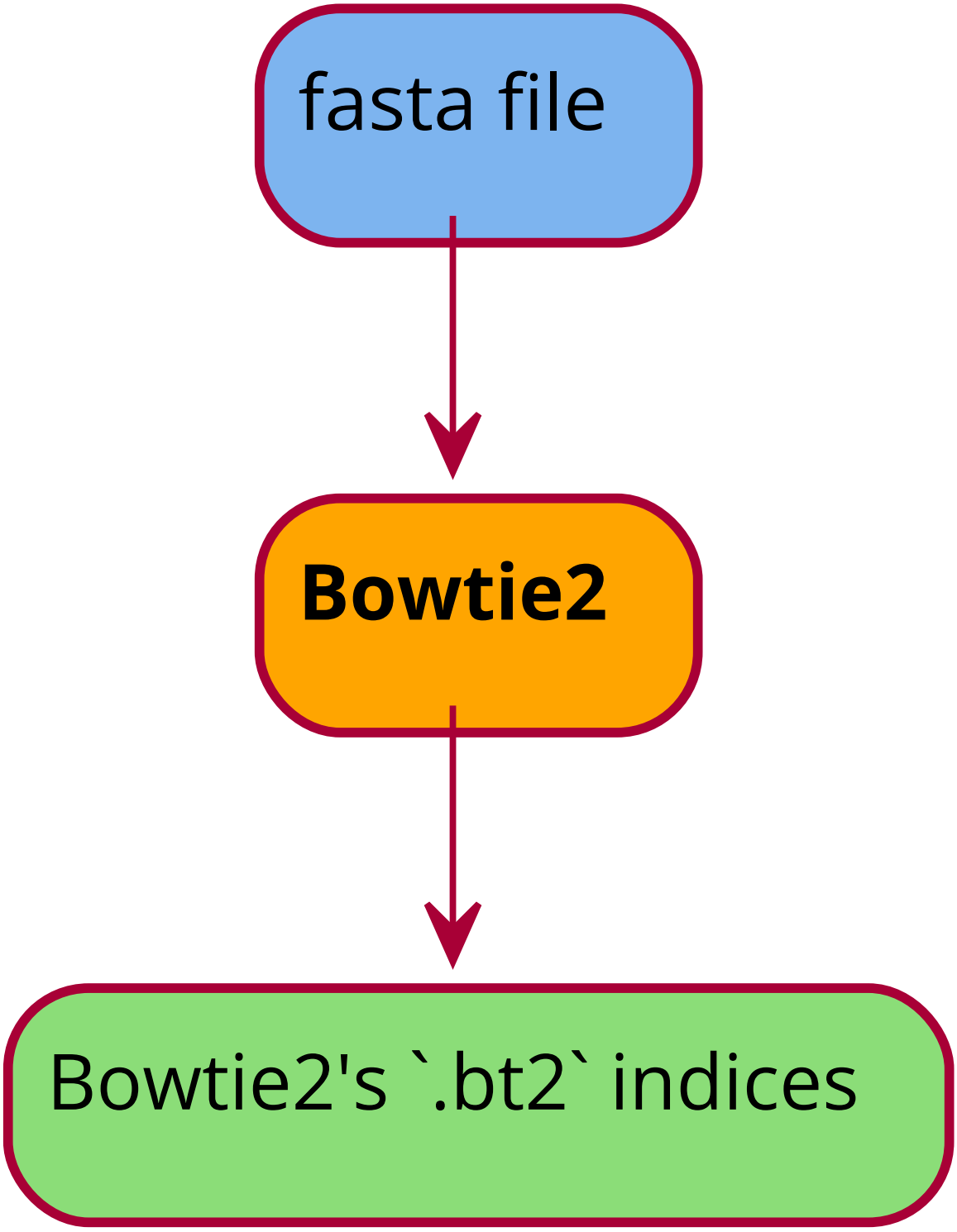
La tabella sottostante presenta una lista dei file presenti nelle cartelle, insieme ad alcune informazioni utili.

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_unicycler.gfa	file dell'assembly in formato .gfa	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_unicycler_assembly.fasta	file dell'assembly ibrido in formato fasta	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID_quast.csv	file delle statistiche di qualità dell'assembly	cartella qc > result

2AS_indexing

Introduzione

L'analisi *2AS_indexing* esegue l'indicizzazione di Bowtie2 su un file FASTA, in modo da generare file di indice `.bt2`. I genomi reference che sono stati indicizzati possono essere selezionati come genomi di organismi ospite per la deplezione nell'analisi deplezione delle sequenze dell'ospite.



Tool

Input

Result

Lancia analisi 2AS_indexing

L'unico software che effettua l'indicizzazione è Bowtie2; gli utenti possono selezionarlo dopo aver scelto l'analisi **2AS_indexing** nell'interfaccia del Lancia Analisi.

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata

L'input per il processo di indicizzazione è un file FASTA; di conseguenza, possono essere usate tutte le principali analisi che producano un file FASTA del genoma:

- step_2AS_import
- step_2AS_mapping
- step_2AS_denovo

Una volta lanciato il software, il sistema genererà un link alla sezione Controllo analisi e l'utente verrà notificato sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

> **Nota:** questa analisi viene usata per produrre i file necessari per la deplezione delle read dell'organismo ospite. Solo gli organismi per i quali questa analisi è stata **lanciata ed importata**, **manualmente o automaticamente a seguito di upload da NCBI**, saranno disponibili come genomi di organismi ospite per la deplezione delle read nell'analisi di *host depletion*.

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 2 sotto-cartelle:

- **meta**: (*metadati*) in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

La tabella che segue elenca i file disponibili nella cartella dei risultati.

Bowtie:

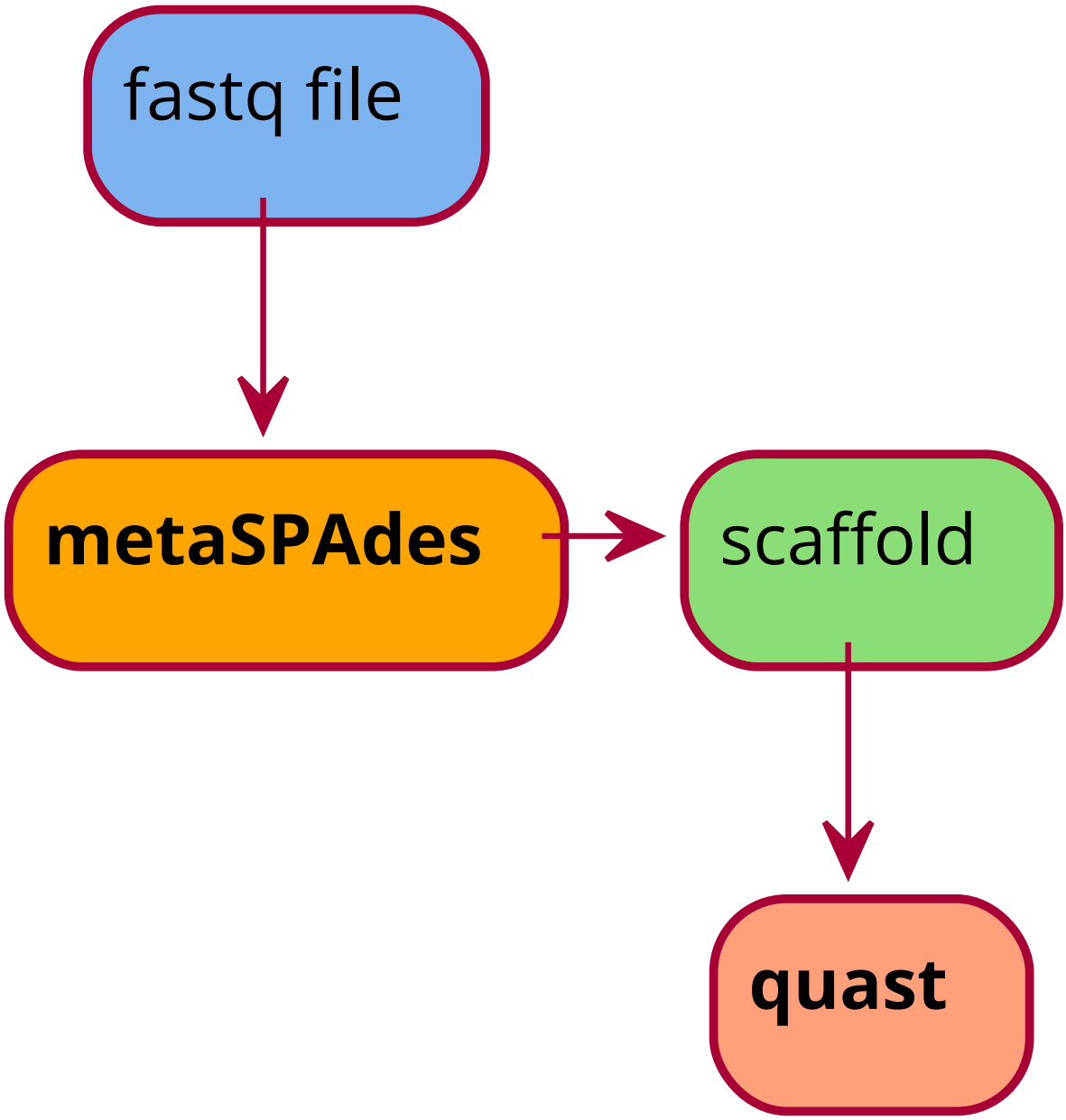
File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID.bt2	file di indice di bowtie2 (`.bt2`)	cartella "result"
DSXXXXXXXX-DTXXXXXX_ID.fasta	file fasta originario sul quale è stato eseguito l'indexing	cartella "result"

2 Metagenomics

2MG_denovo

Introduzione

2MG_denovo effettua il *de novo* assembly per campioni di metagenomica.



Tool

Input

Result

Quality Check tool

Lancia Analisi 2MG_denovo

Una volta selezionata l'analisi **2MG_denovo** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico da usare tra quelli disponibili per l'analisi. Il solo tool utilizzabile per questa analisi è:

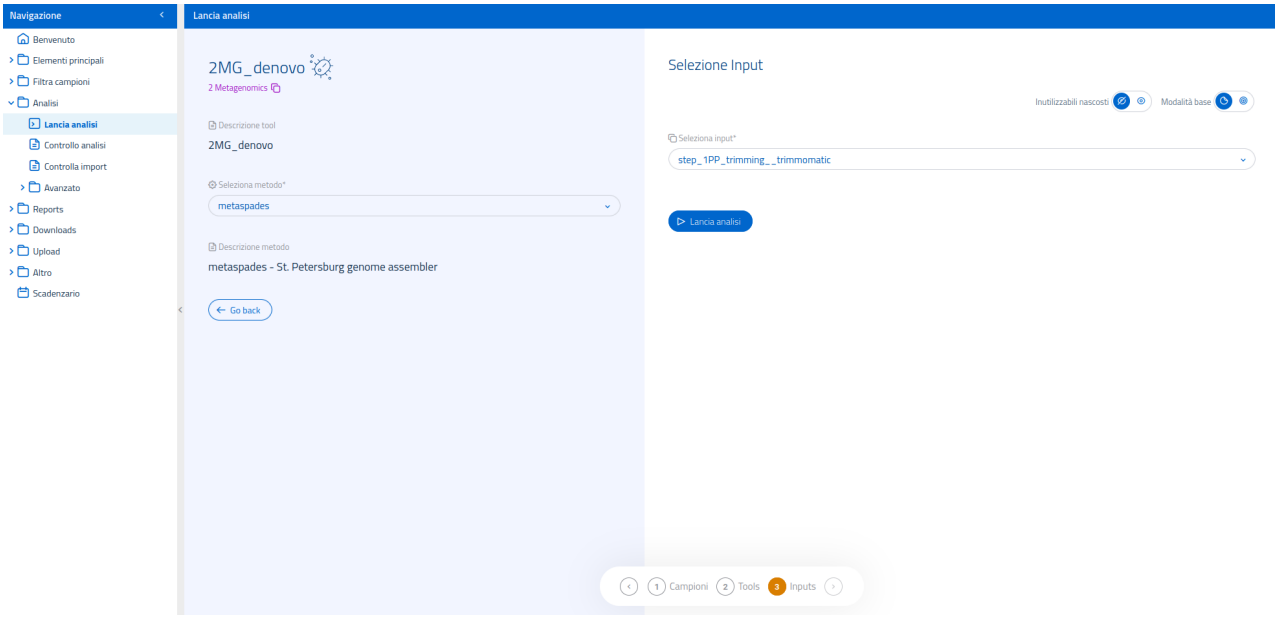
- **metaSPAdes** - St. Petersburg genome assembler

> **Pagina GitHub di metaSPAdes:** https://github.com/ablab/spades

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili sono:

- step_1PP_generated
- step_1PP_trimming
- step_1PP_hostdepl
- step_1PP_downsampling
- step_3TX_class
- step_4AN_AMR



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella verrà visualizzata con la seguente gerarchia: `results > ANNO > ID > 1PP_generated > DSXXXXXXXX-DTXXXXXX_metaspades`. Nella cartella sarà possibile trovare le 2 directories:

- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **qc**: ("quality check") in cui si trovano i files del controllo qualità. A sua volta in questa cartella sono presenti le cartelle meta e result. Per questo accertamento il controllo qualità viene eseguito con **quast**.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

La tabella sottostante elenca i files prodotti da metaspades.

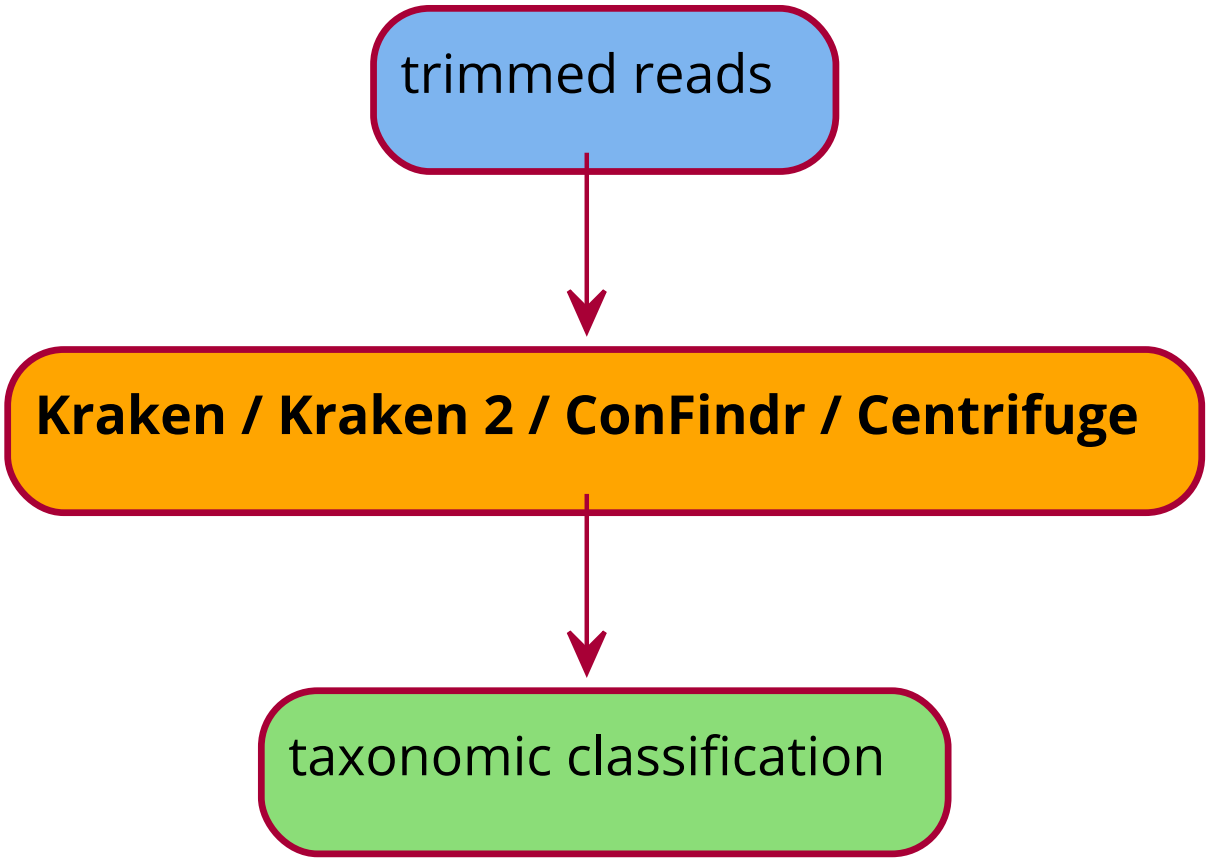
File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_spades_contigs.fasta	file dei contigs del *de novo* assembly	cartella result
DSXXXXXXXX-DTXXXXXX_ID_spades_scaffolds.fasta	file degli scaffolds	cartella result
DSXXXXXXXX-DTXXXXXX_ID_spades_scaffolds_L200.fasta	file contenente gli scaffolds con lunghezza >200bp	cartella result
DSXXXXXXXX-DTXXXXXX_ID_quast.csv	file delle statistiche di qualità dell'assembly	cartella qc > result

3 Taxonomy

3TX_class

Introduzione

L'accertamento 3TX_class esegue la classificazione tassonomica degli organismi d'origine delle reads ed effettua un controllo delle contaminazioni. **3TX_class__kraken** produce appunto i risultati relativi alla contaminazione, consultabili in `Reports > Dettaglio > Check > Check Taxa`.



Tool
Input
Result

Lancia Analisi 3TX_class

Una volta selezionata l'analisi **3TX_class** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico (anche detto "metodo") da usare tra quelli disponibili per l'analisi. I metodi utilizzabili per questo accertamento sono elencati nella lista sottostante.

Per tutti i microrganismi:

- **Kraken** - Kraken taxonomic sequence classification system
- **Kraken 2** - The second version of the Kraken taxonomic sequence classification system

> **Pagina GitHub di Kraken:** https://github.com/DerrickWood/kraken
>
> **Pagina GitHub di Kraken 2:** https://github.com/DerrickWood/kraken2

Solo per i generi *Escherichia*, *Salmonella* e *Listeria*:

- **ConFindr** - designed to find bacterial intra-species contamination in raw Illumina data

> **Pagina GitHub di ConFindr:** https://github.com/OLC-Bioinformatics/ConFindr/tree/main

Solo per single long reads da campioni di **metagenomica** sequenziati con tecnologia nanopore:

- **Centrifuge** - Classifier for metagenomic sequences

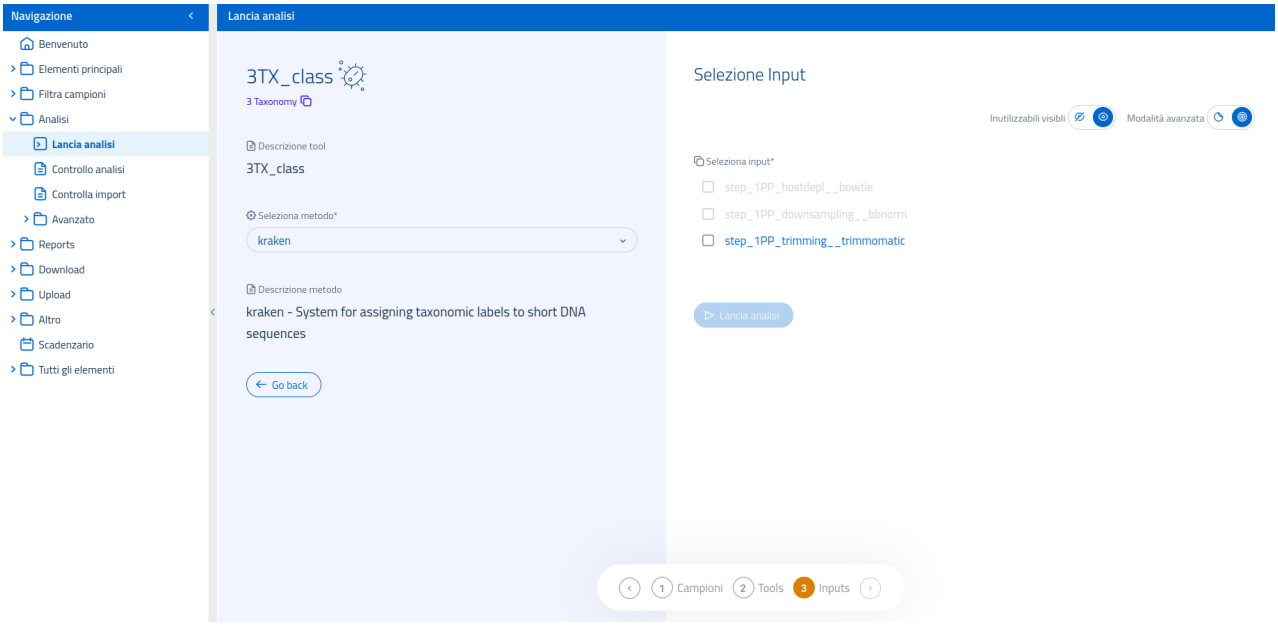
> **Pagina GitHub di ConFindr:** https://github.com/DaehwanKimLab/centrifuge

Nota: Kraken 2 differisce dalla prima versione. Trattandosi di softwares paralleli, sono entrambi disponibili nel sistema di lancio analisi. Per informazioni riguardo il fuzionamento e le differenze tra Kraken e Kraken 2 si invita a consultare il loro manuale ufficiale al link seguente: https://github.com/DerrickWood/kraken2/blob/master/docs/MANUAL.markdown.

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per 3TX_class sono:

- step_1PP_trimming
- step_1PP_hostdepl
- step_1PP_downsampling



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella verrà visualizzata con la seguente gerarchia: `results > ANNO > ID > 3TX_class > DSXXXXXXXX-DTXXXXXX_kraken`, dove "_kraken" può essere sostituito con il nome del tool utilizzato. Nella cartella sarà possibile trovare le 2 directories:

- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

Le tabelle sottostanti elencano i files prodotti dai tool disponibili per 3TX_class. Per maggiori dettagli sui file prodotti dalle due versioni di kraken, si invita a fare riferimento al manuale di kraken e al Kraken2's help sheet.

Kraken

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_kraken.tsv	file di classificazione tassonomica di kraken	cartella result

Kraken 2

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_bracken_genus.tsv	abbondanza stimata delle *reads* a livello di genere di appartenenza rinvenuto	cartella result
DSXXXXXXXX-DTXXXXXX_ID_bracken_genus_report.tsv	report file di classificazione tassonomica a livello di genere	cartella result
DSXXXXXXXX-DTXXXXXX_ID_bracken_species.tsv	abbondanza stimata delle *reads* a livello di specie di appartenenza rinvenuta	cartella result
DSXXXXXXXX-DTXXXXXX_ID_bracken_species_report.tsv	report file di classificazione tassonomica a livello di specie	cartella result
DSXXXXXXXX-DTXXXXXX_ID_kraken.tsv	report file di classificazione tassonomica a livello di ceppo	cartella result
DSXXXXXXXX-DTXXXXXX_ID_kraken.txt.gz	file di testo compresso con le sequenze classificate da kraken2	cartella result

ConFindr

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_contamination.csv	tabella csv delle contaminazioni identificate nel campione da ConFindr	cartella result
DSXXXXXXXX-DTXXXXXX_ID_rmlst.csv	tabella csv degli alleli per ogni gene nel campione	cartella result
DSXXXXXXXX-DTXXXXXX_ID_confindr_report.csv	tabella csv totale dei campioni e delle contaminazioni identificate da ConFindr	cartella result

Centrifuge

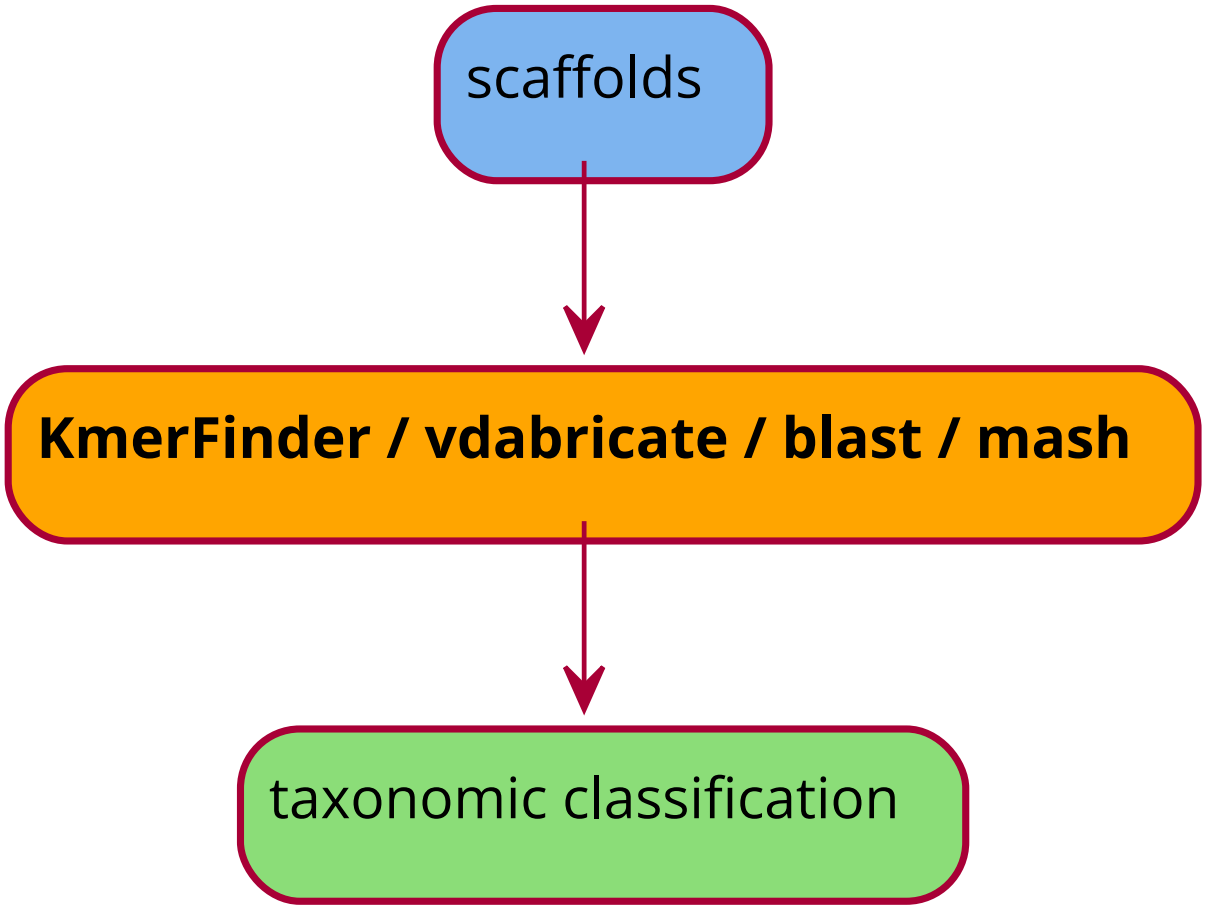
File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_centrifuge.output	file di output tabulare di Centrifuge contenente il *taxa* ID e relativo score e secondo miglior score (aprire con editor di testo o fogli di calcolo)	cartella result
DSXXXXXXXX-DTXXXXXX_ID_centrifuge.report	report tabulare di Centrifuge con la lista degli organismi rinvenuti nel campione e le loro statistiche (aprire con editor di testo o fogli di calcolo)	cartella result

Per ulteriori informazioni sugli output di Centrifuge e sulle caratteristiche del software, si invita a consultare la pagina GitHub ufficiale di Centrifuge.

3TX_species

Introduzione

3TX_species assegna la specie (batterica o virale) più vicina alle sequenze in esame. Se il tool utilizzato è vdabricate, l'accertamento identifica il miglior reference virale da utilizzare per le analisi a valle. **Kmerfinder** è il tool usato per l'identificazione del "best reference" e determina la **specie assegnata**.



Tool
Input
Result

Lancia Analisi 3TX_species

Una volta selezionata l'analisi **3TX_species** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico (anche detto "metodo") da usare tra quelli disponibili per l'analisi. I metodi utilizzabili per questo accertamento sono:

Per l'assegnazione di specie batteriche:

- **KmerFinder** - Prediction of bacterial species using a fast K-mer algorithm

  Pagina GitHub di ConFindr: https://github.com/kcri-tz/kmerfinder

Per l'assegnazione di specie virali:

- **vdabricate** - Mass screening of contigs for viral species of a predefinite database using Abricate

Per tutte le specie:

- **BLAST** - Basic Local Alignment Search Tool: finds regions of similarity between biological sequences

  Pagina GitHub di BLAST: https://github.com/ncbi/blast_plus_docs

- **Mash** - fast genome and metagenome distance estimation using MinHash

  Pagina GitHub di Mash: https://github.com/marbl/Mash

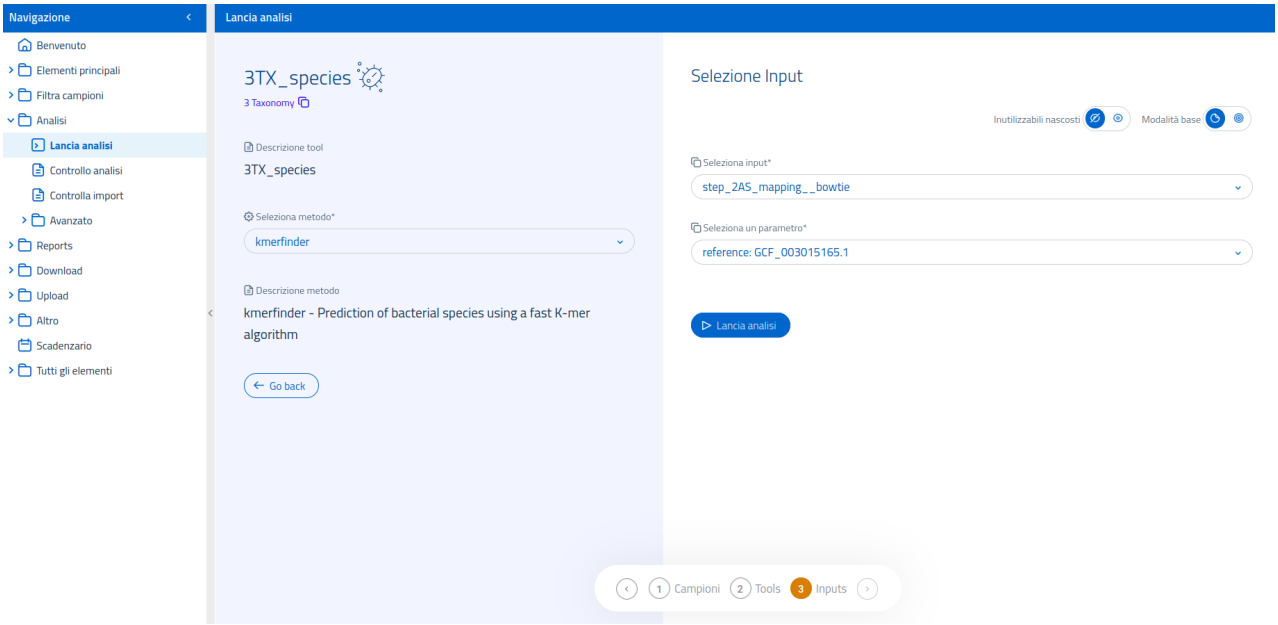
Il tool "vdabricate" può essere utilizzato per l'identificazione del genoma reference migliore per un virus di interesse, in base alle corrispondenze di sequenza trovate nel database delle sequenze virali (DB_Virus).

Per 3TX_species sarà necessario specificare le sequenze di input provenienti da *de novo* assembly o da mapping e il genoma reference usato per il mapping, nel caso venga selezionato quest'ultimo come input.

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per 3TX_species sono:

- step_2AS_mapping
- step_2AS_denovo
- step_2MG_denovo



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella verrà visualizzata con la seguente gerarchia: `results > ANNO > ID > 3TX_species > DSXXXXXXXX-DTXXXXXX_kmerfinder`, dove "_kmerfinder" può essere sostituito con il nome del tool utilizzato. Nella cartella sarà possibile trovare le 2 directories:

- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

Le tabelle sottostanti elencano i files prodotti dai tool disponibili per 3TX_species.

KmerFinder

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_kmerfinder_bacterial.tsv	file con la specie batterica assegnata	cartella results
DSXXXXXXXX-DTXXXXXX_ID_kmerfinder_viral.tsv	file con la specie virale assegnata	cartella results

vdabricate

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_vdabricate_calls.txt	match identificati nel DB_Virus	cartella result

BLAST

File	Descrizione	Posizione
results.csv	tabella riassuntiva dei files dei risultati, con campi separati per ";"	cartella principale
DSXXXXXXXX-DTXXXXXX_ID_blast_nt.asn.gz	dati delle sequenze in formato asn (Abstract Syntax Notation) compresso	cartella result
DSXXXXXXXX-DTXXXXXX_ID_blast_nt.csv	tabella csv dei risultati dell'allineamento di sequenza	cartella result
DSXXXXXXXX-DTXXXXXX_ID_blast_nt.txt	risultati dell'allineamento delle sequenze nell'output standard di BLASTn	cartella result
DSXXXXXXXX-DTXXXXXX_ID_blast_nt_max_target_seqs_5.csv	tabella csv dei risultati dell'allineamento filtrata per la 5 sequenze più vicine	cartella result

Mash

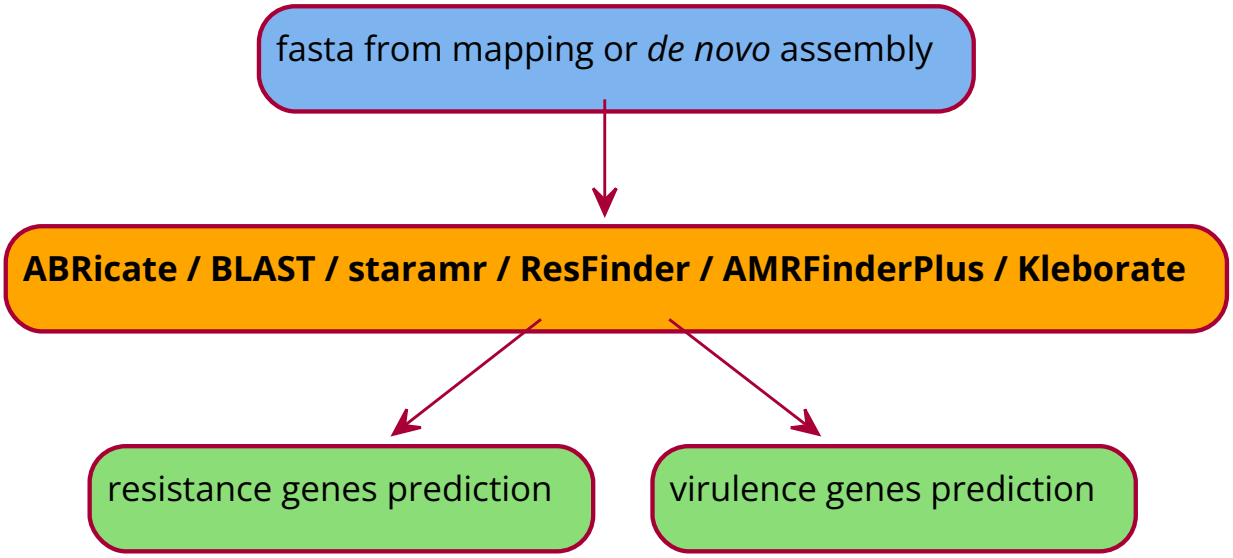
File	Descrizione	Posizione
DS9573634-DT230426121427891_2019.TE.13476.1.1_mash_screen.tsv	tabella intermedia con le percentuali di identità di sequenza tra il genoma del campione e le sequenze testate	cartella result
DS9573634-DT230426121427891_2019.TE.13476.1.1_mash_screening_results.tsv	file tsv con il l'assegnazione finale della specie	cartella result

4 Genome annotation

4AN_AMR

Introduzione

L'accertamento 4AN_AMR (Anti-Microbial Resistance) predice la presenza di geni correlati ad antibiotico-resistenza e di geni di virulenza nella sequenza ricostruita dell'organismo di interesse.



Tool
Input
Result

Lancia Analisi 4AN_AMR

Una volta selezionata l'analisi **4AN_AMR** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico (anche detto "metodo") da usare tra quelli disponibili per l'analisi. In basso sono elencati i metodi utilizzabili per questo accertamento.

Per tutti i microrganismi:

- **ABRicate** - Mass screening of contigs for antimicrobial resistance or virulence genes

  Pagina GitHub di ABRicate: https://github.com/tseemann/abricate

Per tutti i microrganismi:

- **AMRFinderPlus** - AMRFinderPlus identifies acquired antimicrobial resistance genes in bacterial protein and/or assembled nucleotide sequences as well as known resistance-associated point mutations for several taxa

  Pagina GitHub di AMRFinderPlus: https://github.com/ncbi/amr/wiki/Running-AMRFinderPlus#usage

Solo per *Klebsiella Pneumonie* (ma anche K. oxytoca e Escherichia genus):

- **Kleborate** - a tool to screen genome assemblies of Klebsiella pneumoniae and the Klebsiella pneumoniae species complex (KpSC)

  Pagina GitHub di Kleborate: https://kleborate.readthedocs.io/en/latest/Usage.html

Solo per *Listeria monocytogenes*:

- **ResFinder** - identification of acquired antimicrobial resistance genes

  Pagina GitHub di ResFinder: https://github.com/cadms/resfinder

Solo per *Staphylococcus aureus*:

- **BLAST** - Basic Local Alignment Search Tool (blast) finds regions of similarity between biological sequences

  Pagina GitHub di BLAST: https://github.com/ncbi/blast_plus_docs

Solo per *Campylobacter*:

- **staramr** - (*AMR) compiles a summary report of detected antimicrobial resistance genes

  Pagina GitHub di staramr: https://github.com/phac-nml/staramr

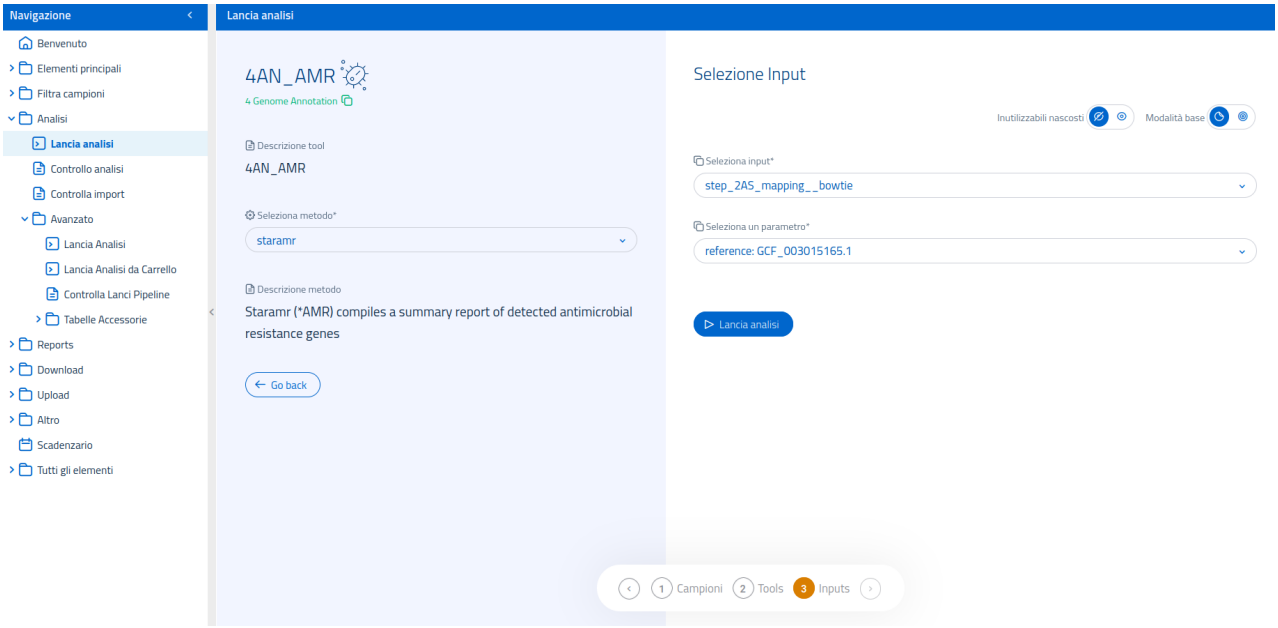
Per 4AN_AMR sarà necessario specificare le sequenze di input provenienti da *de novo* assembly o da mapping e il genoma reference usato per il mapping, nel caso venga selezionato quest'ultimo come input.

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per 4AN_AMR sono:

- step_2AS_mapping
- step_2AS_denovo
- step_2MG_denovo
- step_2AS_import

In aggiunta ai tools per l'indagine di *Anti-Microbial Resistance*, tra i softwares disponibili è elencato anche "**Filtering**", che **non esegue l'AMR**, bensì filtra i risultati dell'AMR (eseguita con uno degli altri programmi disponibili) in base ai parametri di coverage e di identità specificati. Gli input utilizzabili sono step_3TX_species__vdabricate e step_4AN_AMR__abricate.



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella verrà visualizzata con la seguente gerarchia: `results > ANNO > ID > 4AN_AMR > DSXXXXXXXX-DTXXXXXX_abricate`, dove "_abricate" può essere sostituito con il nome del tool utilizzato. Nella cartella sarà possibile trovare le 2 directories:

- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

Le tabelle sottostanti elencano i files prodotti dai tools disponibili per 4AN_AMR.

ABRicate

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_abricate_calls.txt	file con risultati dell'allineamento geni, coverage e chiamate dei database	cartella results
DSXXXXXXXX-DTXXXXXX_ID_abricate.summary	file riassuntivo di abricate con gene e relativo coverage	cartella results
DSXXXXXXXX-DTXXXXXX_ID_output_abricate_AMR.csv	file riassuntivo delle antibiotico-resistenze	cartella results
DSXXXXXXXX-DTXXXXXX_ID_output_abricate_VF.csv	file riassuntivo dei geni di virulenza	cartella results

Per ogni campione nel file .summary, può essere presente un numero riportato nella colonna dei geni. Tale numero indica che il gene corrispondente è presente con un *coverage* orizzontale pari al valore riportato. Nel caso di geni frammentati, per lo stesso gene saranno elencati più numeri separati da ";", corrispondenti ai diversi contigs del *de novo* assembly che hanno prodotto un match per quel gene.

Per ulteriori dettagli riguardo i files prodotti da abricate, si invita a consultare il manuale di ABRicate.

BLAST

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_blast_out.tsv	tabella tsv dei risultati di BLAST	cartella results

staramr

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_staramr_detailed_summary.tsv	tabella tsv dei risultati organizzati per *data type* (gene di resistenza/plasmite/schema MLST...)	cartella results
DSXXXXXXXX-DTXXXXXX_ID_staramr_mlst.tsv	risultati dello schema MLST di 7 alleli per i geni nei *loci* specifici della specie + Sequence Type (ST)	cartella results
DSXXXXXXXX-DTXXXXXX_ID_staramr_plasmidfinder.tsv	risultati tsv con le sequenze appartenenti a plasmidi identificati	cartella results
DSXXXXXXXX-DTXXXXXX_ID_staramr_pointfinder.tsv	risultati dell'AMR in formato tsv con posizione Pointfinder	cartella results
DSXXXXXXXX-DTXXXXXX_ID_staramr_resfinder.tsv	risultati della predizione di resistenze e dell'allineamento geni di resistenza	cartella results
DSXXXXXXXX-DTXXXXXX_ID_staramr_results.xlsx	risultati dell'AMR in formato foglio di calcolo	cartella results
DSXXXXXXXX-DTXXXXXX_ID_staramr_summary.tsv	risultati dell'AMR in formato tsv	cartella results

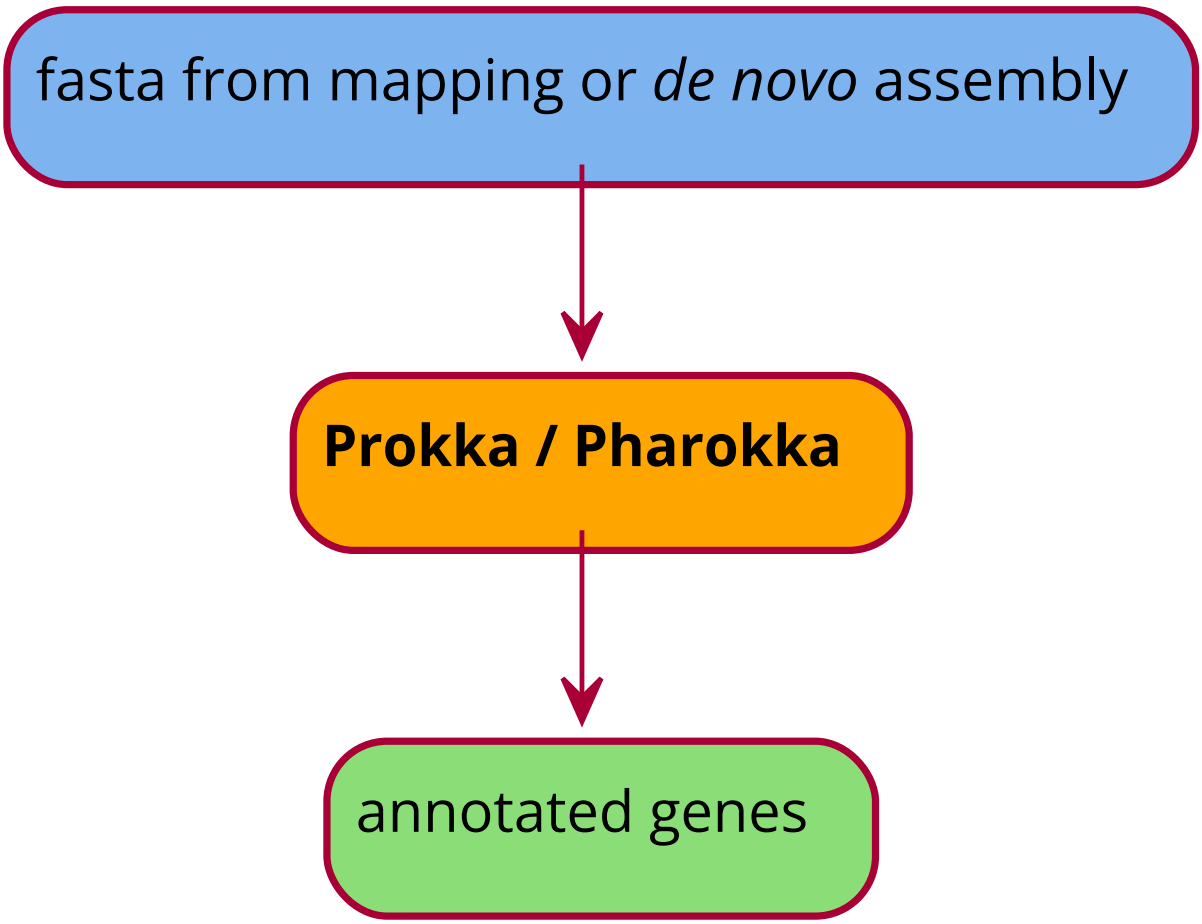
ResFinder

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_resfinder_ResFinder_Hit_in_genome_seq.fsa	file fasta con la sequenza del/i gene/i di resistenza identificati nel genoma del campione con header esteso	cartella results
DSXXXXXXXX-DTXXXXXX_ID_resfinder_ResFinder_Resistance_gene_seq.fsa	file fasta con la sequenza del/i gene/i di resistenza identificati nel genoma del campione con header ridotto	cartella results
DSXXXXXXXX-DTXXXXXX_ID_resfinder_ResFinder_results.txt	file di testo con tabella riassuntiva ed identità di sequenza	cartella results
DSXXXXXXXX-DTXXXXXX_ID_resfinder_ResFinder_results_tab.txt	files tsv con l'elenco dei geni di resistenza identificati (elenca solo i geni presenti)	cartella results
DSXXXXXXXX-DTXXXXXX_ID_resfinder_ResFinder_results_table.txt	files di testo con l'elenco dei geni di resistenza (specifica sia i geni presenti che quelli non rinvenuti)	cartella results
DSXXXXXXXX-DTXXXXXX_ID_resfinder_pheno_table.txt	file di testo con i risultati e tabella dei fenotipi	cartella results
DSXXXXXXXX-DTXXXXXX_ID_resfinder_std_format_under_development.json		cartella results

4AN_genes

Introduzione

L'accertamento 4AN_genes esegue l'annotazione funzionale del genoma (*genome annotation*), identificando le possibili proteine codificate e gli ORF.



Tool
Input
Result

Lancia Analisi 4AN_genes

Una volta selezionata l'analisi **4AN_genes** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico (anche detto "metodo") da usare tra quelli disponibili per l'analisi. I tool utilizzabili per questa analisi sono 2:

- **Prokka** - Tool to annotate bacterial, archaeal and viral genomes
- **Pharokka** - Annotation tool for bacteriophage genomes and metagenomes

> **Pagina GitHub di Prokka:** https://github.com/tseemann/prokka
>
> **Pagina GitHub di Pharokka:** https://github.com/gbouras13/pharokka

Tool Prokka

Il tool Prokka viene usato genericamente per l'annotazione dei genomi di batteri, virus e archea. I suoi parametri prevedono:

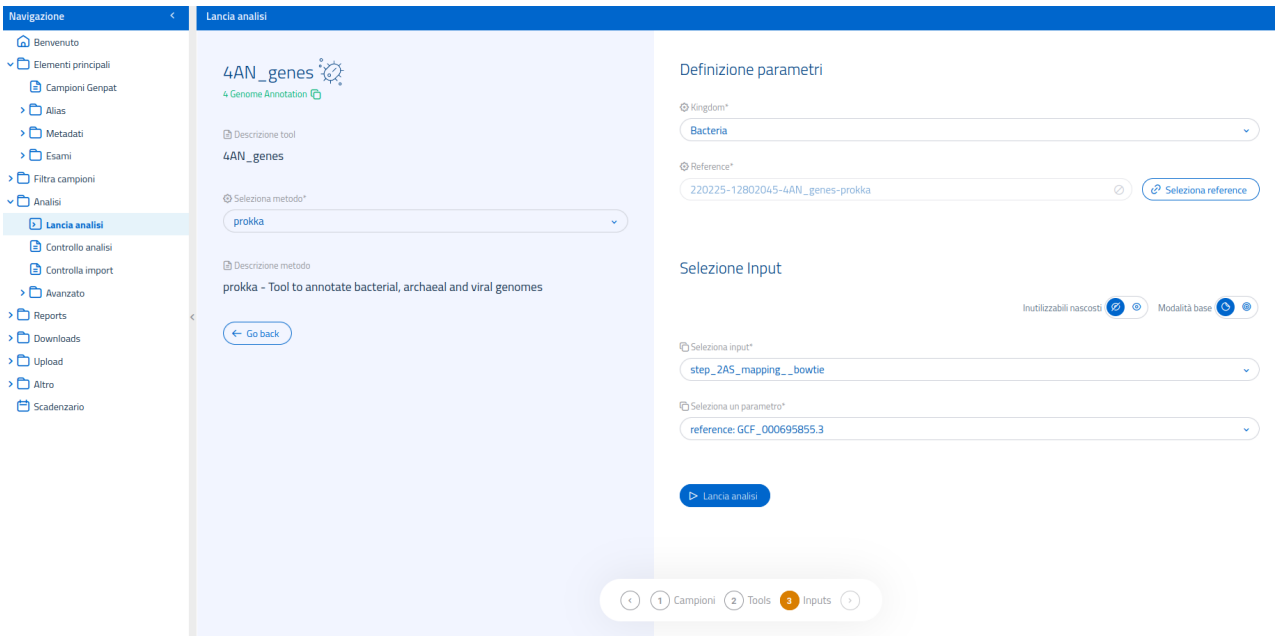
- la selezione di un **regno** tra:
  - virus;
  - batteri;
  - archea;
  - e in aggiunta mitocondri;
- la selezione di un genoma **reference**;
- la selezione dell'analisi di **input**.

Per Prokka è inoltre disponibile la modalità di selezione input avanzata.

I possibili input utilizzabili per 4AN_genes__prokka sono:

- step_2AS_mapping
- step_2AS_denovo
- step_2AS_hybrid
- step_2AS_import

Sarà necessario specificare le sequenze di input provenienti da *de novo* assembly o da mapping e il genoma reference usato per il mapping, nel caso venga selezionato quest'ultimo come input.



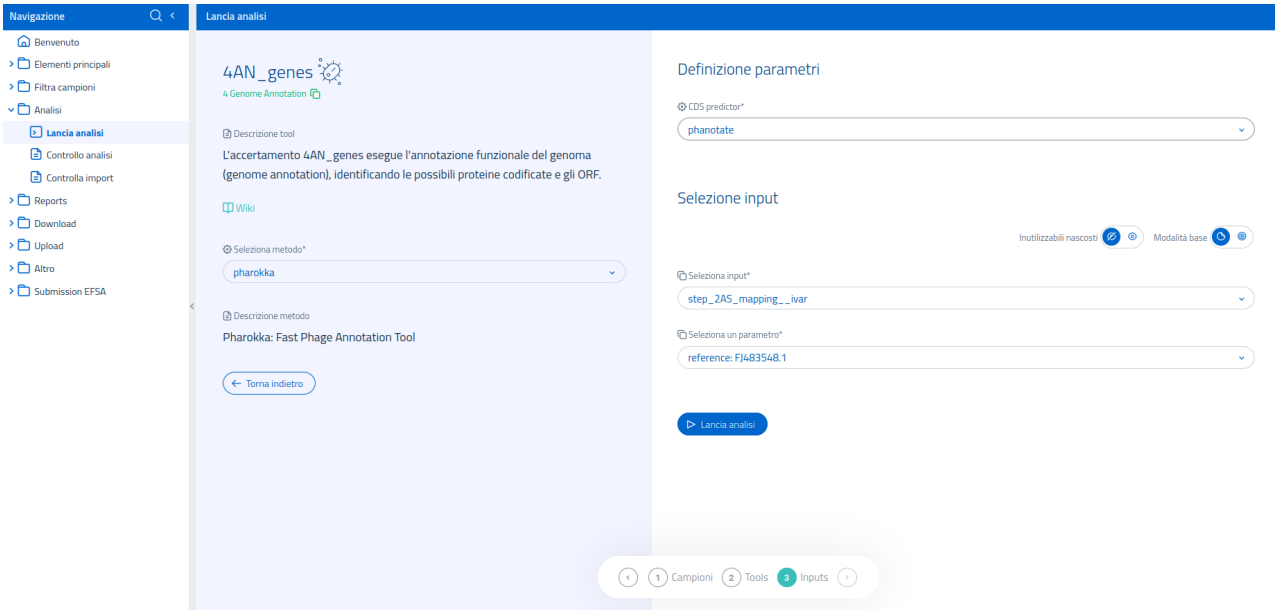
Tool Pharokka

Pharokka è un tool sviluppato **appositamente ed esclusivamente** per l'annotazione dei genomi di batteriofagi. I parametri richiesti sono solo 2:

- il software da usare per la **predizione dei CDS** (sono disponibili i tools **Phanotate** e **Prodigal**; Phanotate è lo strumento di *default*, in quanto specifico per la predizione sui pù piccoli genomi dei fagi);
- la selezione dell'analisi di **input**.

Anche per Pharokka è disponibile la modalità di selezione input avanzata.

- step_2AS_mapping
- step_2AS_denovo
- step_2AS_hybrid



Anche in questo caso sarà necessario specificare il genoma reference usato per il mapping, nel caso venga selezionato quest'ultimo come input dell'analisi.

Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella verrà visualizzata con la seguente gerarchia: `results > ANNO > ID > 4AN_genes > DSXXXXXXXX-DTXXXXXX_prokka`, dove il suffisso dipende dal software selezionato per l'analisi. Nella cartella sarà possibile trovare le 2 directories:

- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

Prokka

La tabella sottostante elenca i files prodotti da Prokka.

File	Descrizione	Posizione
log_errore_controlli_esami.log	log dei messaggi di avvertimento ed errore della run	cartella principale
metadata_samples.tsv	tabella riassuntiva in formato tsv dei metadati dei campioni	cartella principale
results.csv	tabella riassuntiva separata da ";" con informazioni e ID dei campioni	cartella principale
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.err	report delle incongruenze ed errori incontrati durante la run (file di testo)	cartella results
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.faa	sequenze amminoacidiche della traduzione dei geni codificanti identificati (formato faa - fasta aminoacid)	cartella results
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.ffn	sequenze nucleotidiche dei geni codificanti identificati (formato fnn - fasta nucleotide)	cartella results
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.fna	sequenze nucleotidica dei geni codificanti identificati (formato fna)	cartella results
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.fsa	sequenze in formato fsa (fragment analysis data file)	cartella results
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.gbk	file di output in formato GenBank	cartella results
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.gff	file di output in formato gff (General Feature Format)	cartella results
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.log	log della run di prokka	cartella results
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.sqn	file per la sottomissione a GenBank in formato Sequin	cartella results
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.tbl	file di testo con le informazioni di sequenza e dei *loci*	cartella results
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.tsv	file di output tsv che elenca i *loci* e i prodotti proteici dei geni codificanti mappati	cartella results
DSXXXXXXXX-DTXXXXXX_ID_prokka_REFID_result.txt	statistiche sui CDS identificati	cartella results
proteins.faa	sequenze delle proteine identificate in formato faa	cartella results

Per ulteriori dettagli riguardo i files prodotti da Prokka, si invita a consultare il manuale di Prokka.

Pharokka

La tabella sottostante elenca i files prodotti da Prokka.

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_pharokka_REFID.gbk	file genebank in formato gbk con le caratteristiche della sequenza	cartella results
DSXXXXXXXX-DTXXXXXX_ID_pharokka_REFID.gff	file genebank in formato gff con le caratteristiche della sequenza	cartella results
DSXXXXXXXX-DTXXXXXX_ID_pharokka_REFID_cds_final_merged_output.tsv	file tabulare in formato tsv contenente l'annotazione del genoma	cartella results

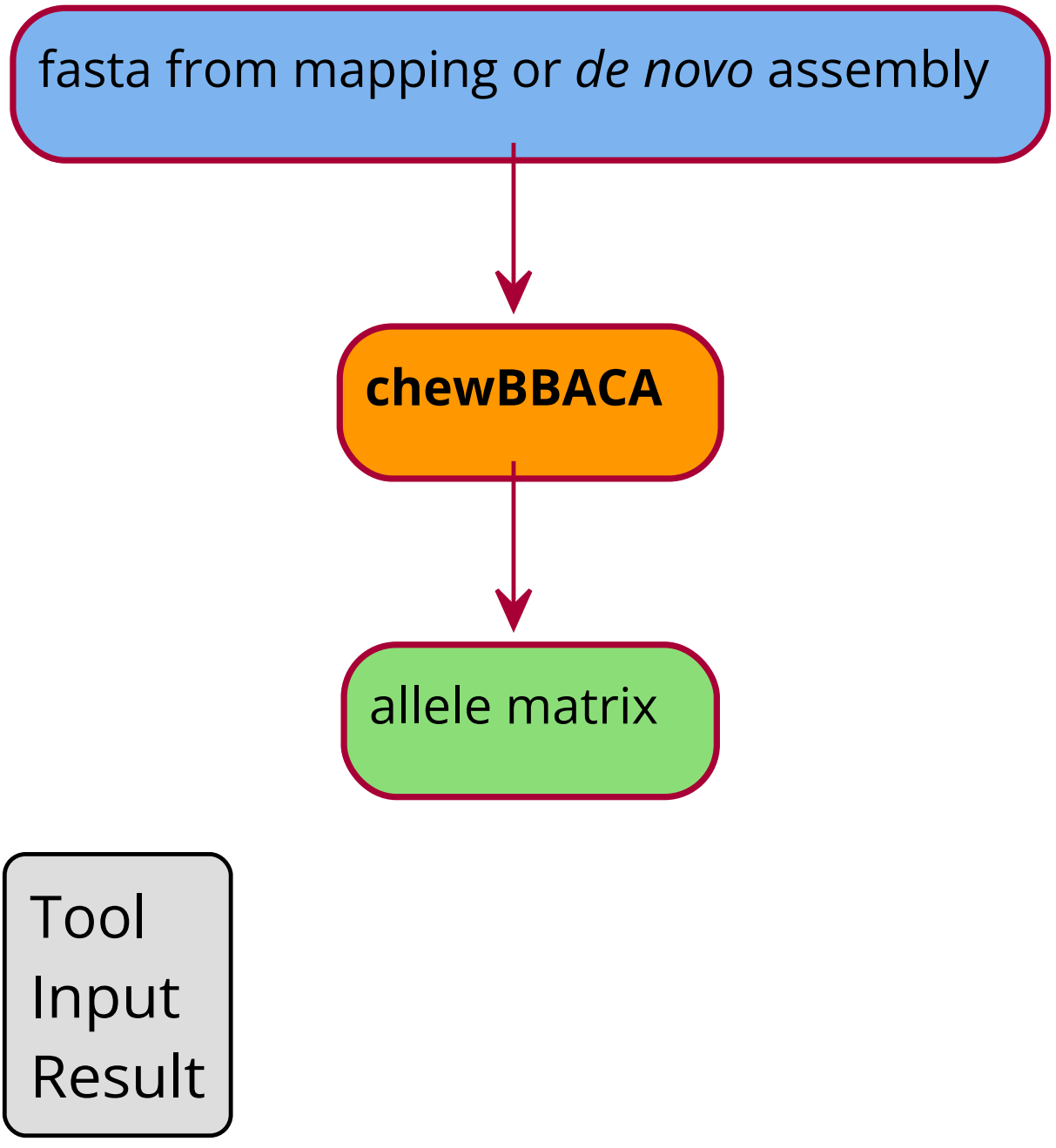
Per ulteriori dettagli riguardo i files prodotti da Pharokka, si invita a consultare il manuale di Pharokka.

4 in silico typing

4TY_cgMLST

Introduzione

4TY_cgMLST esegue il "core genome Multi-Locus Sequence Typing" (cgMLST), un protocollo di caratterizzazione degli isolati batterici, che permette l'identificazione dei cloni all'interno di popolazioni microbiche.



Lancia Analisi 4TY_cgMLST

Una volta selezionata l'analisi **4TY_cgMLST** nella pagina dedicata al lancio di analisi, sarà possibile procedere all'analisi con il tool **chewBBACA** - BSR-Based Allele Calling Algorithm.

Pagina GitHub di chewBBACA: https://github.com/B-UMMI/chewBBACA

Nella tabella sottostante sono riportati gli schemi disponibili per i microrganismi supportati.

Specie	Nome shema	Numero di *loci*	Data ultimo aggiornamento
Mycobacterium bovis	`m_bovis_2891_230720`	2891	20/07/23
Listeria monocytogenes	`l_mono_chewie_1748_220623`	1748	23/06/22
Campylobacter jejuni	`c_jejuni_678_180728`	678	23/07/18
Campylobacter coli	`c_coli_528_220722`	528	22/07/22
Staphylococcus aureus	`s_aureus_ridom_1861_210617`	1861	17/06/21
Brucella melitensis	`b_melitensis_bme_2584_211118`	2584	18/11/21
Klebsiella pneumoniae	`k_pneumoniae_pasteur_629_211129`	629	29/11/21
Escherichia coli	`e_coli_chewie_2360_210531`	2360	31/05/21
Escherichia coli	`e_coli_chewie_7601_210531`	7601	31/05/21
Salmonella enterica	`s_enterica_chewie_3255_210531`	3255	31/05/21
Brucella	`b_wide_1764_220709`	1764	09/07/22
Bacillus anthracis	`b_anthracis_pubmlst_3803_231214`	3803	14/12/23
Acinetobacter baumannii	`a_baumannii_ridom_2390_250612`	2390	12/06/25
Pseudomonas aeruginosa	`p_aeruginosa_ridom_3867_250612`	3867	12/06/25

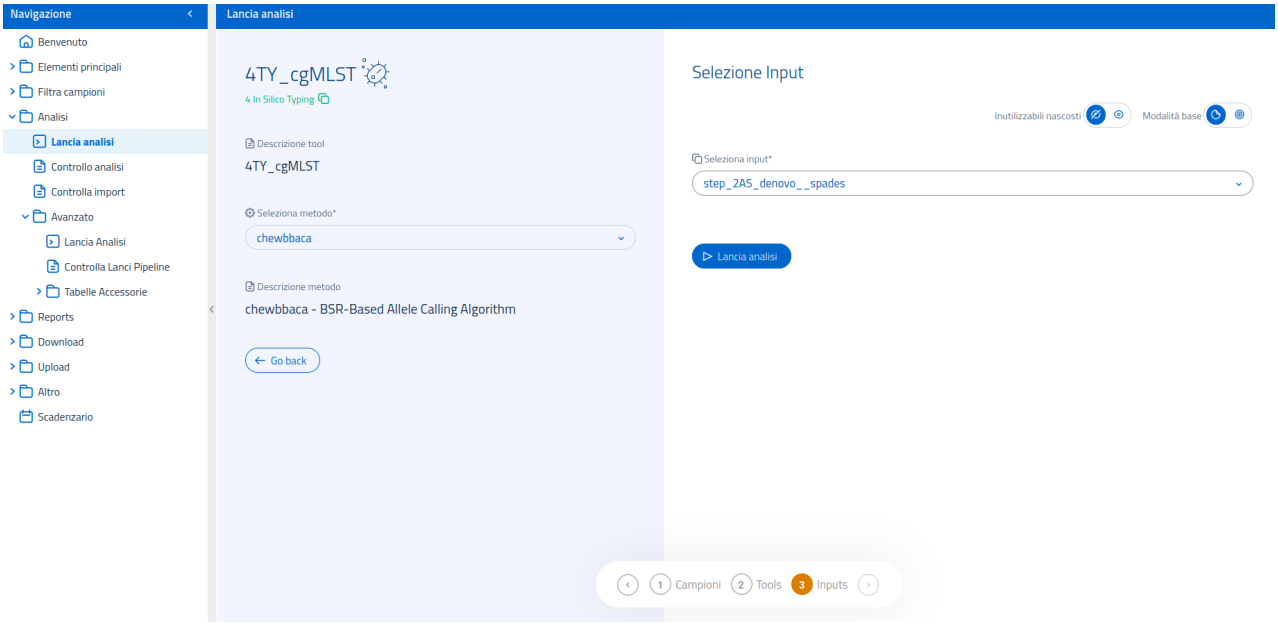
Nota: Il lancio dell'analisi 4TY_cgMLST su un microrganismo per cui non sia disponibile uno schema cgMLST comporterà il fallimento della run.

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per 4TY_cgMLST sono:

- step_2MG_denovo
- step_2AS_denovo
- step_2AS_mapping
- step_2AS_import

Sarà necessario specificare le sequenze di input provenienti da *de novo* assembly o da mapping e il genoma reference usato per il mapping, nel caso venga selezionato quest'ultimo come input.



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella verrà visualizzata con la seguente gerarchia: `results > ANNO > ID > 4TY_cgMLST > DSXXXXXXXX-DTXXXXXX_chewbbaca`, dove il suffisso "_chewbbaca" può essere sostituito con il nome del software utilizzato per l'accertamento. Nella cartella sarà possibile trovare 2 o 3 directories (in base al software utilizzato):

- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

Per la chiamata allelica con chewBBACA i risultati verranno forniti con codifica crc32 (link a crc32 su Wikipedia).

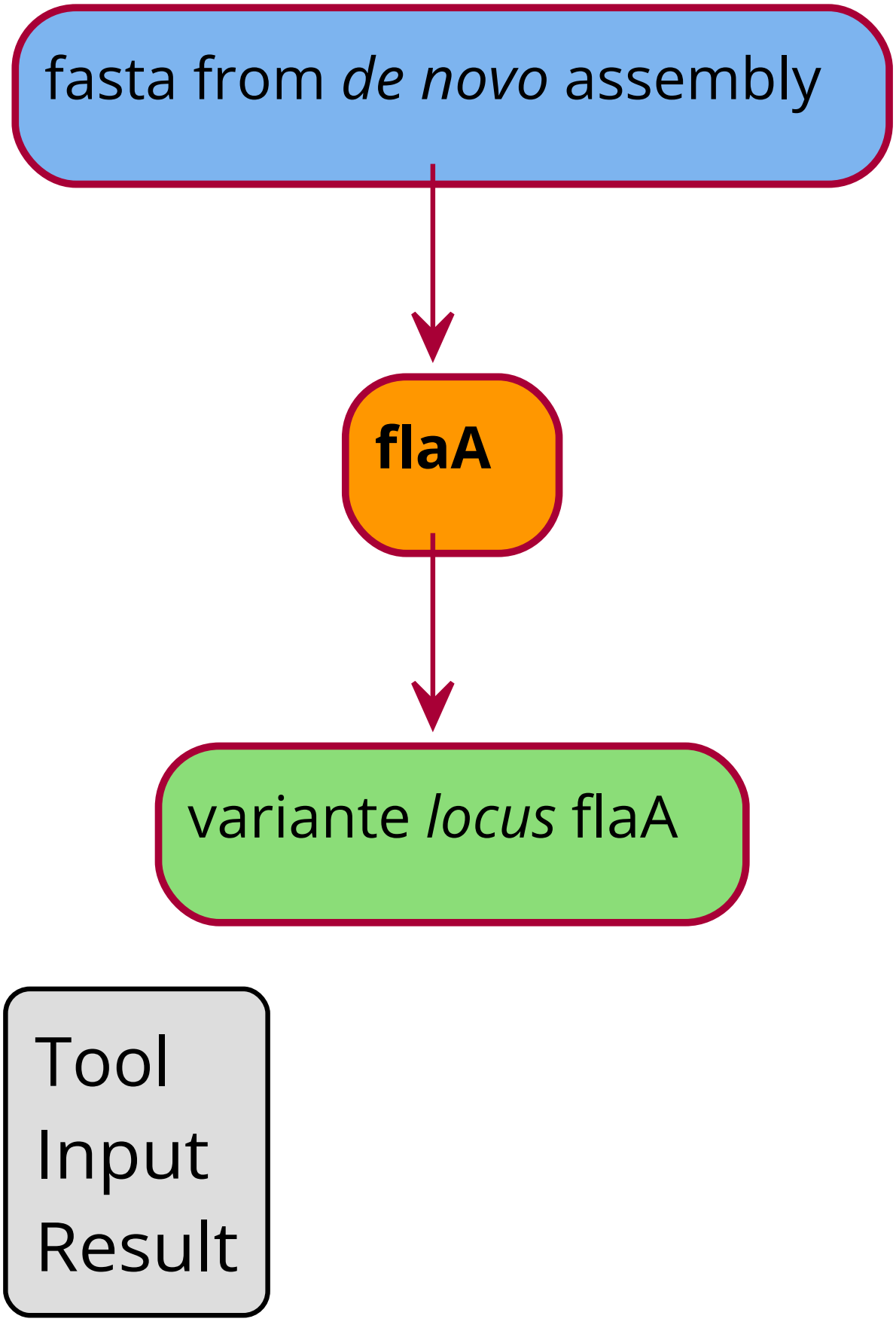
File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_new_alleles.txt	sequenza dei nuovi alleli identificati	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_alleles.tsv	chiamata degli alleli secondo codifica Pasteur pulito ogni volta in formato csv	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_contigsInfo.tsv	informazioni sul contig mappato su ogni *locus*	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_izsam.csv	chiamata degli alleli secondo codifica IZS	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_md5.csv	chiamata degli alleli secondo codifica md5	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_pasteur_2021-05-28.csv	chiamata degli alleli secondo codifica Pasteur in formato pulito ogni volta tsv	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_statistics.tsv	statistiche sul numero di *loci* con codifica EXC, INF, LNF, PLOT, NIPH, ALM, ASM	cartella result
DSXXXXXXXX-ID_import_chewbbaca_check.csv	controllo qualità con le informazioni su calledPerc, calledNum, annotated, new, notFound, discarded	cartella qc/ result

Per maggiori informazioni sulla codifica dei *loci* e sui files prodotti da chewBBACA, si invita a consultare la guida ufficiale di chewBBACA.

4TY_flaA

Introduzione

L'accertamento 4TY_flaA determina *in silico* la variante del *locus* flaA di *Campylobacter*.


fasta from *de novo* assembly
flaA
variante *locus* flaA
Tool
Input
Result

Lancia Analisi 4TY_flaA

Una volta selezionata l'analisi **4TY_flaA** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico (anche detto "metodo") da usare tra quelli disponibili per l'analisi. Il tool utilizzabile per questa analisi è:

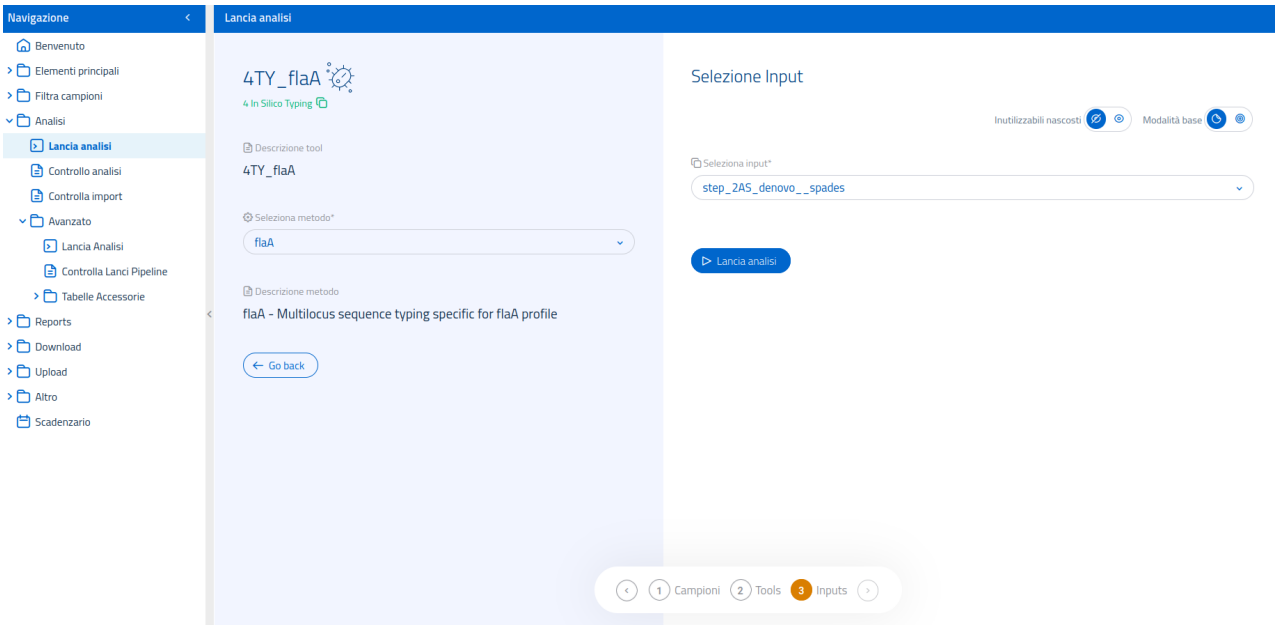
- **flaA** - Multilocus sequence typing specific for flaA profile

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per 4TY_flaA sono:

- step_2MG_denovo
- step_2AS_denovo
- step_2AS_mapping
- step_2AS_import

Sarà necessario specificare le sequenze di input provenienti da *de novo* assembly o da mapping e il genoma reference usato per il mapping, nel caso venga selezionato quest'ultimo come input.



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella verrà visualizzata con la seguente gerarchia: `results > ANNO > ID > 4TY_flaA > DSXXXXXXXX-DTXXXXXX_flaA`. Nella cartella sarà possibile trovare 2 directories:

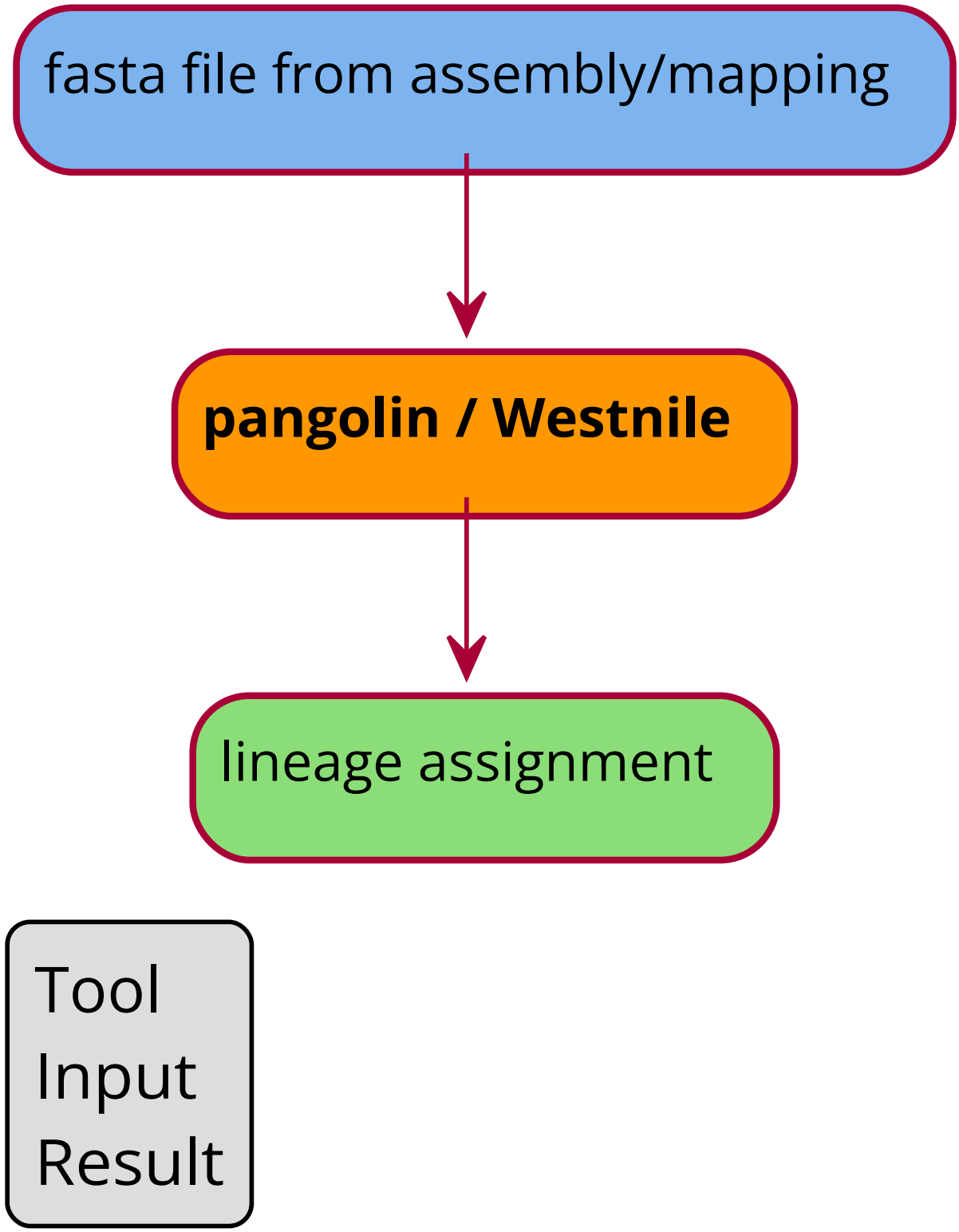
- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_flaA.tsv	variante allelica del locus flaA	cartella result

4TY_lineage

Introduzione

4TY_lineage assegna il lineage per SARS-CoV2 e West Nile Virus ed è esclusivo per i campioni di tali virus.



Lancia Analisi 4TY_lineage

Una volta selezionata l'analisi **4TY_lineage** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico (anche detto "metodo") da usare tra quelli disponibili per l'analisi. I metodi utilizzabili per questo accertamento sono:

Assegnazione del *lineage* di SARS-CoV2:

- **pangolin** - Phylogenetic Assignment of Named Global Outbreak LINeages

  Pagina GitHub di pangolin: https://github.com/cov-lineages/pangolin

Assegnazione del lineage di West Nile Virus:

- **Westnile** - West Nile virus lineage calculation

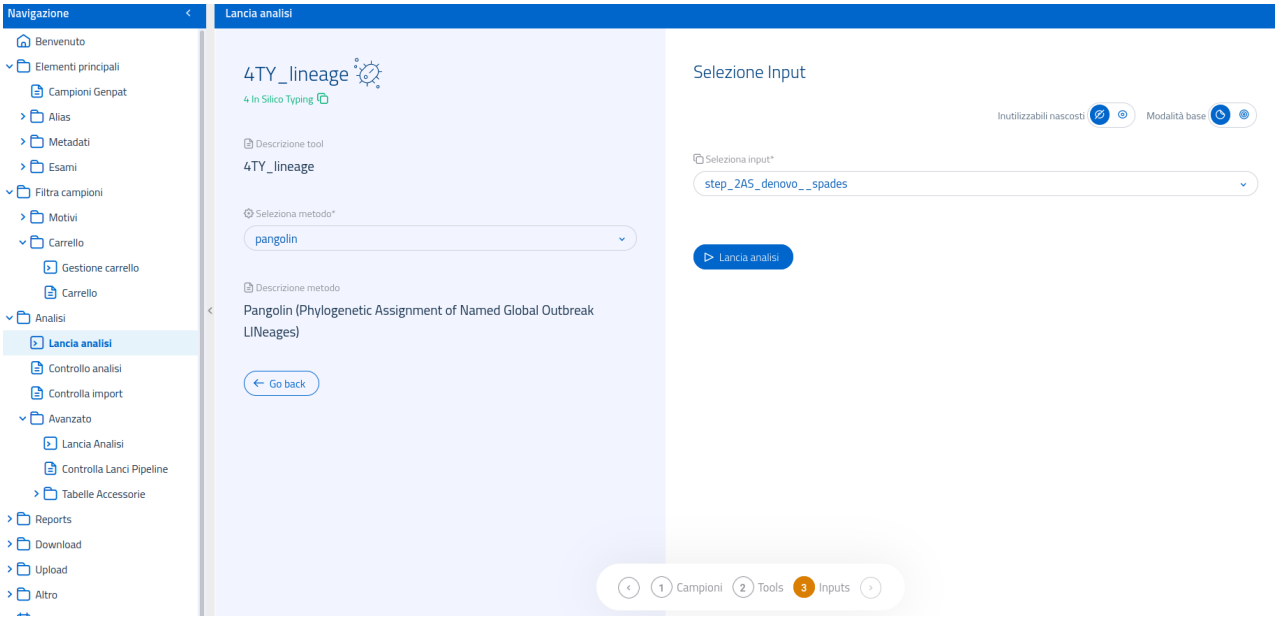
L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per 4TY_lineage sono:

- step_2MG_denovo
- step_2AS_denovo
- step_2AS_mapping
- step_2AS_import
- step_1PP_hostdepl
- step_1PP_downsampling
- step_1PP_trimming
- step_1PP_filtering

Per il tool **Westnile** l'unico input disponibile è il fasta da 2AS_mapping.

Sarà necessario specificare le sequenze di input provenienti da *de novo* assembly o da mapping e il genoma reference usato per il mapping, nel caso venga selezionato quest'ultimo come input.



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella verrà visualizzata con la seguente gerarchia: `results > ANNO > ID > 4TY_lineage > DSXXXXXXXX-DTXXXXXX_pangolin`, dove il suffisso "_pangolin" viene sostituito da "_westnile" nel caso sia stato scelto quest'ultimo come tool. Nella cartella sarà possibile trovare 2 directories:

- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

Pangolin

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_pangolin_lineage_report.csv	file csv con l'assegnazione del lineage	cartella result

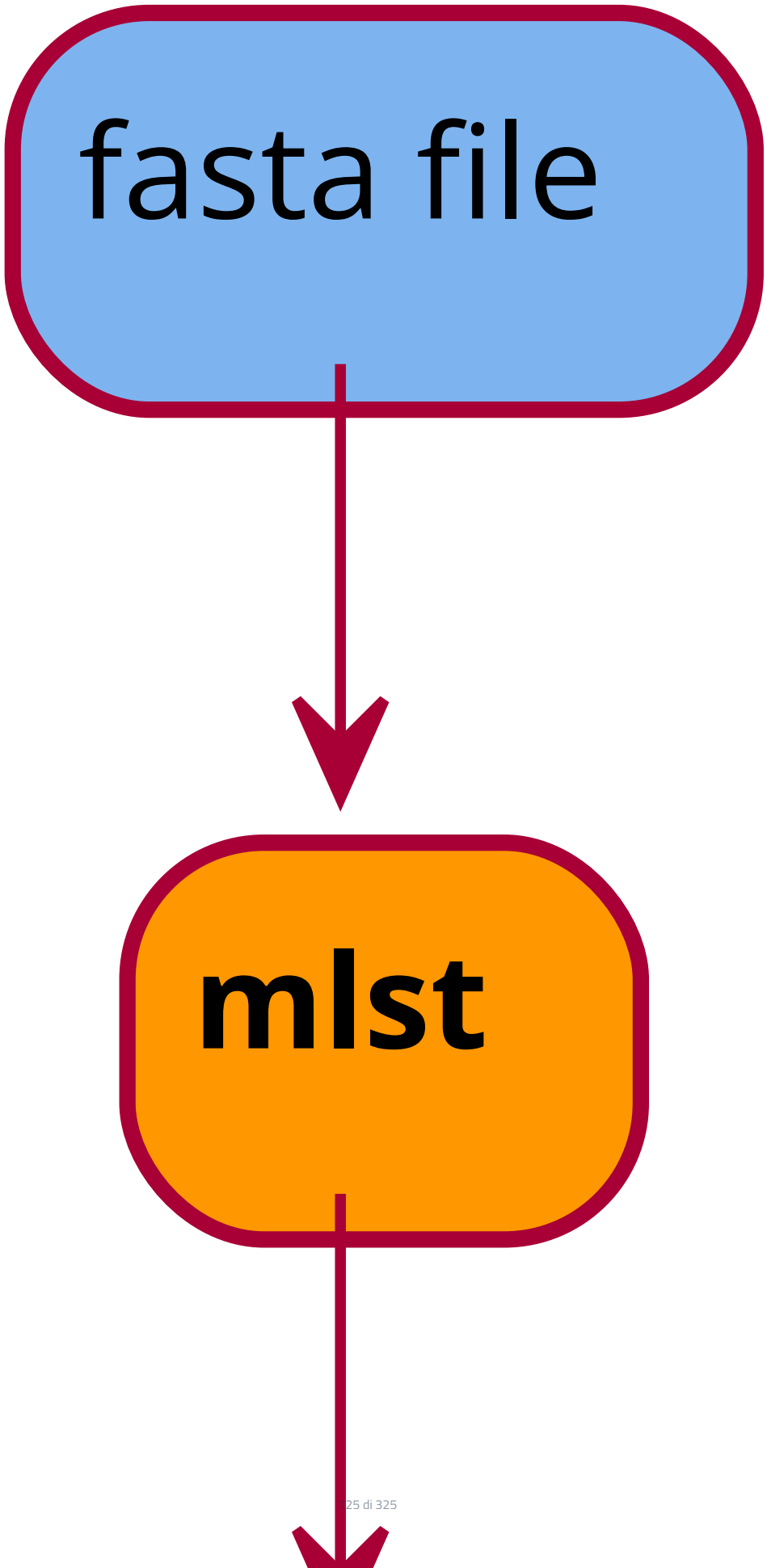
Westnile

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_westnile_lineage.csv	file csv con l'assegnazione del lineage	cartella result
DSXXXXXXXX-DTXXXXXX_ID_westnile_lineage_summary.csv	tabella csv riassuntiva	cartella result

4TY_MLST

Introduzione

4TY_MLST esegue il Multi Locus Sequence Typing *in silico*, assegnando sia il Sequence Type (ST) che il Clonal Complex (CC) dove possibile.


fasta file
mlst

Lancia Analisi 4TY_MLST

Una volta selezionata l'analisi **4TY_MLST** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico da usare tra quelli disponibili per l'analisi. Il tool utilizzabile per questa analisi è:

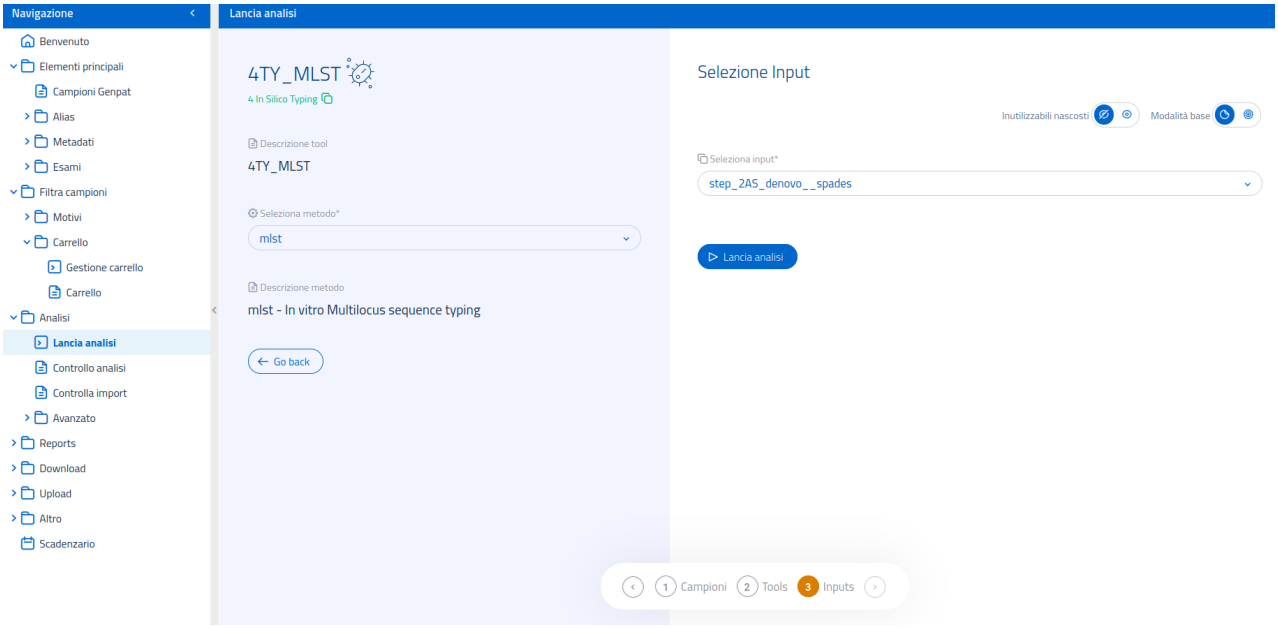
- **mlst** - *In vitro* Multi-Locus Sequence Typing

> **Pagina GitHub di mlsts:** https://github.com/tseemann/mlst

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per 4TY_MLST sono:

- step_2MG_denovo
- step_2AS_denovo
- step_2AS_mapping
- step_2AS_import



Sarà necessario specificare le sequenze di input provenienti da *de novo* assembly o da mapping e il genoma reference usato per il mapping, nel caso venga selezionato quest'ultimo come input.

Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella verrà visualizzata con la seguente gerarchia: `results > ANNO > ID > 4TY_MLST > DSXXXXXXXX-DTXXXXXX_mlst`. Nella cartella sarà possibile trovare 2 directories:

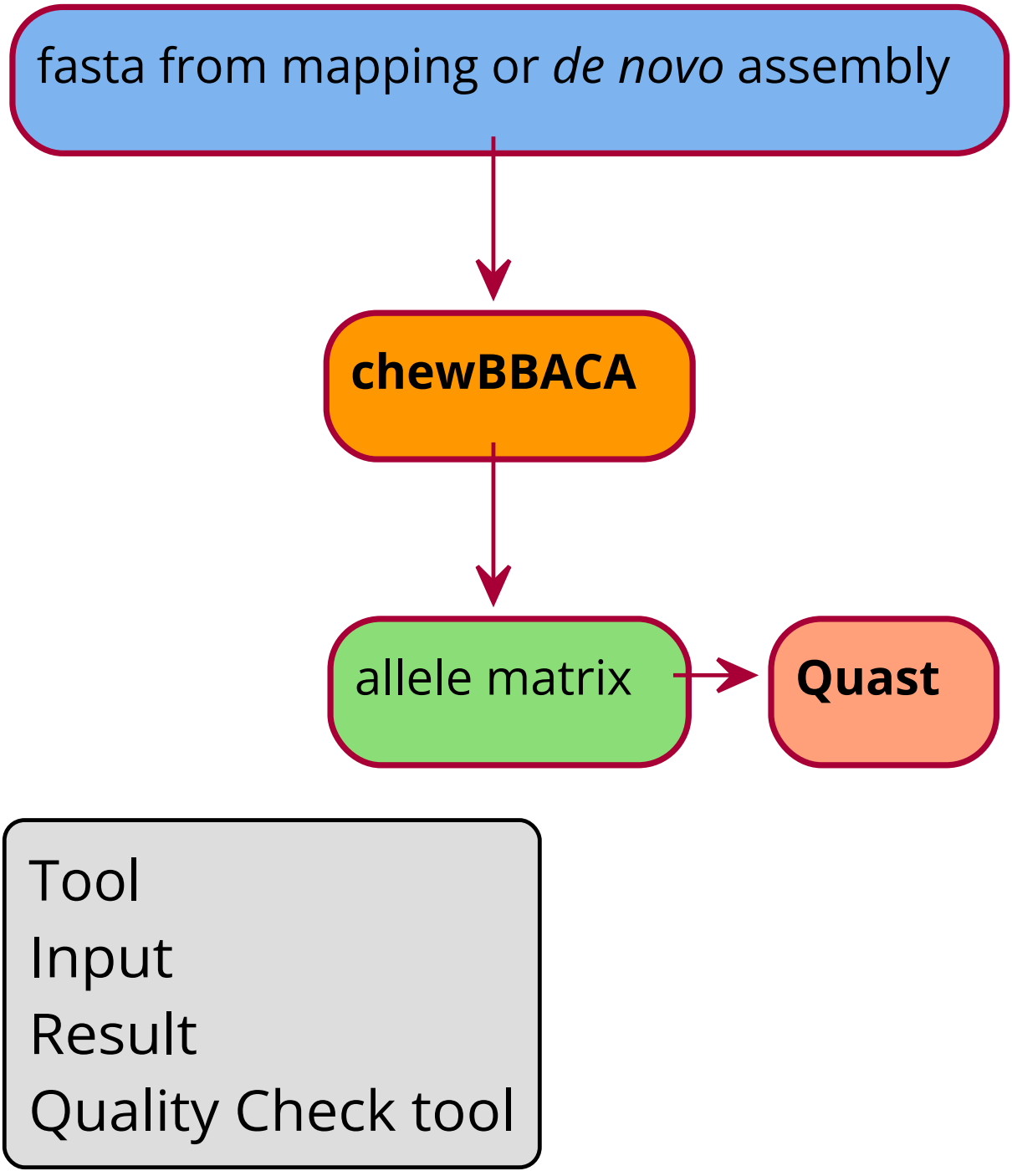
- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_mlst.tsv	risultato delle chiamate dei *loci* per il MLST	cartella result
DSXXXXXXXX-DTXXXXXX_ID_mlst_cc.csv	risultato con l'assegnazione dei CC	cartella result

4TY_wgMLST

Introduzione

4TY_wgMLST esegue il "whole genome Multi-Locus Sequence Typing" (wgMLST) *in silico* per le specie batteriche *Campylobacter jejuni* e *Campylobacter coli*. A differenza del protocollo cgMLST (core genome MLST), la caratterizzazione avviene sul *whole genome*.



Lancia Analisi 4TY_wgMLST

Una volta selezionata l'analisi **4TY_wgMLST** nella pagina dedicata al lancio di analisi, sarà possibile selezionare il software bioinformatico (anche detto "metodo") da usare tra quelli disponibili per l'analisi. L'unico metodo utilizzabile per questo accertamento è **chewBBACA** - BSR-Based Allele Calling Algorithm.

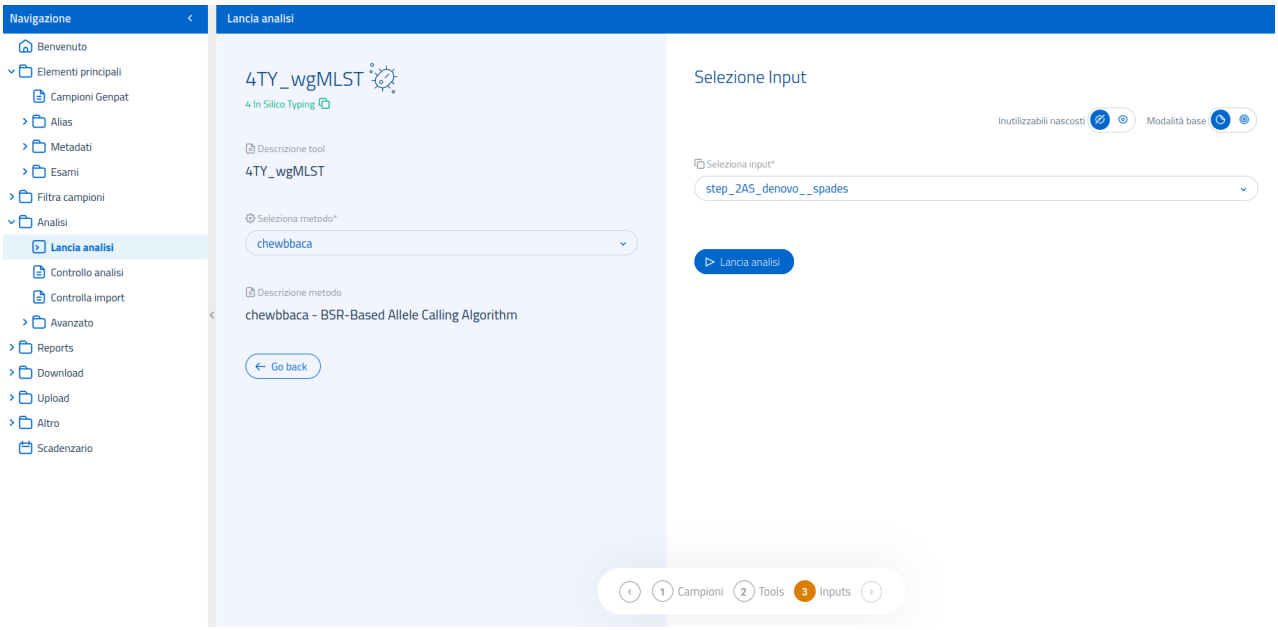
> **Nota:** Sono disponibili schemi per wgMLST solo per *Campylobacter jejuni* e *Campylobacter coli*. Il lancio dell'analisi su un altro microrganismo, per cui non sia presente lo schema wgMLST, comporterà il fallimento della run.

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per 4TY_wgMLST sono:

- step_2MG_denovo
- step_2AS_denovo
- step_2AS_mapping
- step_2AS_import

Sarà necessario specificare le sequenze di input provenienti da *de novo* assembly o da mapping e il genoma reference usato per il mapping, nel caso venga selezionato quest'ultimo come input.



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella verrà visualizzata con la seguente gerarchia: `results > ANNO > ID > 4TY_wgMLST > DSXXXXXXXX-DTXXXXXX_chewbbaca`. Nella cartella sarà possibile trovare le 3 directories seguenti:

- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **qc**: ("quality check") in cui si trovano i files del controllo qualità. A sua volta in questa cartella sono presenti le cartelle meta e result. Per questo accertamento il controllo qualità viene effettuato da **Quast**.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

Per la chiamata allelica con chewBBACA i risultati verranno forniti con codifica disponibile secondo 3 diversi schemi:

- **schema IZS:** ogni allele è identificato da un ID numerico progressivo. L'assegnazione dell'ID avviene considerando tutti i *loci*, quindi la numerazione **NON** ricomincia da 1 ad ogni nuovo *locus*.
- **schema MD5:** ogni allele è identificato da un codice alfanumerico di 16 caratteri (MD5), ottenuto applicando una funzione di "hash" alla sequenza stessa dell'allele; il che lo rende interscambiabile con altri metodi.
- **schema CRC32:** controllo basato sul Cyclic Redundancy Check (CRC). Ad ogni allele è assegnato un codice numerico di controllo (*check value*) determinato dal resto della divisione polinomiale del blocco di dati. Tale valore è utilizzato nello schema come identificatore dell'allele.

Si noti che il file di esempio "DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_ccoli_curated_210722.csv" riportato in tabella potrà riportare come suffisso l'abbreviazione per *C. coli* o *C. jejuni*, a differenza della specie analizzata.

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_new_alleles.txt	sequenza dei nuovi alleli identificati	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_alleles.tsv	chiamata degli alleli secondo schema Pasteur pulito ogni volta in formato csv	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_ccoli_curated_210722.csv	tabella csv degli alleli. Righe: campioni, colonne: *loci*	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_contigsInfo.tsv	informazioni sul contig mappato su ogni *locus*	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_crc32.csv	tabella csv degli alleli codificati con schema CRC32. Righe: campioni, colonne: *loci*	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_izsam.csv	chiamata degli alleli secondo schema IZS	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_md5.csv	chiamata degli alleli secondo schema md5	cartella result
DSXXXXXXXX-DTXXXXXX_ID_chewbbaca_results_statistics.tsv	statistiche sul numero di *loci* con codifica EXC, INF, LNF, PLOT, NIPH, ALM, ASM	cartella result
DSXXXXXXXX-DTXXXXXX_ID_import_chewbbaca_check.csv	controllo qualità con le informazioni su calledPerc, calledNum, annotated, new, notFound, discarded	cartella qc/ result

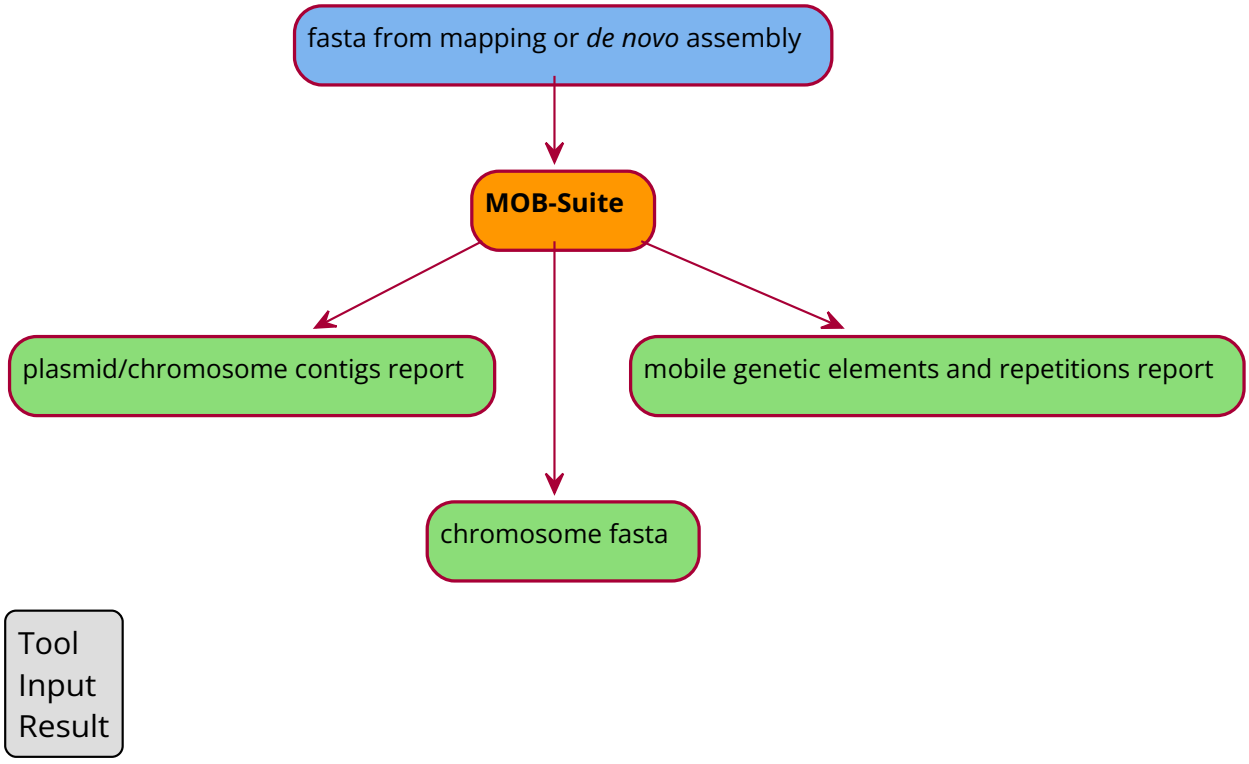
Per maggiori informazioni sulla codifica dei *loci* e sui files prodotti da chewBBACA, si invita a consultare la guida ufficiale di chewBBACA.

4TY_plasmid

Introduzione

L'analisi 4TY_plasmid esegue il software MOB-Suite per tipizzare e ricostruire le sequenze plasmidiche, a partire dagli assembly da Whole Genome Sequencing (WGS).

Pagina GitHub di MOB-Suite: https://github.com/phac-nml/mob-suite



Lancia Analisi 4TY_plasmid

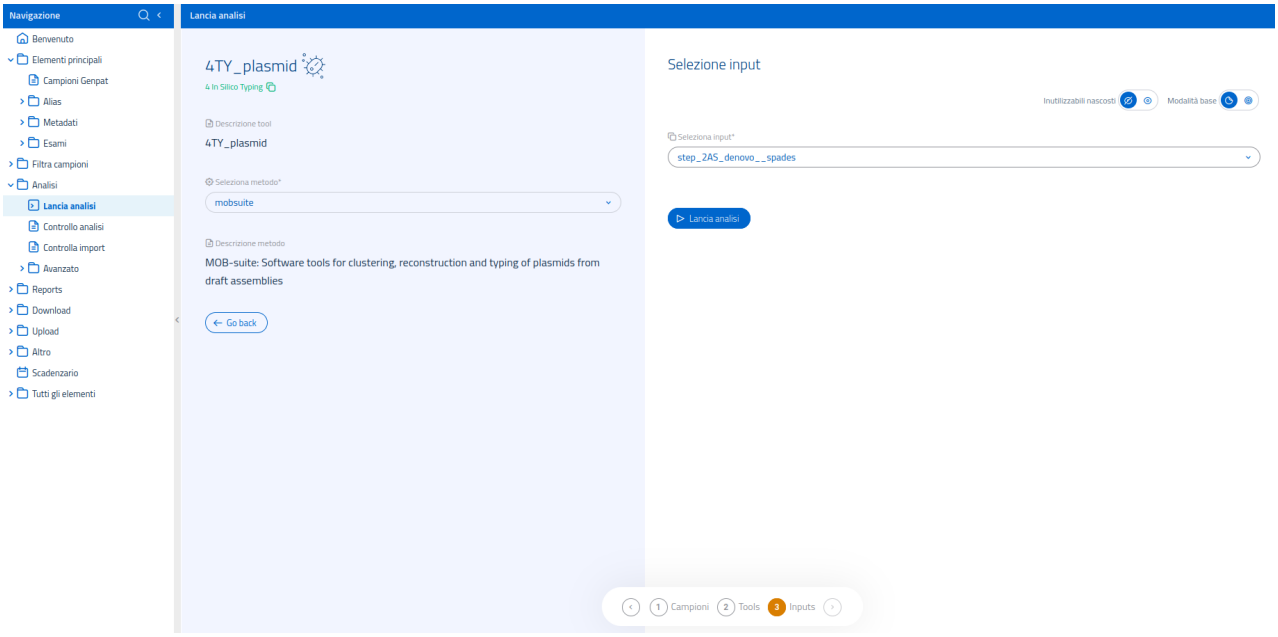
Una volta selezionata l'analisi **4TY_plasmid** nella pagina dedicata al lancio di analisi. Il software bioinformatico disponibile per l'analisi è **MOB-suite** - *Software tools for clustering, reconstruction and typing of plasmids from draft assemblies*.

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per 4TY_plasmid sono:

- step_2MG_denovo
- step_2AS_denovo
- step_2AS_mapping
- step_2AS_hybrid
- step_2AS_import

Sarà necessario specificare le sequenze di input provenienti da *de novo* assembly o da mapping e il genoma reference usato per il mapping, nel caso venga selezionato quest'ultimo come input.



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sull'apposito tasto all'interno della scheda dell'analisi. Nella cartella sarà possibile trovare 2 directories:

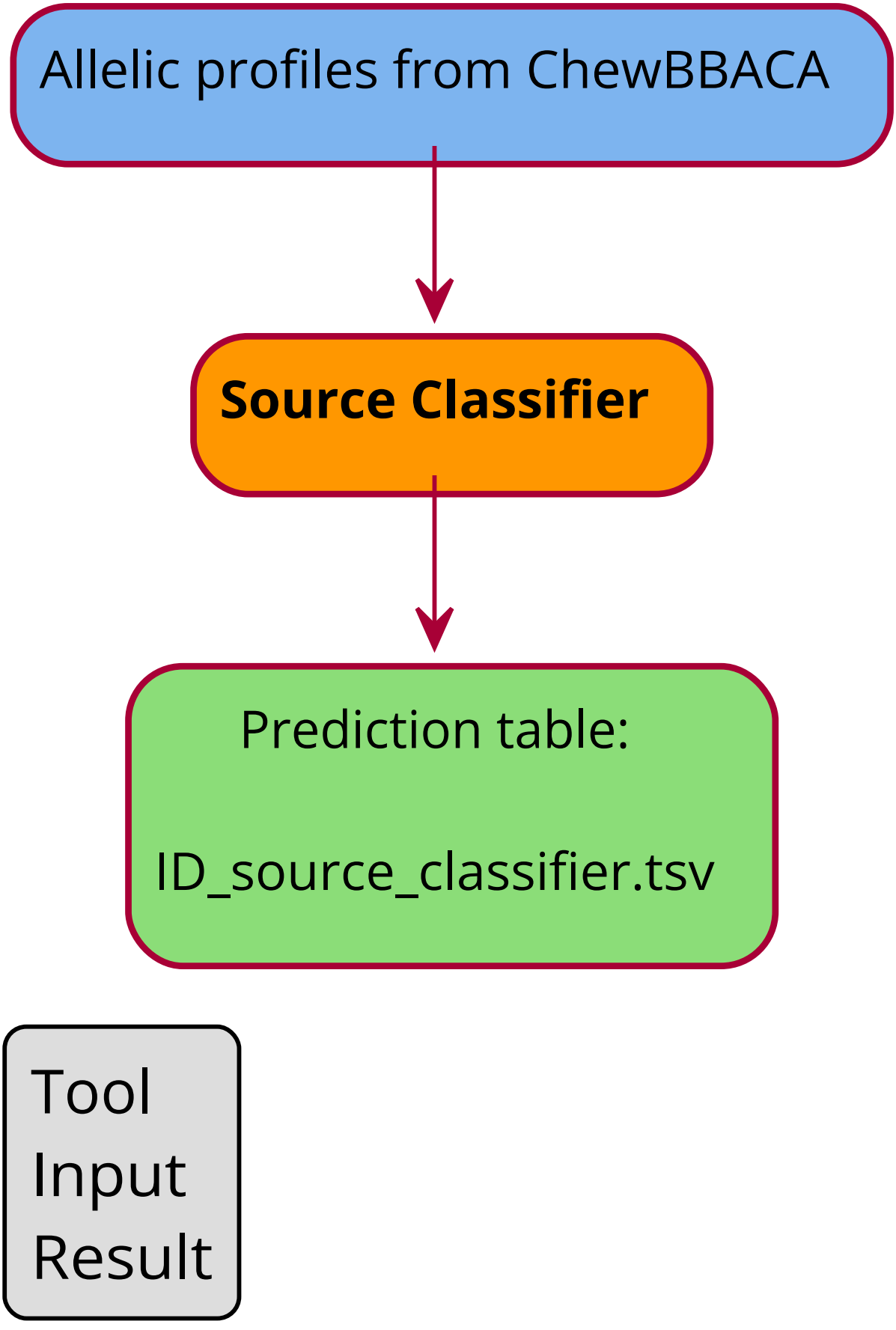
- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_mobsuite_chromosome.fasta	fasta con i contigs appartenenti al cromosoma	cartella result
DSXXXXXXXX-DTXXXXXX_ID_mobsuite_contig_report.txt	file tabulare con l'assegnazione dei contig al cromosoma o ad un gruppo plasmidico	cartella result
DSXXXXXXXX-DTXXXXXX_ID_mobsuite_mge.report.txt	blast HSP degli elementi ripetuti e degli elementi genetici mobili (MGE) rilevati	cartella result

4TY_ML

Introduzione

L'analisi ML (Machine Learning) consente di effettuare la predizione della sorgente alimentare (Source Attribution) di *L. monocytogenes*. L'analisi usa un classificatore binario (sorgenti predette: "meat" o "non-meat"), prodotto come *deliverable* del progetto di ricerca corrente.


Allelic profiles from ChewBBACA
Source Classifier
Prediction table:
ID_source_classifier.tsv
Tool
Input
Result

Lancia Analisi 4TY_ML

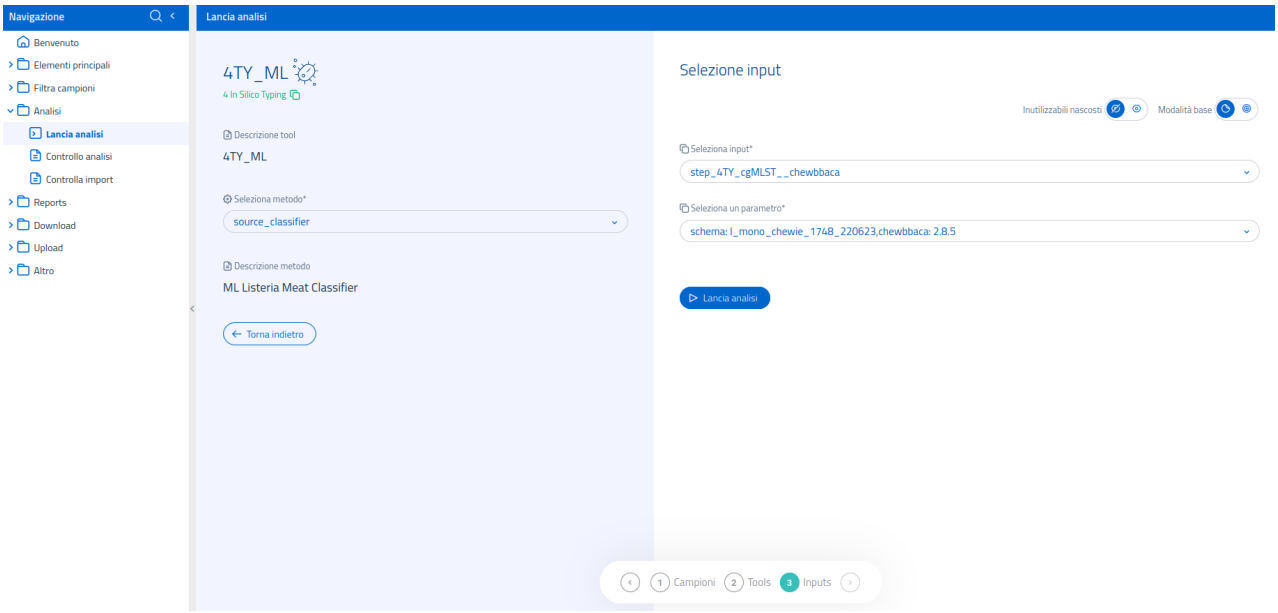
L'analisi **4TY_ML** prevede l'uso di un modello già pre-addestrato come source-classifier, disponibile attualmente solo per *L. monocytogenes*. La classificazione per fenotipo sarà binaria, ovvero il modello è addestrato per predire quali campioni, tra le *L. monocytogenes* date in input, proviene da una matrice alimentare "carne" e quali da una matrice "non carne". La predizione si basa sulla matrice di alleli da cgMLST.

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per 4TY_ML sono:

- step_4TY_wgMLST
- step_4TY_cgMLST

Entrambi i tipi di input prevedono l'uso di ChewBBACA come software di esecuzione, inoltre per **4TY_cgMLST** è possibile usare alternativamente la versione 2.8.4 con schema Pasteur o 2.8.5 con schema crc32.



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella di output

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella verrà visualizzata con la seguente gerarchia: `results > ANNO > ID > 4TY_ML > DSXXXXXXXX-DTXXXXXX_source_classifier`. Nella cartella sarà possibile trovare 2 directories:

- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

Nella tabella seguente è riportata la posizione ed una breve descrizione dei principali files prodotti dall'analisi.

File	Descrizione	Posizione
DSXXXXXXXX-DTXXXXXX_ID_source_classifier.tsv	tabella tsv con i risultati della predizione per il campione	cartella result
metadata.csv	tabella csv dei metadati dei campioni usati come input	cartella results

Visualizzazione dei risultati

I risultati dell'analisi **4TY_ML** possono essere visualizzati anche tramite un'apposita tabella di check, ovvero **Check ML**, disponibile sotto `Reports > Dettaglio > Checks > Check ML` del menù di navigazione laterale.

5.4.4 Multi sample

SNPs-based clustering

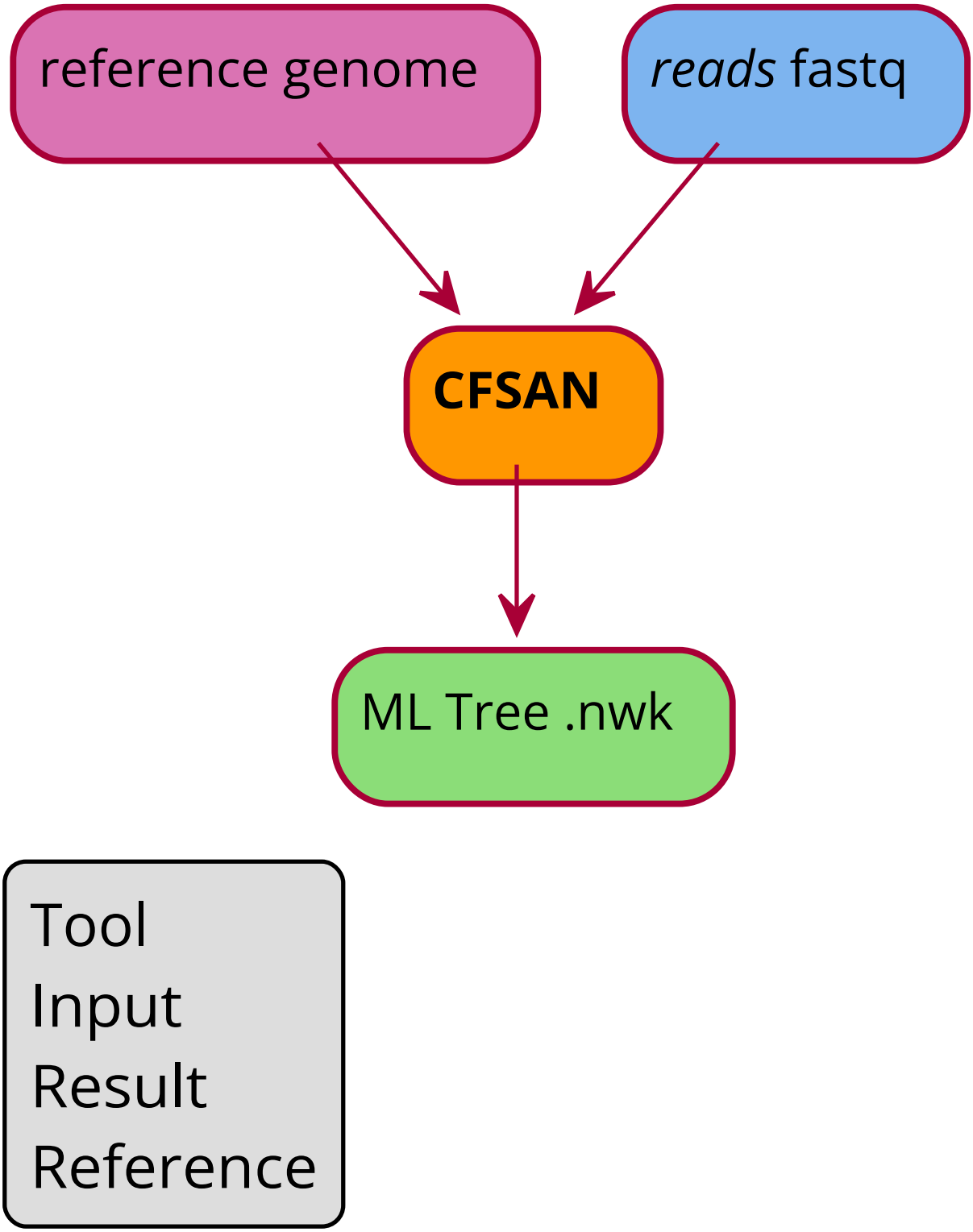
CFSAN

Introduzione

CFSAN effettua un'analisi filogenetica *reference-based* e conseguente costruzione di un dendrogramma di distanza tra microrganismi, usando l'algoritmo "Maximum Likelihood" (ML), a partire da una matrice di Polimorfismi a Singolo Nucleotide (SNPs).

Maggiori informazioni sul software sono disponibili nella guida ufficiale di CFSAN.

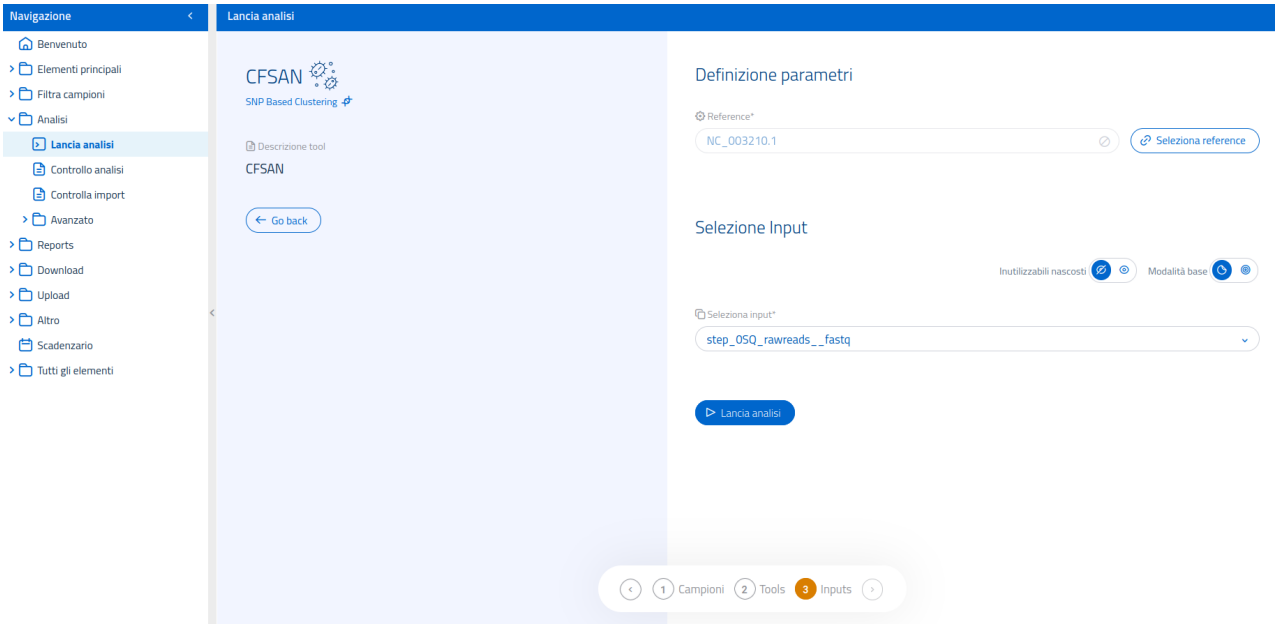
Pagina GitHub di CFSAN: https://github.com/CFSAN-Biostatistics/snp-pipeline



Lancia Analisi CFSAN

Una volta selezionata l'analisi **CFSAN** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma dell'analisi scelta.

Nella schermata di definizione parametri sarà necessario specificare un genoma reference; inoltre l'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi. Nel caso di CFSAN, gli input sono sempre *raw reads* in formato FASTQ, ma possono essere interne (**step_0SQ_rawreads__fastq**) o importate in piattaforma (**step_0SQ_rawreads__external**).



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di verificare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Dalla scheda dell'analisi completata è possibile, oltre all'esplorazione della cartella di output, anche la visualizzazione dell'albero nwk e l'accesso diretto ad alcuni dei files di log e metadati.

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 2 sotto-cartelle:

- **meta**: ("metadati") in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

La tabella in basso presenta la lista dei principali file di interesse presenti nelle cartelle, insieme ad alcune informazioni utili.

File	Descrizione	Posizione
metrics.tsv	tabella .tsv riassuntiva, con valori e statistiche dei campioni e relativi SNPs	cartella "result"
referenceSNP.fasta	file FASTA contenente le basi della sequenza reference per tutte le posizioni degli SNPs	cartella "result"
snp_distance_matrix.tsv	tabella .tsv con matrice di distanza, calcolata sulle differenze in SNP, per ogni coppia di campioni	cartella "result"
snp_distance_pairwise.tsv	tabella .tsv con le *pairwise distances* in base alle differenze in SNPs per ogni coppia di campioni	cartella "result"
snpma.fasta.nwk	file dell'albero in formato .nwk (newick) calcolato con algoritmo ML	cartella "result"
snpma.vcf	elenco degli SNPs e delle relative posizioni, per ogni campione, in formato VCFv4.2	cartella "result"

Per ulteriori informazioni in merito a tutti gli output di CFSAN, si invita a consultare l'apposita sezione del manuale ufficiale CFSAN.

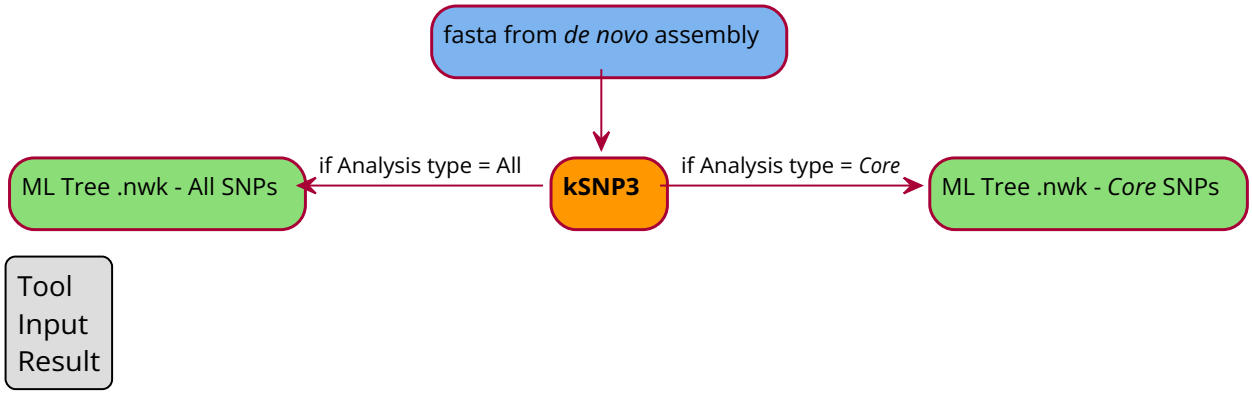
kSNP3

Introduzione

kSNP3 identifica i Polimorfismi a Singolo Nucleotide (SNPs) nei campioni, esegue un'analisi filogenetica senza allineamento (*reference-free*) e costruisce un albero filogenetico con algoritmo Maximum Likelihood (ML), a partire dalla matrice di SNPs dei campioni.

Una guida all'utente completa è disponibile nell'apposita sezione della guida di kSNP3.

- Sito ufficiale
- pagina GitHub



Lancia Analisi kSNP3

Una volta selezionata l'analisi **kSNP3** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma dell'analisi scelta.

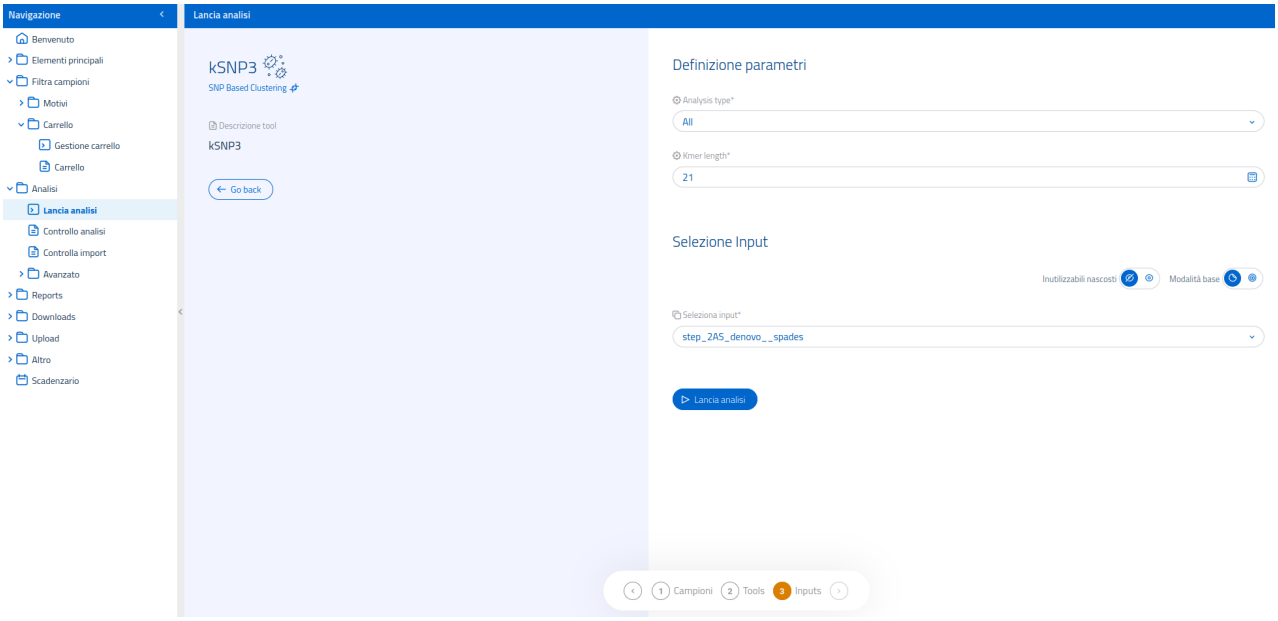
Nella sezione di definizione parametri sarà possibile scegliere se:

1. effettuare l'analisi filogenetica utilizzando tutti gli SNPs ("All") o solo i *Core* SNPs (opzione "**Analysis type**");
2. cambiare la lunghezza dei kmer (opzione "**kmer length**", default = 21).

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

Gli input utilizzabili per kSNP3 sono:

- step_2AS_denovo
- step_2MG_denovo



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di verificare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Dalla scheda dell'analisi completata è possibile, oltre all'esplorazione della cartella di output, anche la visualizzazione dell'albero .nwk e l'accesso diretto ad alcuni dei files di log e metadati.

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 2 sotto-cartelle:

- **meta**: ("metadati") in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

> Si tenga presente che, in base al parametro usato per **Analysis type**, il nome dei file di output sarà diverso:
>
> - SNPs_all_matrix (se è stato selezionato "All" per *Analysis type*);
> - core_SNPs_matrix (se è stato selezionato "Core" per *Analysis type*).
>
> Nella tabella sottostante sono riportati i nomi nel caso di esecuzione su tutti gli SNPs ("All").

File	Descrizione	Posizione
SNPs_all_matrix.fasta	file multifasta contenente gli *headers* con il nome del campione e la riga per la sequenza di SNPs concatenati. "N" indica l'assenza di uno SNP in quel campione	cartella "results"
SNPs_all_matrix.fasta.nwk	file dell'albero in formato nwk (newick) calcolato con algoritmo ML	cartella "results"

> Per ulteriori informazioni riguardo la formattazione degli output di kSNP3, si invita a consultare l'apposita sezione della guida utente kSNP3.

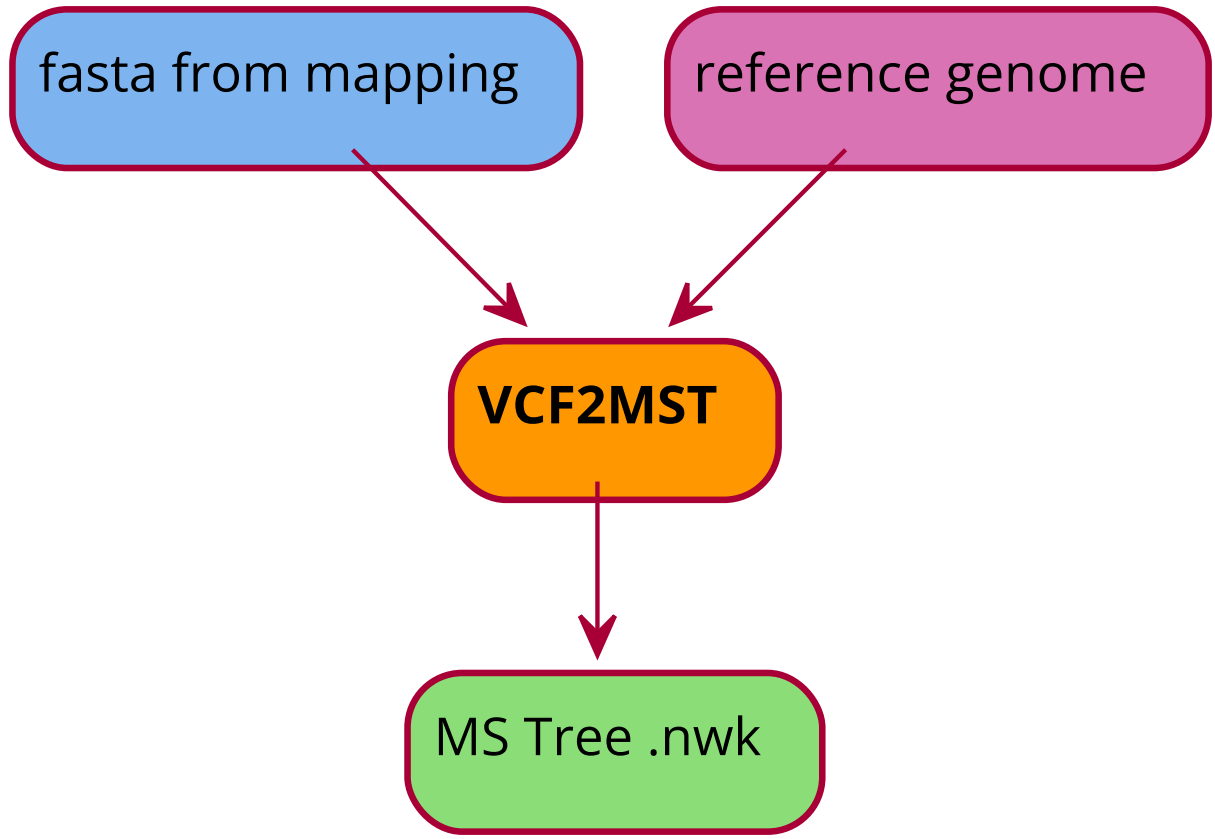
VCF2MST

Introduzione

VCF2MST costruisce un albero filogenetico di tipo Minimum Spanning Tree (MST) a partire da file VCF contenenti Polimorfismi a Singolo Nucleotide (SNPs), senza ricorrere ainferenze filogenomiche basate sull'allineamento delle sequenze.

Il programma costruisce il Minimum Spanning Tree basandosi sulle distanze di Hamming, che misurano il numero di sostituzioni necessarie perchè una sequenza (stringa) si trasformi in un'altra.

Maggiori informazioni sul software sono reperibili nella pagina GitHub di "VCF2MST - Hamming Distance based Minimum Spanning Tree from Samples vcf using graptree" o consultandone l'articolo.



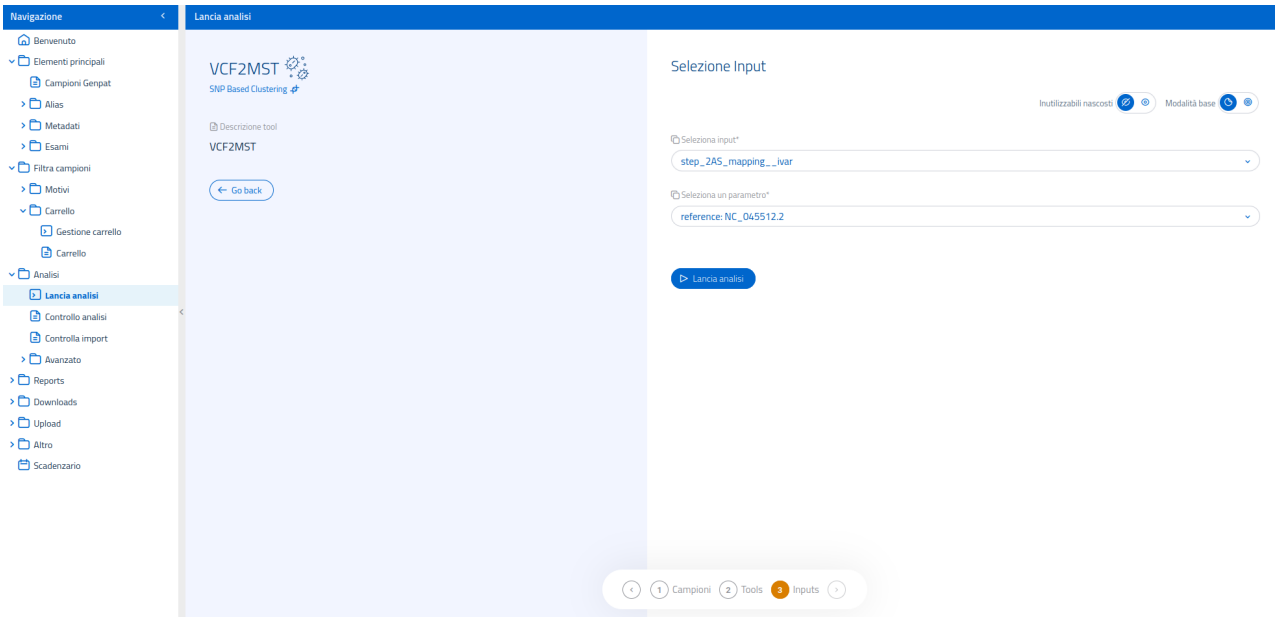
Tool
Input
Result
Reference

Lancia analisi VCF2MST

Una volta selezionata l'analisi **VCF2MST** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input per VCF2MST ricadono nella categoria step_2AS_mapping. Nella schermata di selezione input sarà anche presente un campo per il genoma reference (il reference viene inserito automaticamente).



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di verificare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Dalla scheda dell'analisi completata è possibile, oltre all'esplorazione della cartella dei risultati, anche la visualizzazione dell'albero .nwk tramite l'integrazione SPREAD in GenPat e l'accesso diretto al file .nwk ed ai metadati.

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 2 sotto-cartelle:

- **meta**: ("metadati") in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

La tabella in basso riporta alcune informazioni sul file di output prodotto da VCF2MST.

File	Descrizione	Posizione
tree.nwk	file dell'albero MST in formato .nwk (newick)	cartella "result"

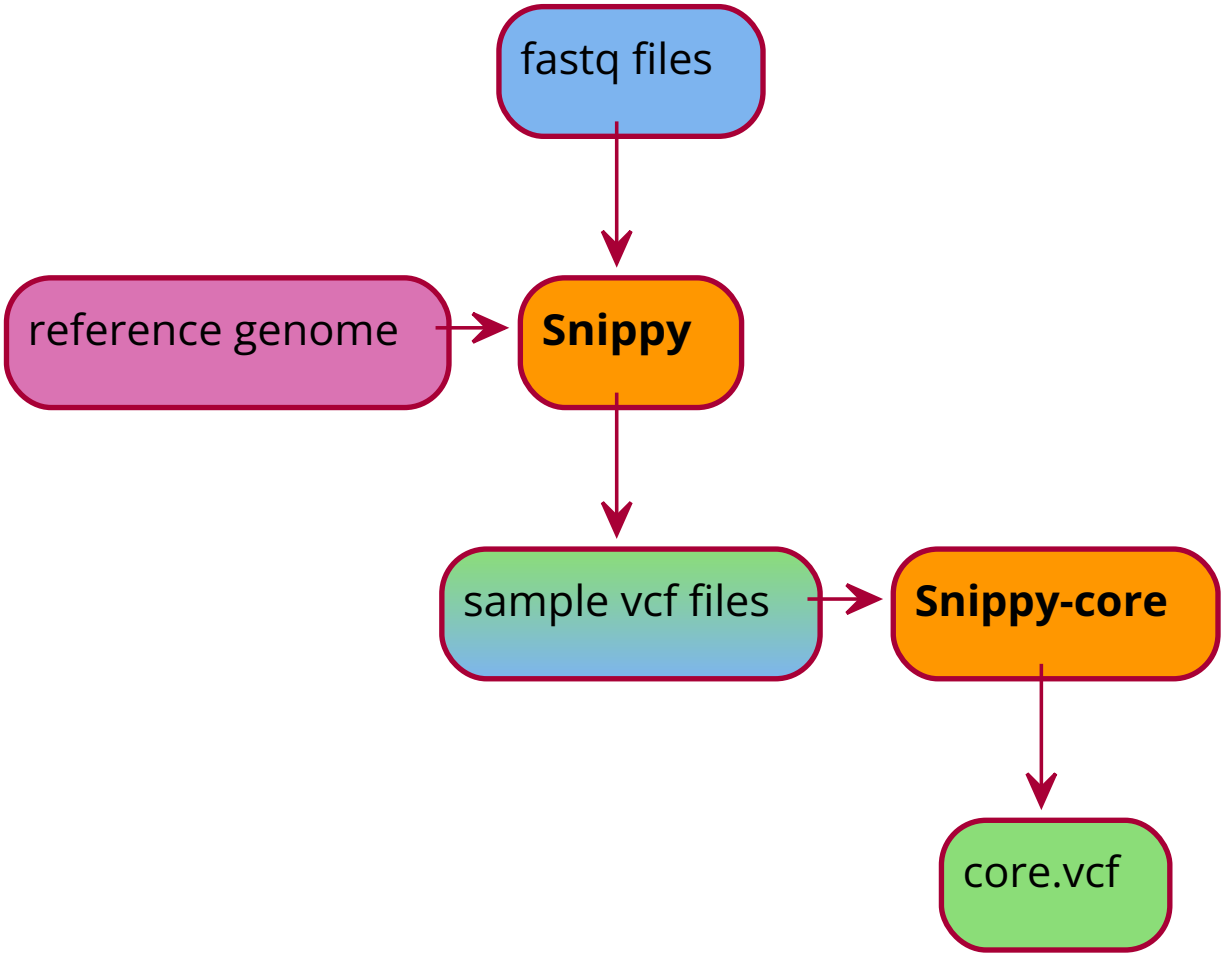
Snippy-core

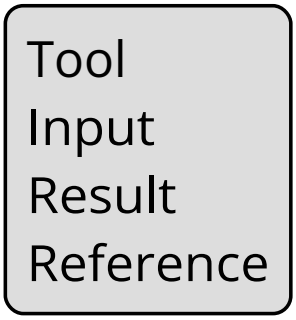
Introduzione

Snippy-core effettua l'estrazione dei *core SNPs* dai file VCF prodotti da Snippy (*Rapid haploid variant calling and core genome alignment*).

Se non è stato precedentemente eseguito un allineamento con Snippy, l'analisi Snippy-core effettuerà, in un primo momento, il lancio del tool Snippy, per produrre - tramite allineamento ad un reference GeneBank con formato `.gb` o `.gbk` - i file VCF dei campioni selezionati; successivamente, essa calcolerà il VCF-*core* per i campioni del dataset, usando Snippy-core (da cui la pipeline prende il nome). Il `core.vcf` così generato contiene le varianti (SNPs e indels) *core* tra quelle originariamente presenti nei VCF dei singoli campioni. Nel caso siano stati già eseguiti gli allineamenti con Snippy, l'analisi esegue solo il software Snippy-core.

Pagina GitHub di Snippy & Snippy-core: https://github.com/tseemann/snippy





Lancia Analisi Snippy-core

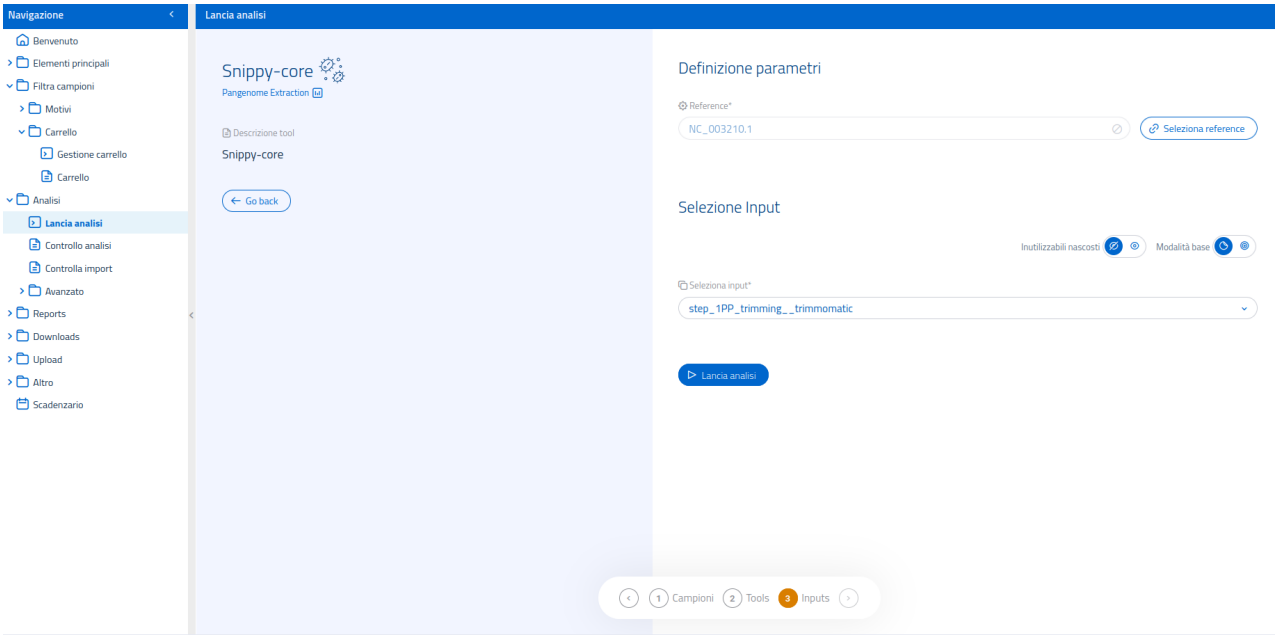
Una volta selezionata l'analisi **Snippy-core** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma dell'analisi scelta. Il lancio è specifico per la pipeline di *Snippy* + *Snippy-core* e non vi sono altri tool utilizzabili.

Nell'interfaccia per la definizione dei parametri, sarà necessario selezionare un reference. Verrà sempre usato un reference `.gb` per permettere l'esecuzione di Snippy-core, dopo Snippy.

La sezione dedicata alla selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per Snippy-core sono:

- step_1PP_trimming
- step_1PP_generated
- step_1PP_hostdepl
- step_1PP_downsampling



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di verificare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella dei risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results`, è possibile trovare 2 sotto-cartelle:

- **meta**: ("metadati") in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

La tabella in basso presenta la lista dei principali file di interesse presenti nelle cartelle, insieme ad alcune informazioni utili.

File	Descrizione	Posizione
core.aln	file FASTA con le sequenze dei campioni e del reference	cartella "result"
core.full.aln	file FASTA con le sequenze dei campioni allineate al reference	cartella "result"
core.ref.fa	file FASTA con la sequenza nucleotidica del genoma reference	cartella "result"
core.tab	tabella .tsv delle varianti. Colonne: cromosoma, posizione, nucleotide nel reference, nucleotide nel campione 1, nucleotide nel campione 2...	cartella "result"
core.txt	file di testo con tabella riassuntiva delle caratteristiche delle sequenze di campioni e reference	cartella "result"
core.vcf	file VCFv4.2 delle varianti. Comprende un header informativo e la tabella con il tipo di variante e una matrice binaria di presenza/assenza della variante nei campioni	cartella "result"

Gene-by-gene based clustering

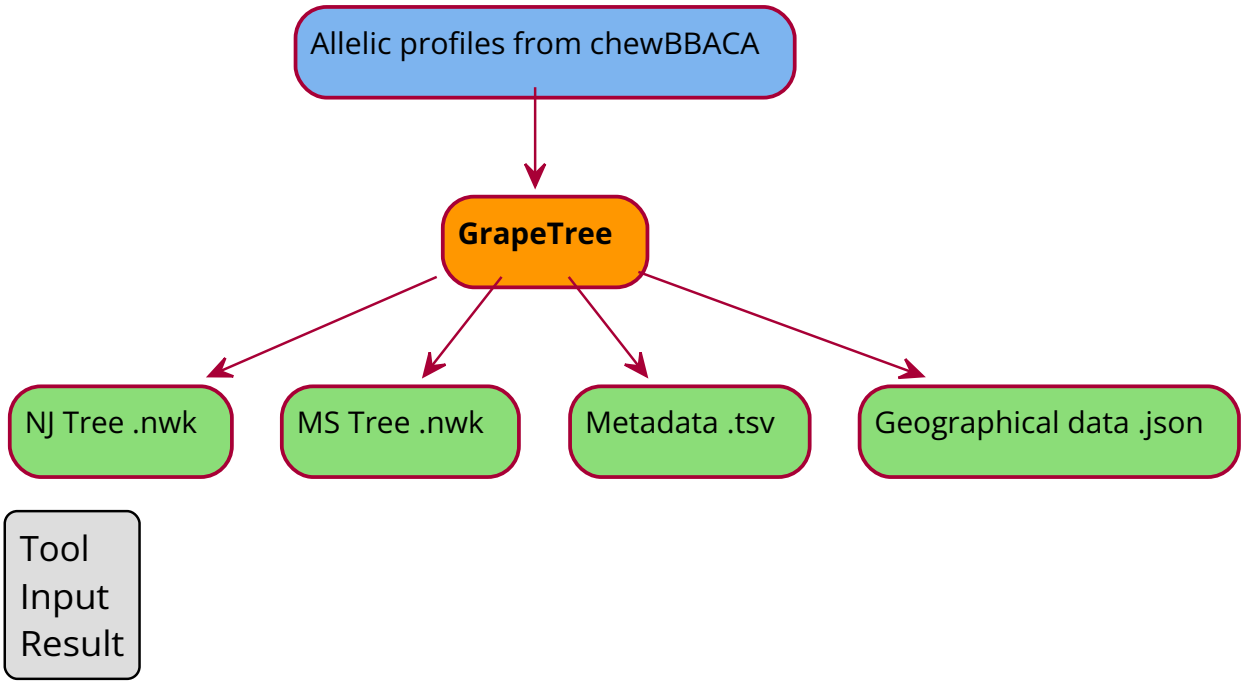
GrapeTree

Introduzione

L'analisi GrapeTree costruisce grafici ad albero usando i 2 algoritmi "Minimum Spanning Tree" (MSTree) e "Neighbor Joining" (NJ) a partire dai profili allelici dei campioni forniti, così come identificati nell'analisi cgMLST.

Pagina github di GrapeTree: https://github.com/achtman-lab/GrapeTree

In piattaforma è possibile sia calcolare gli alberi, sia visualizzarli grazie all'integrazione della versione stand-alone del sistema interattivo di visualizzazione di GrapeTree.



Una guida completa all'uso e all'interfaccia del sistema di visualizzazione in piattaforma è disponibile nella pagina dedicata a SPREAD.

Per ulteriori informazioni sul software, si invita a consultare l'articolo e la documentazione ufficiale del sistema EnteroBase, di cui GrapeTree fa parte.

Lancia analisi Grapetree

Una volta selezionata l'analisi **Grapetree** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma: non sarà necessario scegliere un tool dal momento che le analisi multi-campione usano sempre un software specifico.

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per Grapetree sono:

- step_4TY_wgMLST
- step_4TY_cgMLST

Per entrambi gli input possibili, è necessario che il programma utilizzato, per il calcolo dei profili allelici, sia **chewBBACA**. Si consiglia, dunque, di prestare particolare attenzione e selezionare tale software per il lancio dell'analisi 4TY_cgMLST o 4TY_wgMLST, se si intende lanciare GrapeTree come analisi a valle.

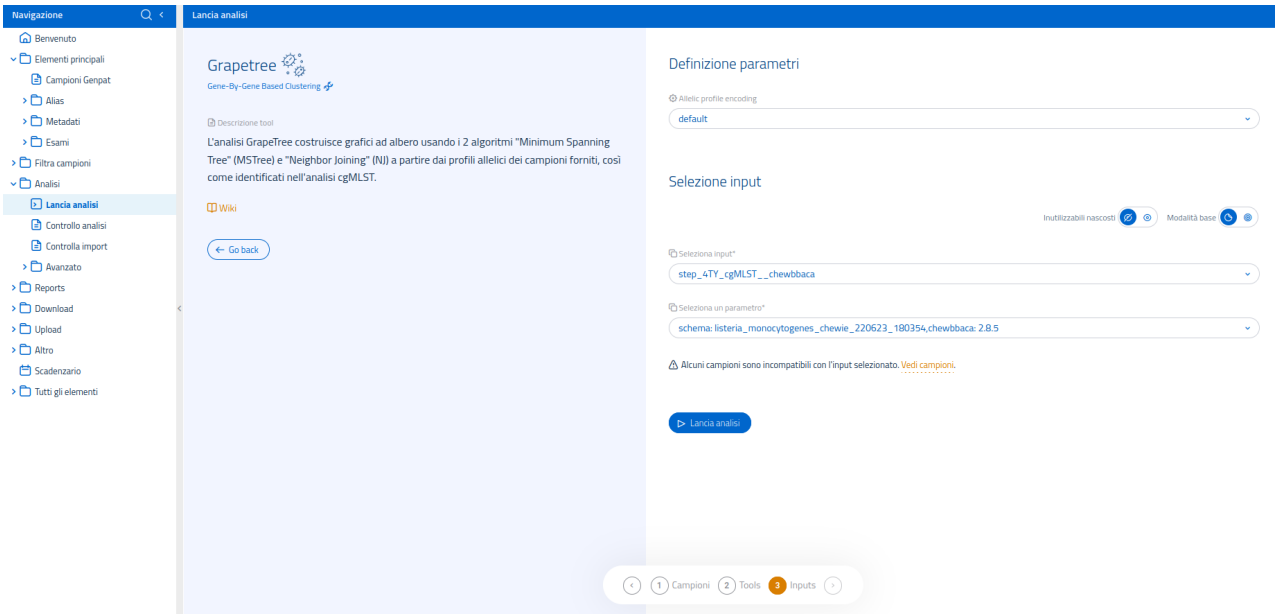
Nella schermata di definizione parametri è possibile:

- selezionare la codifica: le possibilità comprendono la codifica **crc32** e la codifica **izsam**. La selezione di **default** corrisponde a *crc32*;
- specificare lo schema per la specie microbica su cui è stato eseguito 4TY_cgMLST__chewbbaca, se vengono selezionati come input i profili allelici da cgMLST. Il campo è precompilato con lo schema appropriato alla specie batterica i cui codici sono presenti nel carrello, o nel tag attivo, usato per il lancio dell'analisi. Non è pertanto necessario alcun intervento ulteriore, da parte dell'utente.

Nota: di default GrapeTree usa, come input, l'output della **versione 2.8.5** del database di ChewBBACA, con codifica **crc32**. Allo scopo di mantenere retrocompatibilità con gli esami precedenti, rimane possibile eseguire GrapeTree sui vecchi profili calcolati con ChewBBACA 2.8.4; in tal caso è necessario selezionare **izsam** come codifica.

Si noti che tali associazioni devono essere rispettate affinché l'analisi venga effettuata correttamente. **Qualunque altra combinazione comporterà il fallimento dell'analisi**.

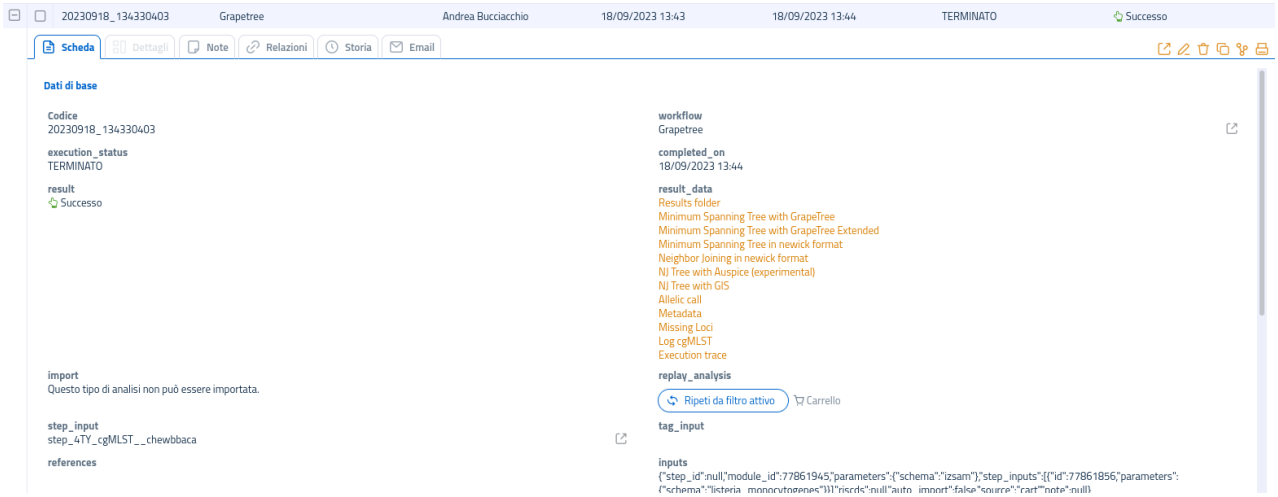
Si noti inoltre che, ai fini della validità delle analisi effettuate, è necessario usare la versione ChewBBACA 2.8.5.



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di verificare lo stato del processo. Il sistema invierà una notifica all'utente sia quando l'analisi viene lanciata con successo, sia al termine dell'esecuzione.

Nella scheda dell'analisi completata, oltre alla possibilità di esplorare la cartella di output, saranno presenti alcune opzioni come l'accesso ad alcuni dei file prodotti e la visualizzazione diretta degli alberi MSTree e NJ tramite l'integrazione di GrapeTree in piattaforma.

Per visualizzare l'albero sarà necessario navigare nella scheda del risultato di GrapeTree e selezionare il link **Minimum Spanning Tree with Grapetree**:



Nota: I risultati delle analisi multi-campione non possono essere importati. Si invita a sfruttare l'apposita funzionalità per la conservazione delle analisi non importabili o, in alternativa, l'opzione per il download diretto del file newick (`.nwk`) e del file dei metadati, in modo da poter conservare i file e visualizzare l'albero in qualunque momento, sia all'interno della piattaforma che con un software esterno.

Cartella dei risultati

Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output, **Result folder**, può essere esplorata cliccando sul link presente all'interno della scheda dell'analisi. All'interno della successiva cartella `Results` vi sono due sotto-cartelle:

- **meta** (*metadati*), in cui vengono salvati i file di log e di configurazione del processo eseguito;
- **result**, in cui sono salvati i file con i risultati prodotti dall'analisi.

La tabella in basso mostra la lista dei principali file di interesse presenti nelle cartelle, insieme ad alcune informazioni utili.

File	Descrizione	Posizione
cgMLST.nwk	file dell'albero in formato .nwk (newick) calcolato con algoritmo MSTree	cartella "result"
cgMLST.tsv	tabella .tsv degli alleli identificati da chewBBACA nel cgMLST (chiamata allelica). Righe: campioni; Colonne: *loci*.	cartella "result"
cgMLST_NJ.nwk	file dell'albero in formato .nwk (newick) calcolato con algoritmo NJ	cartella "result"
missing_loci.tsv	elenco dei *loci* non in comune tra i campioni e quindi scartati per la costruzione dell'albero	cartella "result"
cgMLST_dists_matrix.csv	matrice di distanza allelica da cgMLST in formato .csv	cartella "result"

I file in formato .nwk contengono tutti i dati dell'albero e possono essere visualizzati tramite GrapeTree o un altro software esterno per la visualizzazione dei dendrogrammi.

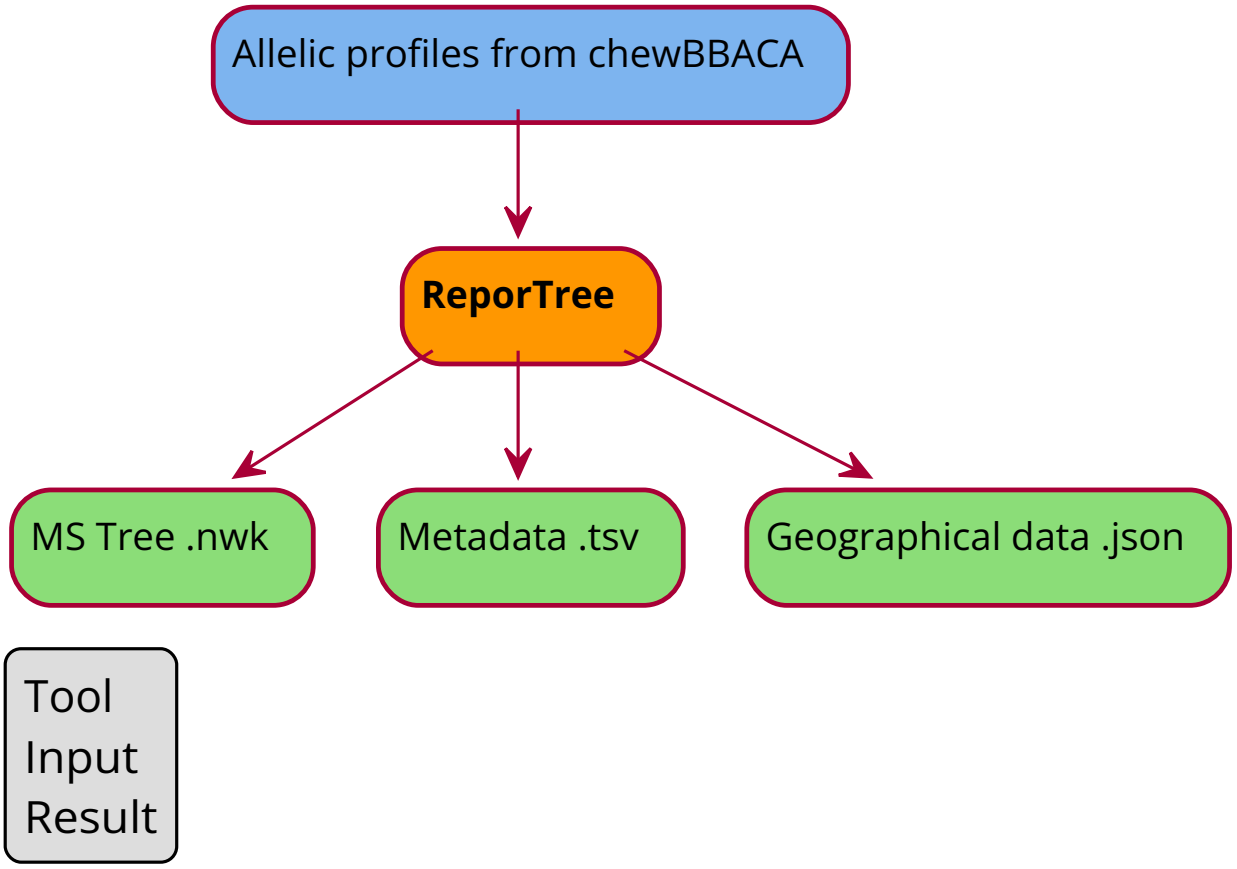
ReporTree

Introduzione

ReporTree costruisce grafici di distanza tra microrganismi a partire dai profili allelici dei campioni forniti. L'albero viene costruito con l'algoritmo "Minimum Spanning Tree" (MSTree o MST).

Pagina github di ReporTree https://github.com/insapathogenomics/ReporTree

Una volta calcolato l'albero, è possibile visualizzarlo direttamente in piattaforma (usando SPREAD) tramite il sistema integrato di visualizzazione dendrogrammi, accessibile dalla scheda dell'analisi completata.



Lancia analisi ReporTree

Una volta selezionata l'analisi **ReporTree** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma: non vi è necessità di scegliere il tool, in quanto l'analisi è specifica per l'uso di ReporTree.

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per ReporTree sono:

- step_4TY_wgMLST
- step_4TY_cgMLST

Per entrambi gli input possibili, è necessario che il software utilizzato per il calcolo dei profili allelici sia **chewBBACA**. Si ponga quindi attenzione nel selezionare tale software per il lancio dell'analisi 4TY_cgMLST o 4TY_wgMLST, se si intende lanciare ReporTree come analisi a valle.

Nella schermata di definizione parametri è possibile:

- selezionare la codifica: le possibilità comprendono la codifica **izsam** e la codifica **crc32**. La selezione di **default** corrisponde a crc32.
- specificare lo schema per la specie microbica su cui è stato eseguito 4TY_cgMLST__chewbbaca, se vengono selezionati come input i profili allelici da cgMLST. Il campo è precompilato con lo schema appropriato alla specie batterica i cui codici sono presenti nel carrello o tag attivo, usato per il lancio dell'analisi. Non è pertanto necessario alcun intervento ulteriore, da parte dell'utente.

Nota: di default ReporTree usa come input l'output della **versione 2.8.5** del database di ChewBBACA, con encoding **crc32**. Allo scopo di mantenere retrocompatibilità con gli esami precedenti, rimane possibile eseguire ReporTree sui vecchi profili calcolati con ChewBBACA 2.8.4; in tal caso è necessario selezionare **izsam** come encoding.

Si noti che tali associazioni devono essere rispettate perchè l'analisi venga effettuata correttamente. **Qualunque altra combinazione comporterà il fallimento dell'analisi**.

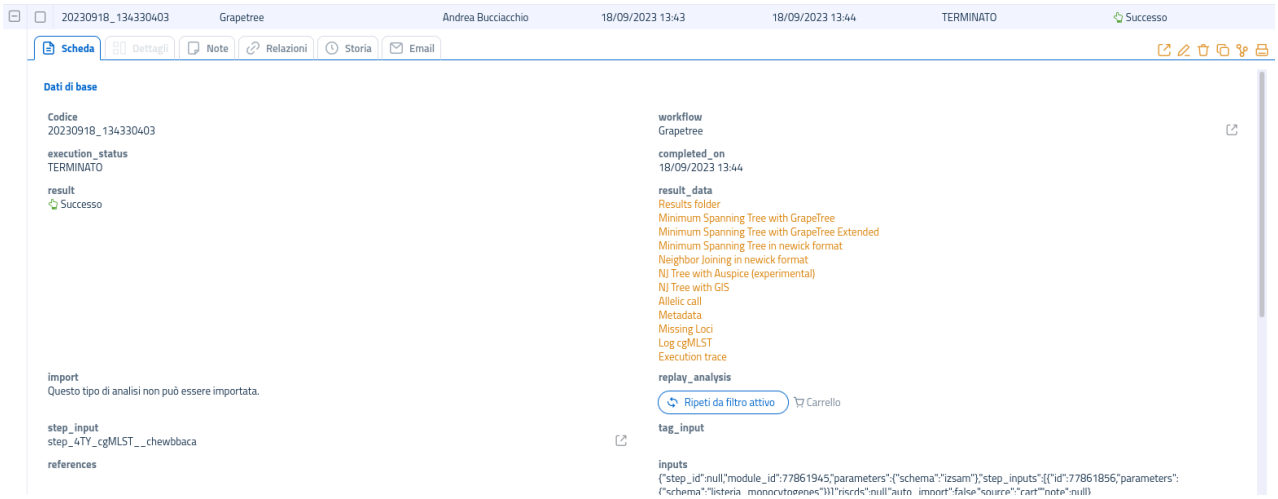
Si noti inoltre che, ai fini della validità delle analisi effettuate, è necessario usare ChewBBACA 2.8.5.

La scheda per la definizione dei parametri offre un elevato livello di personalizzazione, con campi in cui è possibile specificare i valori desiderati. L'ultimo campo, **Custom parameters**, permette di aggiungere ulteriori argomenti, elencati come *flags*, in riga di comando, in UNIX-style o GNU-style. Per maggiori informazioni in merito a ReporTree e al suo utilizzo, si invita a consultare la guida ufficiale di ReporTree.

Lancia analisi ReporTree
Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di verificare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Nella scheda dell'analisi completata saranno presenti alcune opzioni, oltre alla possibilità di esplorare la cartella di output, come l'accesso diretto ad alcuni dei files di log e metadati, l'apertura del file nwk e la visualizzazione diretta dell'albero MSTree tramite l'integrazione di GrapeTree in piattaforma.

Per visualizzare l'albero sarà necessario navigare nella scheda del risultato di GrapeTree e selezionare il link **Minimum Spanning Tree with Grapetree**:



Nota: I risultati delle analisi multi-campione non possono essere importati. Si invita a sfruttare l'apposita funzionalità per la conservazione delle analisi non importabili o, in alternativa, l'opzione per il download diretto del file newick (`.nwk`) e del file dei metadati, in modo da poter conservare i file e visualizzare l'albero in qualunque momento, sia all'interno della piattaforma che con un software esterno.

Cartella dei risultati

Per consultare la guida sul download dei file dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results` è possibile trovare 2 sotto-cartelle:

- **meta**: ("metadata") in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

La tabella in basso presenta la lista dei principali file di interesse presenti nelle cartelle, insieme ad alcune informazioni utili.

File	Descrizione	Posizione
gt_clusterComposition.tsv	Tabella con elenco dei cluster e loro composizione	cartella "result"
gt_dist_grapetree.tsv	Matrice di distanza	cartella "result"
reportree_metadata.tsv	Tabella dei metadati dei campioni	cartella "result"
gt_metadata_w_partitions.tsv	Tabella dei metadati dei campioni + cluster di appartenenza in base alle partizioni	cartella "result"
gt.nwk	Albero MST in formato newick	cartella "result"
gt_dist_hamming.tsv	Tabella con le distanze di Hamming	cartella "result"
result_alleles_all.tsv	Tabella delle chiamate alleliche	cartella "result"
gt_dist.tsv	Elenco delle distanze *pairwise* tra campioni	cartella "result"
gt_partitions_summary.tsv	Elenco delle partizioni dell'albero e relativi campioni nei cluster	cartella "result"
gt_loci_report.tsv	Report sulle statistiche dei *loci* usati	cartella "result"
gt_partitions.tsv	Tabella dei campioni e relativa appartenenza a cluster in base al partizionamento	cartella "result"
gt_loci_used.txt	Lista dei *loci*	cartella "result"

Il file in formato newick (.nwk) contiene tutti i dati dell'albero e può essere visualizzato tramite **SPREAD** o un altro software esterno per la visualizzazione di dendrogrammi o grafici di distanza.

Pipeline Surveillance

Introduzione

La pipeline *Surveillance* esegue ReporTree usando come campioni di interesse quelli specificati come input; inoltre, la matrice di distanza e l'albero sono calcolati usando tutti i campioni della specie indicata presenti in piattaforma.

La pipeline "Surveillance" è automatizzata per la sorveglianza di *Listeria monocytogenes*: essa viene lanciata automaticamente dalla piattaforma ogniqualvolta viene calcolato un nuovo profilo allelico di *L. monocytogenes*.

Ciò include: - i profili allelici calcolati al termine della pipeline NGSmanager sia *automatica*, al caricamento di nuovi campioni sulla piattaforma, sia lanciata *manualmente*; - i lanci manuali dell'analisi 4TY_cgMLST.

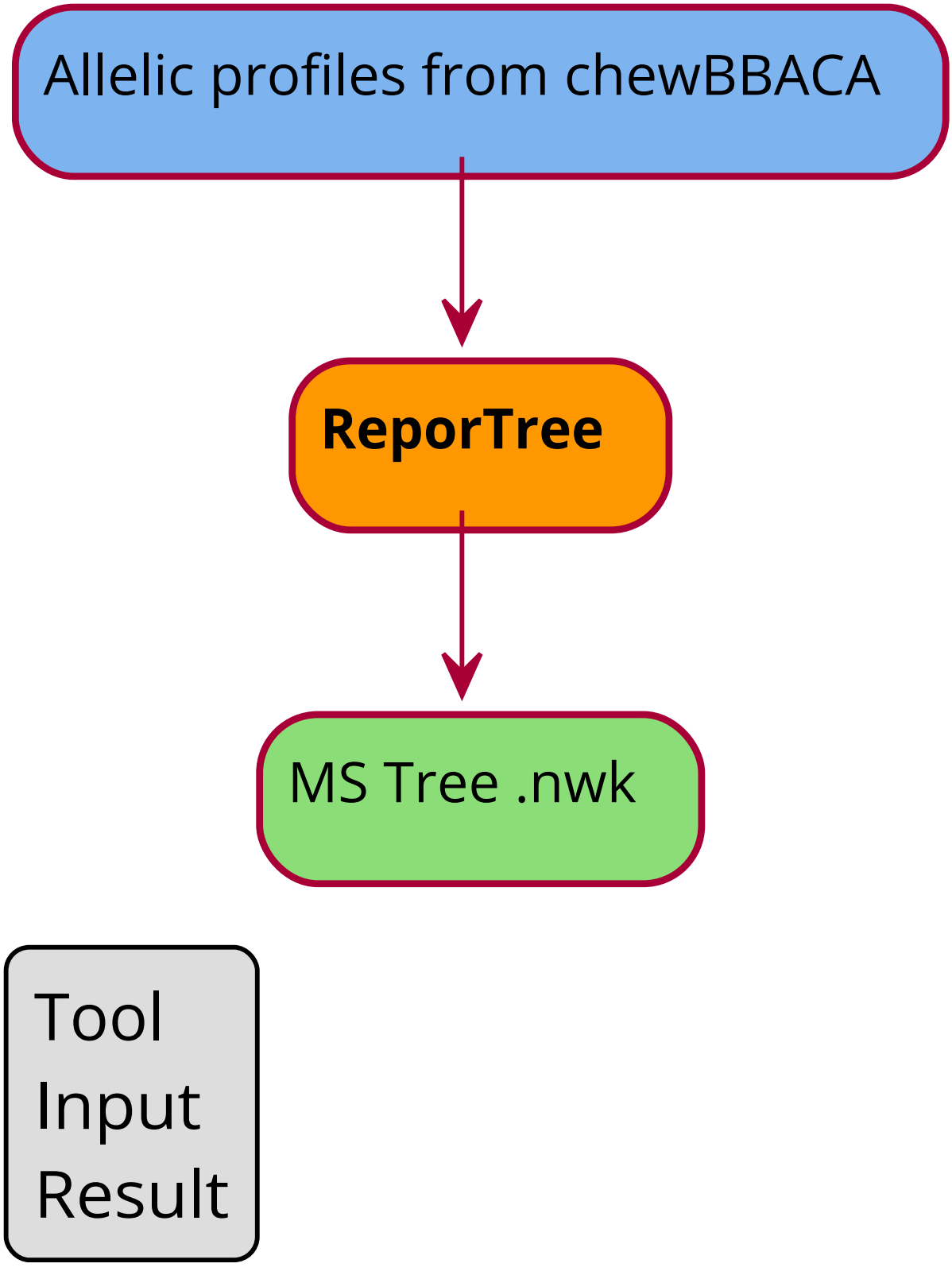
Surveillance può essere lanciata manualmente dal menù **Lancia analisi**: - da utenti IZS Listeria e LNR Listeria per il lancio su specie *L. monocytogenes*; - da utenti IZS Campylobacter per il lancio su specie *Campylobacter coli* e *Campylobacter jejuni*.

Possono essere usati profili allelici cgMLST e wgMLST.

Pagina GitHub di ReporTree https://github.com/insapathogenomics/ReporTree

Citazione: *Mixão V., Pinto M., Sobral D., Di Pasquale A., Gomes J.P., Borges V., (2023), 'ReporTree: a surveillance-oriented tool to strengthen the linkage between pathogen genetic clusters and epidemiological data'. Genome Medicine. doi: 10.1186/s13073-023-01196-1*

Per ogni uso e riferimento riguardo le opzioni del tool, si invita a consultare la pagina Wiki di ReporTree.



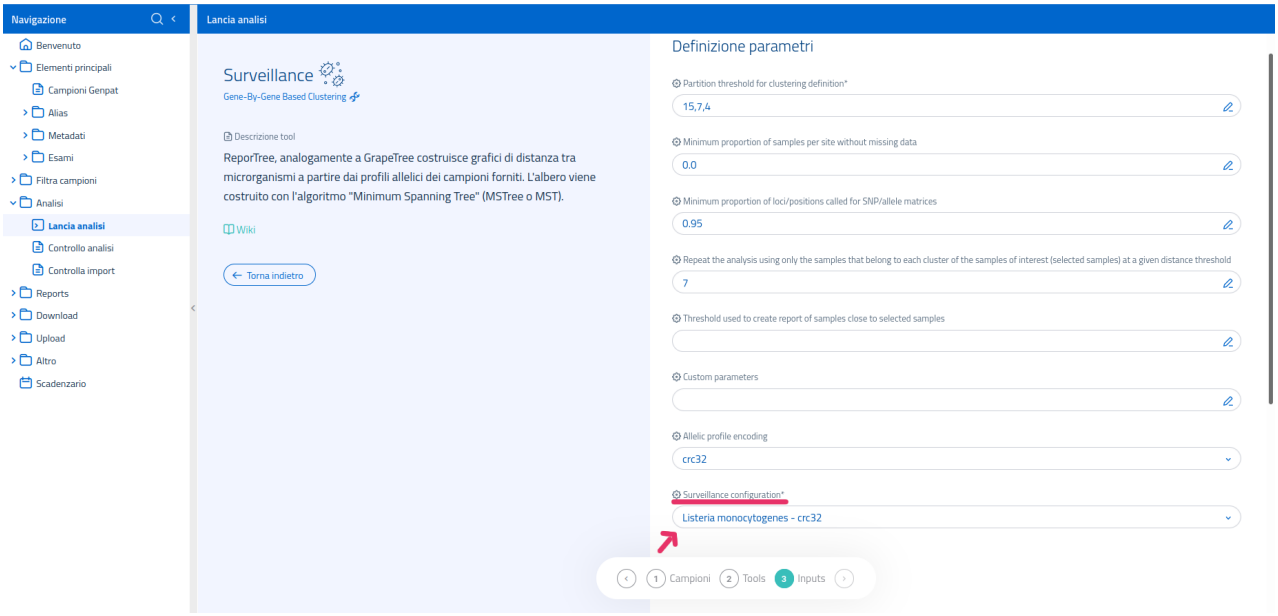
Lancia pipeline Surveillance

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Surveillance** nella pagina lancia di analisi, il sistema passerà ad un'interfaccia di conferma.

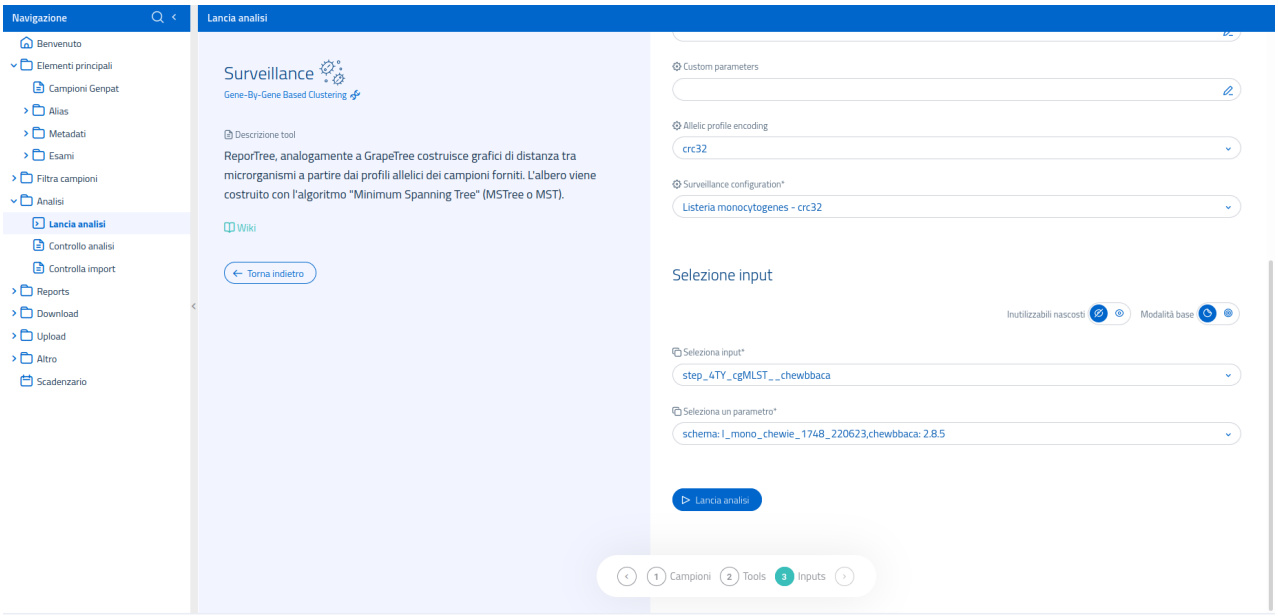
La pipeline **Surveillance** eseguirà ReporTree con le seguenti variazioni:

- Mentre un normale lancio di ReporTree viene eseguito esclusivamente sui campioni presenti nel carrello, il lancio di ReporTree tramite pipeline *Surveillance* viene eseguito su tutti i campioni presenti in piattaforma, appartenenti alla specie indicata nell'apposito campo (immagine **A**).
- Un normale lancio di ReporTree richiede di specificare i codici dei campioni di interesse tra quelli usati come input nel carrello o nel tag; nel lancio di ReporTree, tramite pipeline Surveillance, l'input è costituito **solo dai campioni di interesse** (ovvero, nel carrello attivato o nel tag attivo devono essere presenti *solo* tali campioni) che, quindi, non dovranno essere esplicitati durante la definizione dei parametri.

A



B



Poichè la pipeline usa ReporTree, gli input sono gli stessi:

- step_4TY_wgMLST
- step_4TY_cgMLST

Per ogni altro dettaglio si faccia riferimento alle sezioni pertinenti dell'analisi ReporTree e alla guida ufficiale del tool:

- Parametri per il lancio
- Cartella dei risultati
- Guida ufficiale di ReporTree

ReporTree VCF

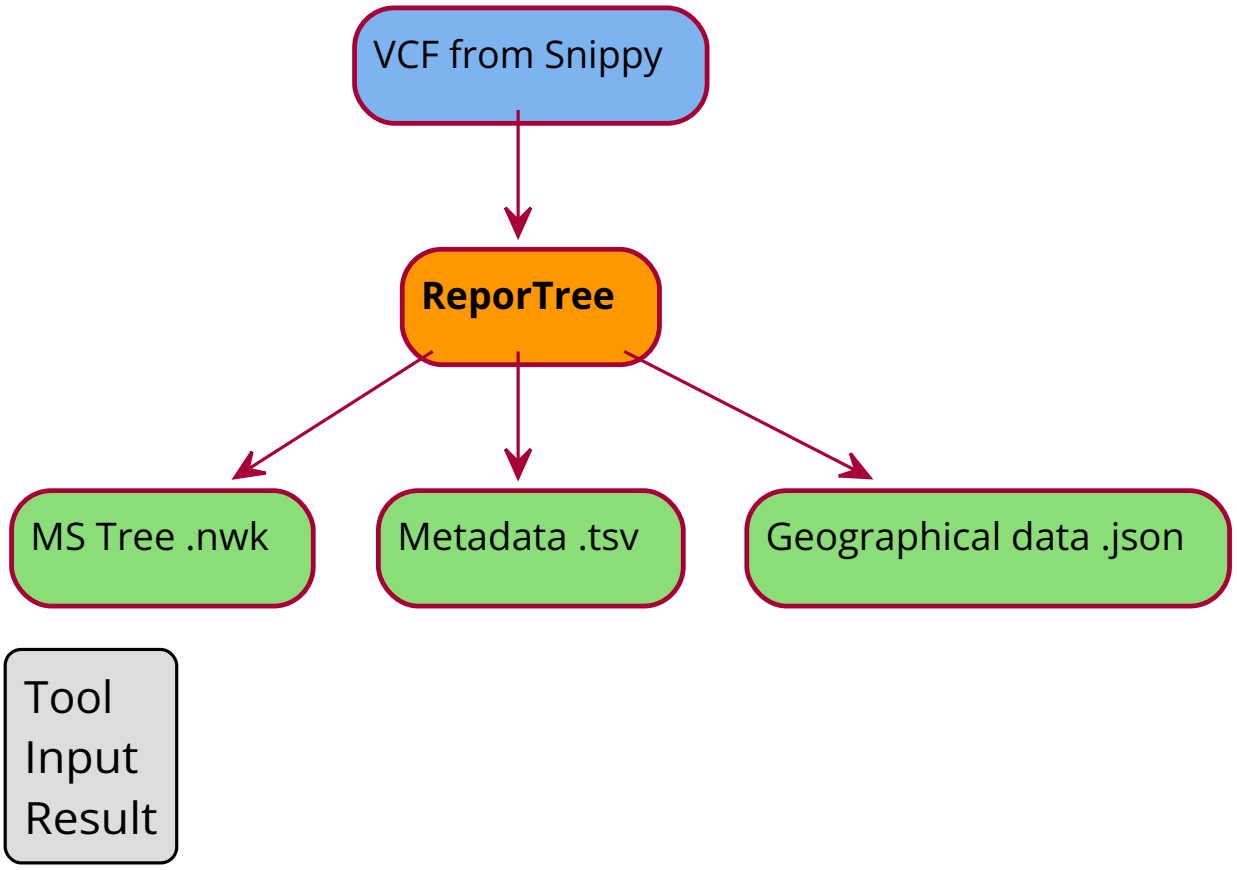
Introduzione

ReporTree-VCF è un'implementazione di ReporTree che permette all'utente di usare come input il formato *variant calling* (VCF), disponibile per ReporTree, per produrre una matrice ed un grafico di distanza tra microrganismi.

L'albero viene costruito con l'algoritmo "Minimum Spanning Tree" (MSTree o MST).

Pagina github di ReporTree: https://github.com/insapathogenomics/ReporTree

Una volta calcolato l'albero, è possibile visualizzarlo direttamente in piattaforma (usando SPREAD) tramite il sistema integrato di visualizzazione dendrogrammi, accessibile dalla scheda dell'analisi completata.



Lancia analisi ReporTree-VCF

Il file di input VCF può essere prodotto con l'analisi 2AS_mapping eseguita con **Snippy** o **Ivar**.

Si noti che, nonostante anche la pipeline Snippy-core produca il file VCF, trattandosi di un'analisi multicampione, non è possibile importarla; di conseguenza, non si possono usare i file da essa prodotti come input per ReporTree-VCF.

Per ogni altro uso e caratteristica, l'analisi ed il procedimento sono identici alle altre istanze di ReporTree in piattaforma. Si rimanda quindi alla pagina principale per ReporTree.

ReporTree MA

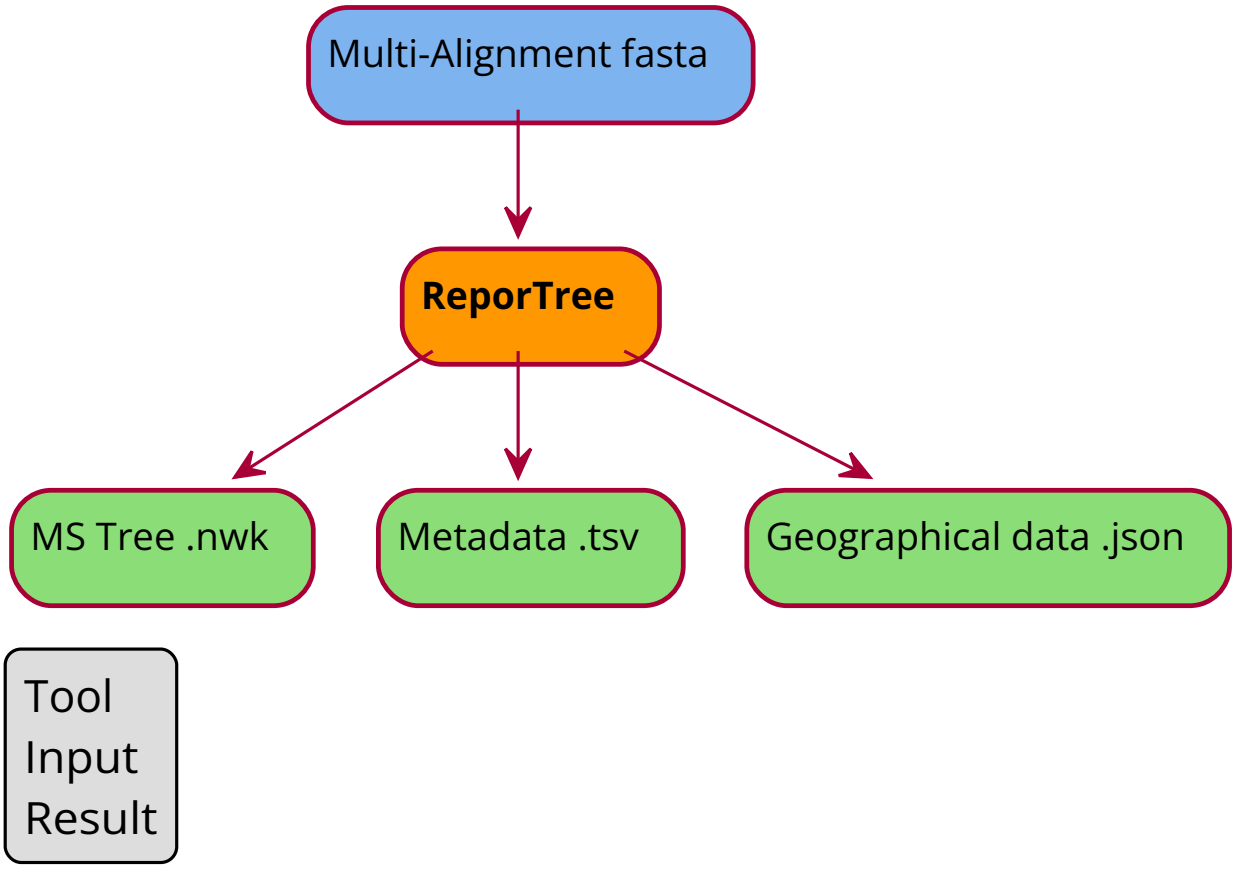
Introduzione

ReporTree-MA è un'implementazione di ReporTree che permette all'utente di usare, come input, il file FASTA del multiallineamento, disponibile per ReporTree, per produrre una matrice ed un grafico di distanza tra microrganismi.

L'albero viene costruito con l'algoritmo "Minimum Spanning Tree" (MSTree o MST).

Pagina github di ReporTree: https://github.com/insapathogenomics/ReporTree

Una volta calcolato l'albero, è possibile visualizzarlo direttamente in piattaforma (usando SPREAD) tramite il sistema integrato di visualizzazione dendrogrammi, accessibile dalla scheda dell'analisi completata.



Il file di multiallineamento viene prodotto direttamente da ReporTree usando, come input, i singoli file FASTA dei campioni. I possibili input sono, quindi:

- step_2AS_mapping
- step_2AS_denovo
- step_2AS_hybrid
- step_2AS_import

Per ogni altro uso e caratteristica, l'analisi ed il procedimento sono identici alle altre istanze di ReporTree disponibili in piattaforma. Si rimanda, quindi, alla pagina principale per ReporTree.

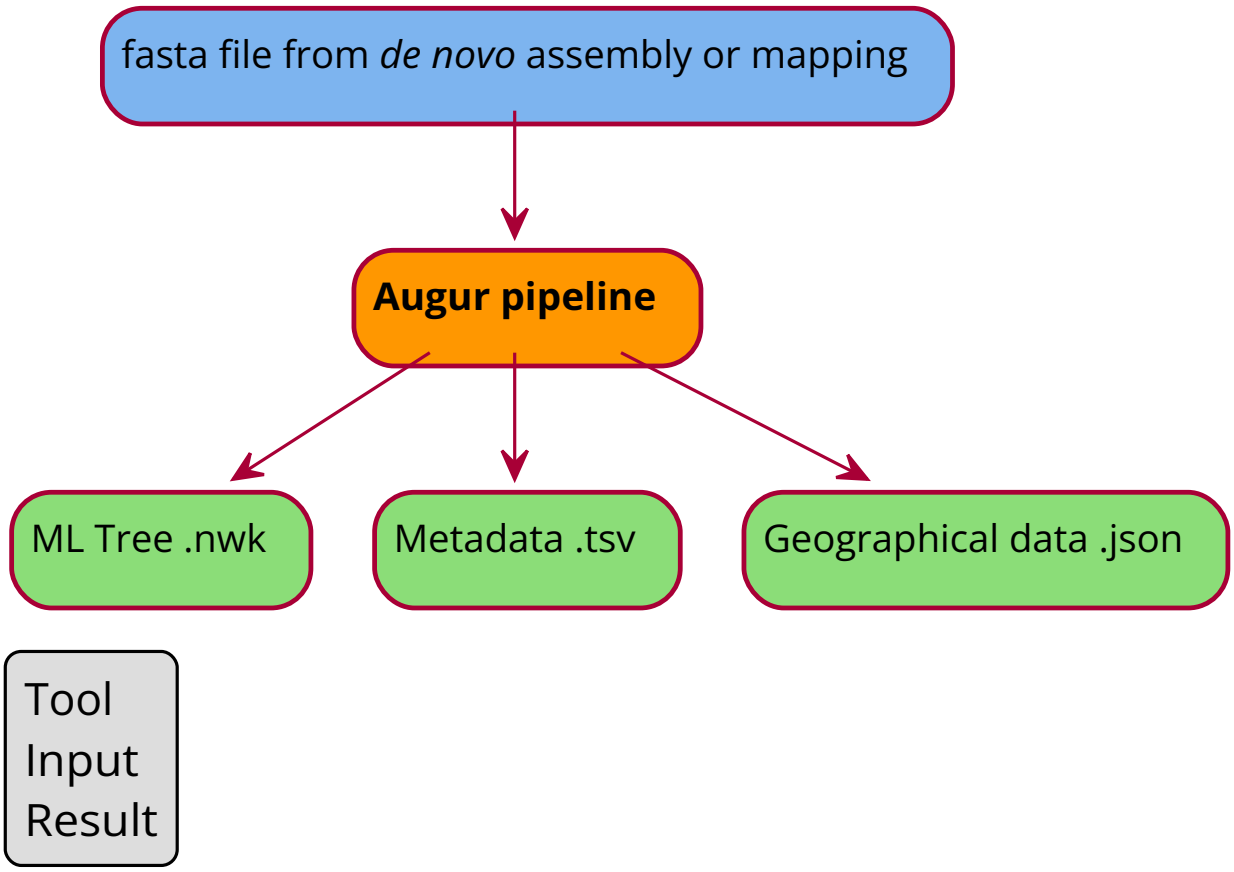
Augur

Introduzione

L'integrazione in piattaforma di "*Augur: A bioinformatics toolkit for phylogenetic analysis*" permette di eseguire la **pipeline Augur** per produrre un albero filogenetico (con algoritmo Maximum Likelihood - ML), a partire dai file FASTA del mapping o *de novo* assembly.

Pagina github di Augur: https://github.com/nextstrain/augur/tree/7cb384877cb61e26781dec947f81fee29462c008
Pagina di documentazione ufficiale di Augur: https://docs.nextstrain.org/projects/augur/en/stable/index.html

Una volta calcolato l'albero, è possibile visualizzarlo direttamente in piattaforma (usando Auspice), tramite il sistema integrato di visualizzazione dendrogrammi, accessibile dalla scheda dell'analisi completata.



Lancia analisi Augur

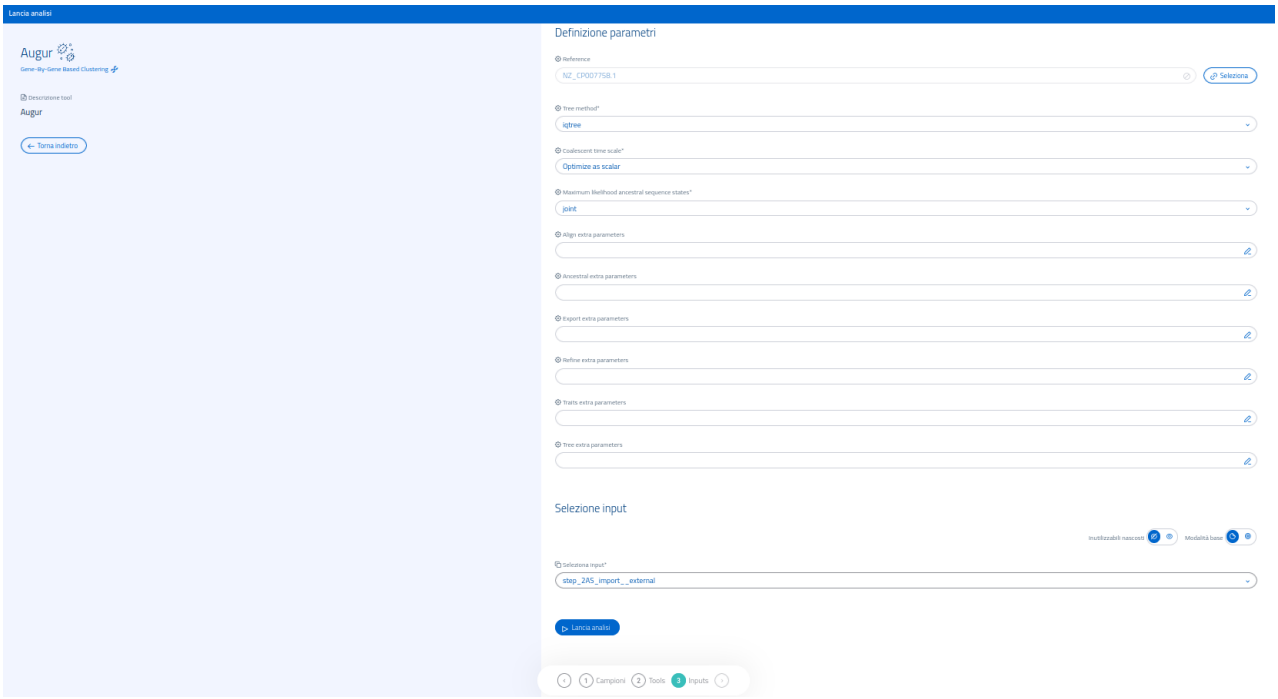
Una volta selezionata l'analisi **Augur** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma: non vi è necessità di scegliere il tool, in quanto l'analisi è specifica per l'uso di Augur.

L'interfaccia per la selezione dell'input mette a disposizione la modalità avanzata di selezione input, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

I possibili input utilizzabili per Augur comprendono i risultati di tutte le analisi di mapping o *de novo* assembly.

Nella schermata di definizione dei parametri è possibile:

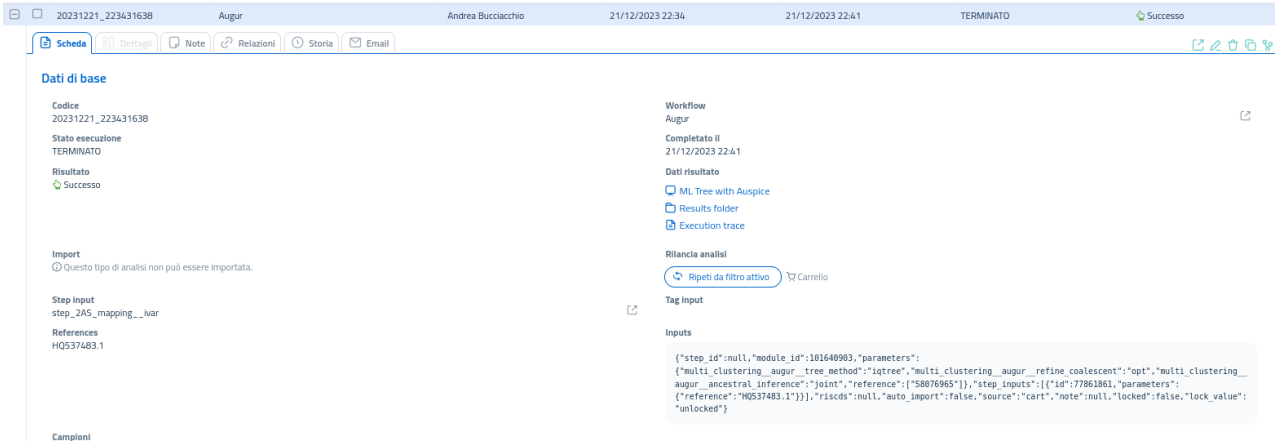
- selezionare il reference da usare come radice dell'albero;
- scegliere il software con cui costruire il file .nwk (iqtree, fasttree o raxm);
- scegliere una modalità per la scala temporale;
- selezionare una modalità per l'algoritmo Maximum Likelihood nel trattare la sequenza ancestrale (*reference*);
- aggiungere parametri addizionali come testo (i campi riconosceranno i parametri per il lancio del tool da terminale. Consultare la guida ufficiale di Augur per l'elenco dei parametri disponibili).



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di verificare lo stato del processo. Il sistema invierà una notifica all'utente sia quando l'analisi viene lanciata con successo, sia al termine dell'esecuzione.

Nella scheda dell'analisi completata, oltre alla possibilità di esplorare la cartella di output, saranno presenti alcune opzioni come l'accesso diretto ad alcuni dei file di log e metadati, l'apertura del file .nwk e la *visualizzazione diretta* dell'albero Maximumm Likelihood tramite l'integrazione in piattaforma di Auspice.

Per visualizzare l'albero sarà necessario navigare nella scheda risultato di GrapeTree e selezionare il link **ML Tree with Auspice**:



Nota: I risultati delle analisi multi-campione non possono essere importati. Si invita a sfruttare l'apposita funzionalità per la conservazione delle analisi non importabili o, in alternativa, l'opzione per il download diretto del file newick (`.nwk`) e del file dei metadati, in modo da poter conservare i file e visualizzare l'albero in qualunque momento, sia all'interno della piattaforma sia attraverso un software esterno.

Cartella dei risultati

Per consultare la guida sul download dei file dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output, **Result folder**, può essere esplorata cliccando sul link presente all'interno della scheda dell'analisi. All'interno della successiva cartella `Results` vi sono due sotto-cartelle:

- **meta** (*metadati*), in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**, in cui sono salvati i file con i risultati prodotti dall'analisi.

La tabella in basso mostra la lista dei principali file di interesse presenti nelle cartelle, insieme ad alcune informazioni utili.

File	Descrizione	Posizione
aa_muts.json	file json usato da Auspice per le mutazioni negli amminoacidi	cartella "result"
alignment.fasta	FASTA del multiallineamento	cartella "result"
augur_metadata.tsv	tabella dei metadati in formato tsv prodotta da Augur	cartella "result"
auspice.json	file json usato da Auspice per la costruzione delle corrispondenze tra foglie del dendrogramma e nodi sulla mappa	cartella "result"
branch_lengths.json		cartella result
nt_muts.json	file json usato da Auspice per le mutazioni nella sequenza nucleotidica	cartella "result"
traits.json	file json usato da Auspice per popolare i campi dei metadati	cartella "result"
tree.nwk	file dell'albero in formato .nwk (newick) calcolato con algoritmo Maximum Likelihood	cartella "result"
metadata.csv	tabella dei metadati in formato csv	cartella "result"

Il file in formato nwk contiene tutti i dati dell'albero e può essere visualizzato tramite Auspice o un altro software esterno per la visualizzazione di dendrogrammi.

Pangenome extraction

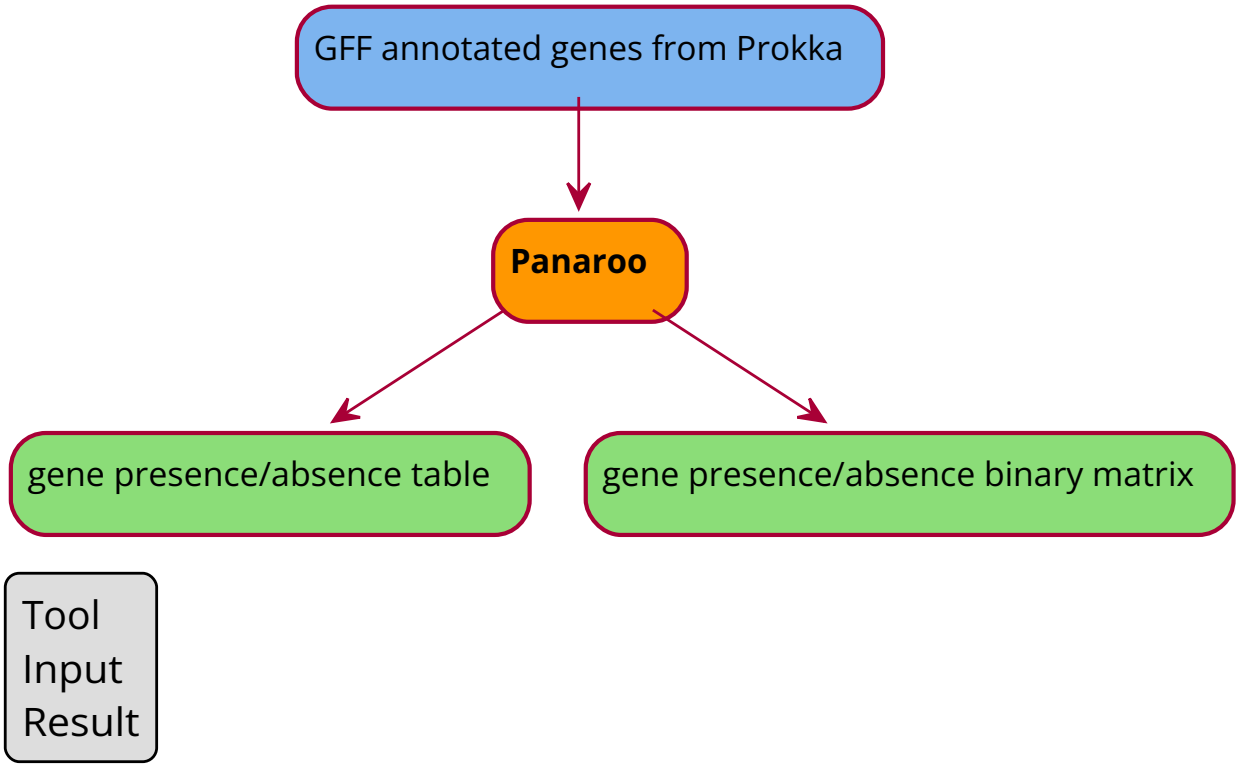
Panaroo

Introduzione

Panaroo elabora il pan-genome in matrici di presenza/assenza dei geni annotati nei campioni, a partire dall'annotazione funzionale del genoma, effettuata da Prokka.

Per approfondimenti sul funzionamento di Panaroo si rimanda alla guida ufficiale.

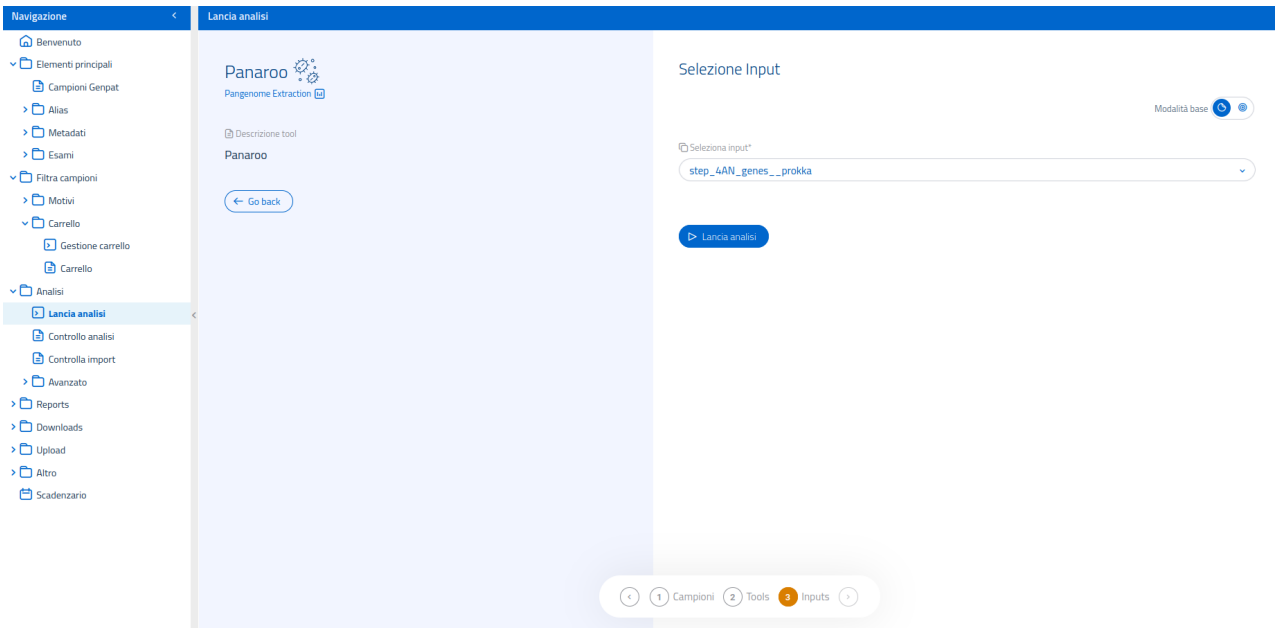
Pagina GitHub di Panaroo: https://github.com/gtonkinhill/panaroo



Lancia analisi Panaroo

Una volta selezionata l'analisi **Panaroo** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma: non vi è necessità di scegliere il tool, in quanto l'analisi è specifica per l'uso di *Panaroo - An updated pipeline for pangenome investigation.*

Nell'interfaccia per la selezione dell'input è possibile confermare l'input per il programma: Panaroo è progettato per l'elaborazione degli output di Prokka; è necessario quindi usare come input i file di annotazione da 4AN_genes. Poichè non vi è possibilità di scelta, il campo è precompilato e non è necessario alcun intervento da parte dell'utente.



Dopo aver lanciato l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di verificare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella dei risultati

> Per consultare la guida sul download dei file dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results` è possibile trovare 2 sotto-cartelle:

- **meta**: ("metadata") in cui vengono salvati i file di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i file con i risultati prodotti dall'analisi.

La tabella in basso presenta la lista dei principali file generati da Panaroo, insieme ad alcune informazioni utili.

Ulteriori informazioni sui file di output di Panaroo sono reperibili nella guida ufficiale agli output di Panaroo.

File	Descrizione	Posizione
combined_DNA_CDS.fasta	file FASTA con le sequenze nucleotidiche dei geni annotati e di quelli identificati da Panaroo	cartella "result"
combined_protein_CDS.fasta	file fasta con le sequenze aminoacidiche prodotte dai geni annotati e di quelli identificati da Panaroo	cartella "result"
combined_protein_GFF3cdhit_out.txt	log della fase CD-HIT di Panaroo	cartella "result"
combined_protein_cdhit_out.txt.clstr	informazioni sui cluster da CD-HIT	cartella "result"
core_alignment_header.embl		cartella "result"
core_gene_alignment.aln	file di allineamento	cartella "result"
final_graph.gml	grafico del pan-genome	cartella "result"
gene_data.csv	tabella .csv delle corrispondenze tra sequenze dei geni annotati e i codici usati internamente da Panaroo	cartella "result"
gene_presence_absence.Rtab	matrice binaria di presenza/assenza dei geni nei campioni	cartella "result"
gene_presence_absence.csv	file .csv che descrive la presenza dei geni nei campioni	cartella "result"
gene_presence_absence_roary.csv	file .csv che descrive la presenza dei geni nei campioni, con più informazioni (sul modello di Roary)	cartella "result"
pan_genome_reference.fa	*reference pan-genome*, in formato FASTA, per tutti i geni rinvenuti nel dataset	cartella "result"
pre_filt_graph.gml	grafico del pan-genome grezzo	cartella "result"
struct_presence_absence.Rtab	matrice binaria che descrive la presenza/assenza di eventi di riarrangiamento del genoma	cartella "result"
summary_statistics.txt	file riassuntivo delle statistiche	cartella "result"

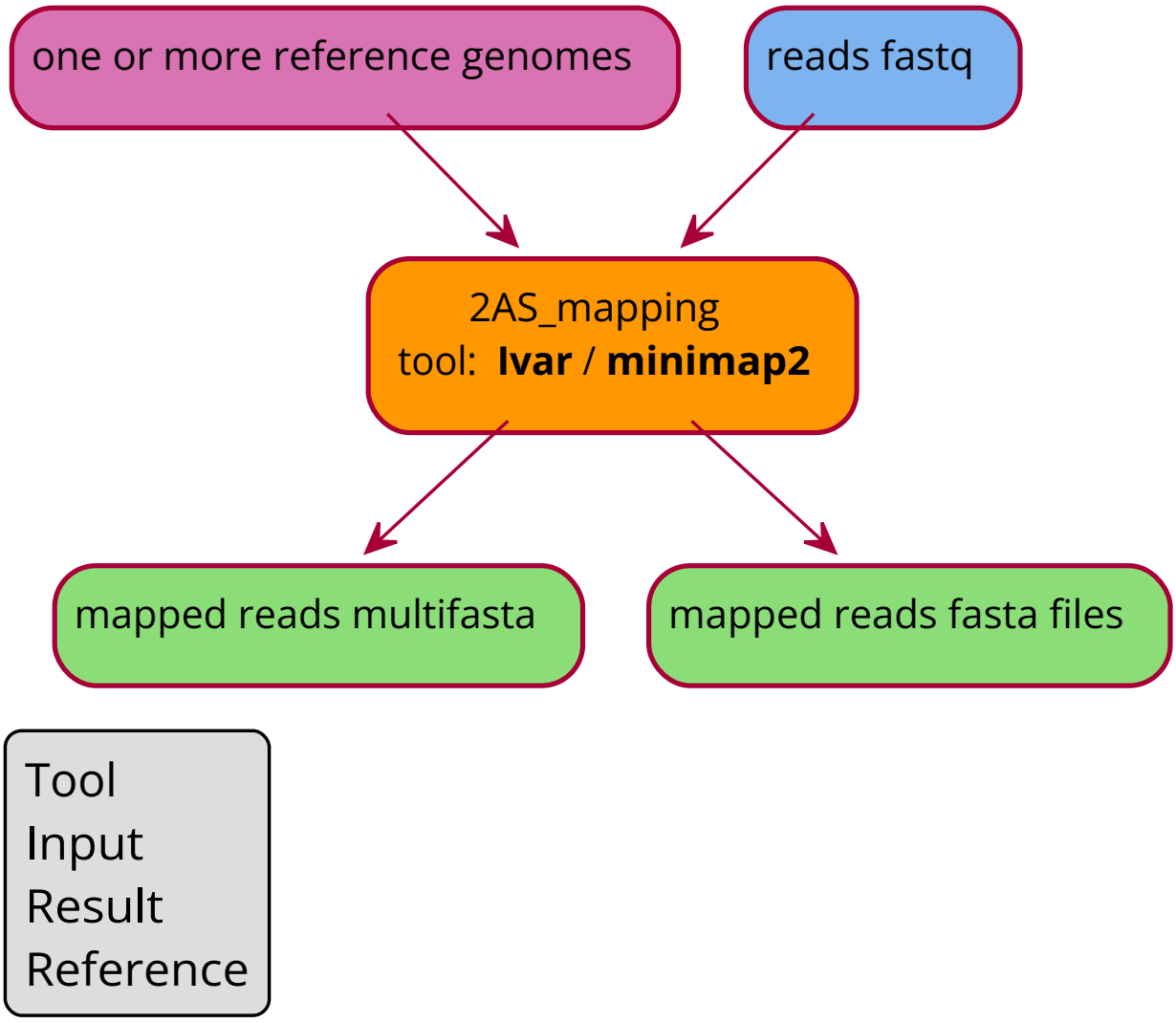
Per la visualizzazione del grafico prodotto da Panaroo, gli autori del programma consigliano l'uso di Cytoscape. Ulteriori informazioni sulla visualizzazione del grafico del pan-genome sono disponibili nella sezione della documentazione ufficiale di Panaroo.

5.4.5 Pipelines

Pipeline Mapping Virus Segmentati

Introduzione

La pipeline Mapping Virus Segmentati permette di eseguire il mapping dei segmenti di genoma dei virus segmentati su più *references*. Questo permette di eseguire il mapping di ogni segmento sul reference appropriato. La pipeline produce, oltre ai singoli allineamenti, anche il multifasta completo.



Lancia Pipeline Mapping Virus Segmentati

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Mapping Virus Segmentati** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

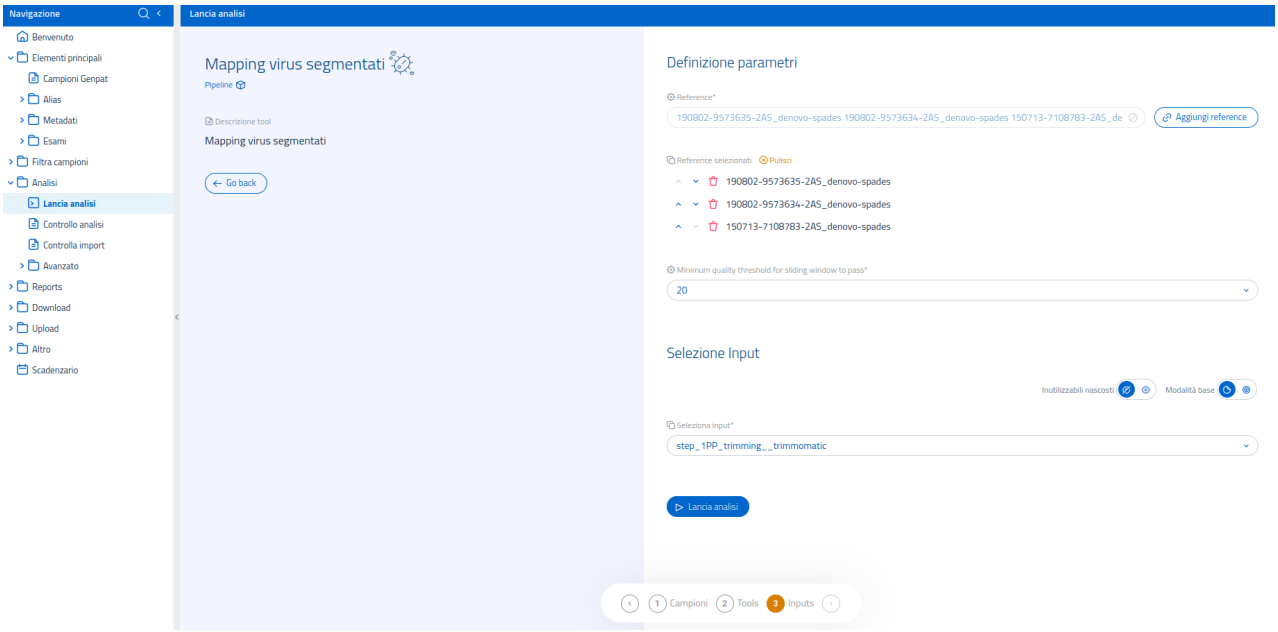
La pipeline **Mapping Virus Segmentati** esegue il mapping di tutti i segmenti di genoma da virus segmentati forniti, usando più *references*. I software usati saranno:

- **Ivar** (2AS_mapping__ivar) nei casi di Illumina *short reads* o IonTorrent
- **minimap2** (2AS_mapping__minimap2) per *long reads* Nanopore

Gli input utilizzabili sono i risultati delle analisi di preprocessamento:

- *reads* deplete (step_1PP_hostdepl)
- *downsampled reads* (step_1PP_downsampling)
- *reads* trimmate (step_1PP_trimming)
- *reads* filtrate (step_1PP_filtering)

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Reference multipli

I reference multipli richiesti dalla pipeline devono essere forniti dall'utente tramite il sistema di selezione dedicato, la cui guida all'uso è consultabile nell'apposita sezione della Wiki sul lancio di analisi.

Risultati

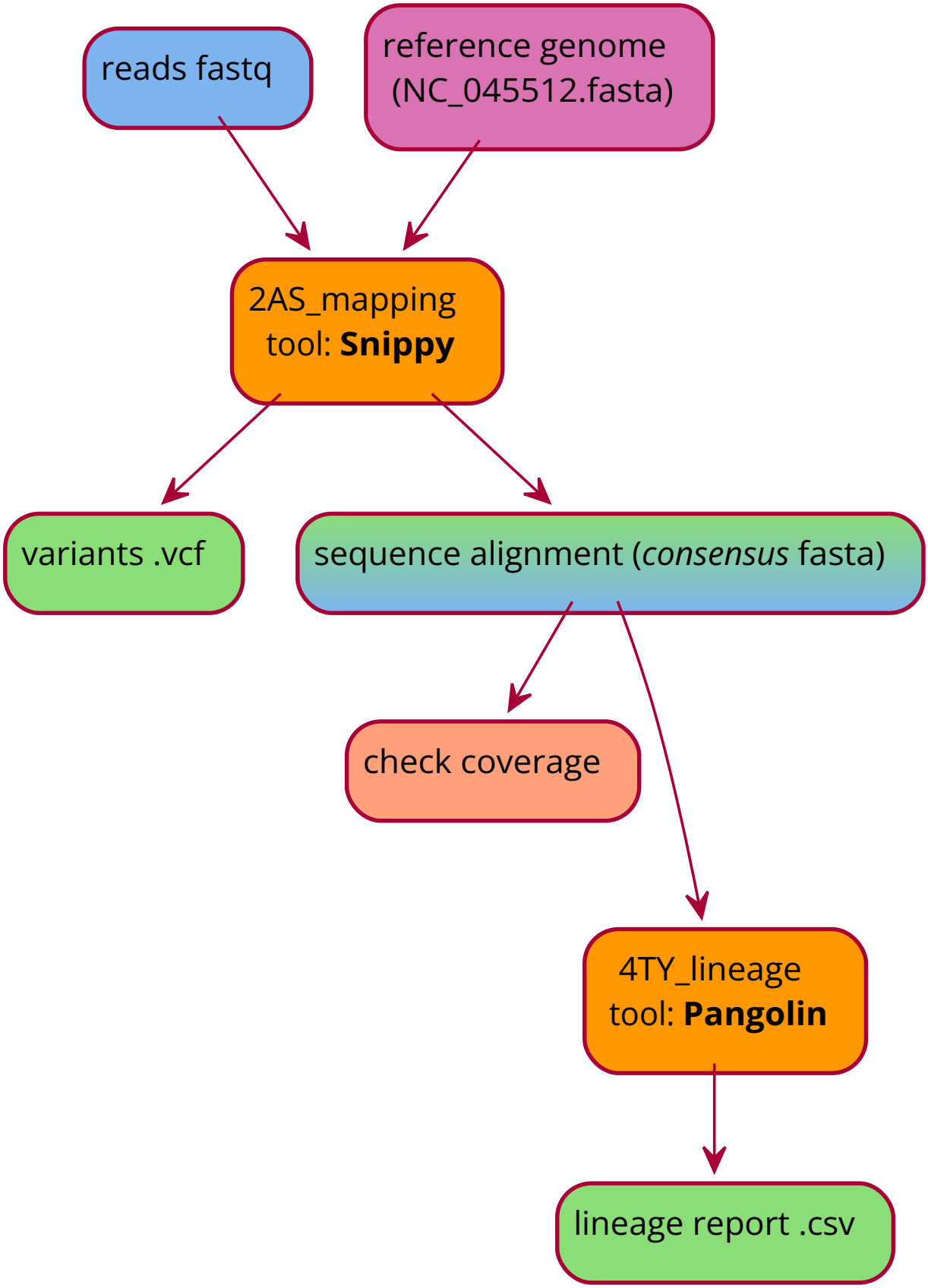
I files prodotti dalla pipeline "Mapping Virus Segmentati" saranno gli stessi prodotti dalle analisi che la compongono, organizzati con la stessa gerarchia. Come riferimento si rimanda quindi alla relativa sezione dell'analisi 2AS_mapping. A tali files si aggiunge il multifasta dei segmenti di genoma mappati, le cui informazioni sono riportate nella seguente tabella:

File	Descrizione	Posizione
`DSXXX-DTXXX_20XX.XX.XXXX.X_tool_aggregate.fasta`	file multifasta con tutte le sequenze dei segmenti mappate ai *references*	`DSXXX-DTXXX_tool > result`
`DSXXX-DTXXX_20XX.XX.XXXX.X_tool_reference.fasta`	fasta con la consensus generata dal mapping su un singolo *reference*. I risultati di ogni mapping singolo è salvato in una cartella identificata da un numero sequenziale addizionale, aggiunto al codice base della cartella	`DSXXX-DTXXX[sequential-number]_tool > result`
`DSXXX-DTXXX_20XX.XX.XXXX.X_tool_reference.bam` e `.bam.bai`	`bam` e relativo indice generati dal mapping su un singolo *reference*. I risultati sono salvati come nel caso del fasta.	`DSXXX-DTXXX[sequential-number]_tool > result`
warnings.log	file di testo con avvertenze nel caso di numero di *references* usato non corretto	cartella principale

Pipeline Covid Emergency

Introduzione

La pipeline **Covid Emergency** permette di effettuare rapidamente l'assembly con reference (2AS_mapping) e l'assegnazione del *lineage* (4TY_lineage), per campioni di SARS-CoV2. Il genoma reference che viene utilizzato per il mapping è sempre NC_045512 (Wuhan-Hu-1). Poichè il tool usato per il mapping è Snippy, la pipeline produce anche il VCF delle varianti.



Tool
Input
Result
Reference

Lancia Pipeline Covid Emergency

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Covid Emergency** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

La pipeline Covid Emergency è composta da 2 fasi:

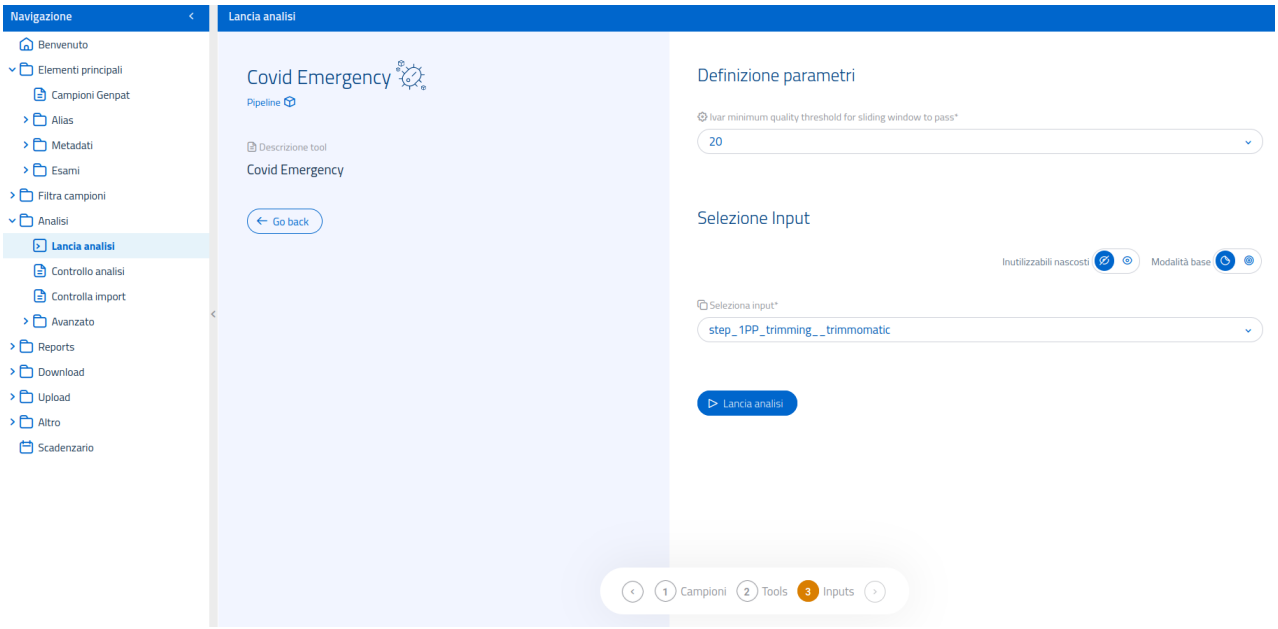
1. Mapping eseguito con **Snippy** (2AS_mapping__snippy);
2. Assegnazione del lineage eseguita con **Pangolin** (4TY_lineage__pangolin).

Gli input utilizzabili sono pertanto gli stessi disponibili per il lancio di un mapping con Snippy, ovvero le reads in formato fastq:

- step_1PP_hostdepl
- step_1PP_downsampling
- step_1PP_trimming

L'interfaccia per la selezione dell'input mette a disposizione, anche in questo caso, la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

Nella sezione dedicata alla definizione dei parametri, il campo per la selezione della soglia di qualità permette di specificare un valore minimo di qualità (20 o 14) delle basi predette.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

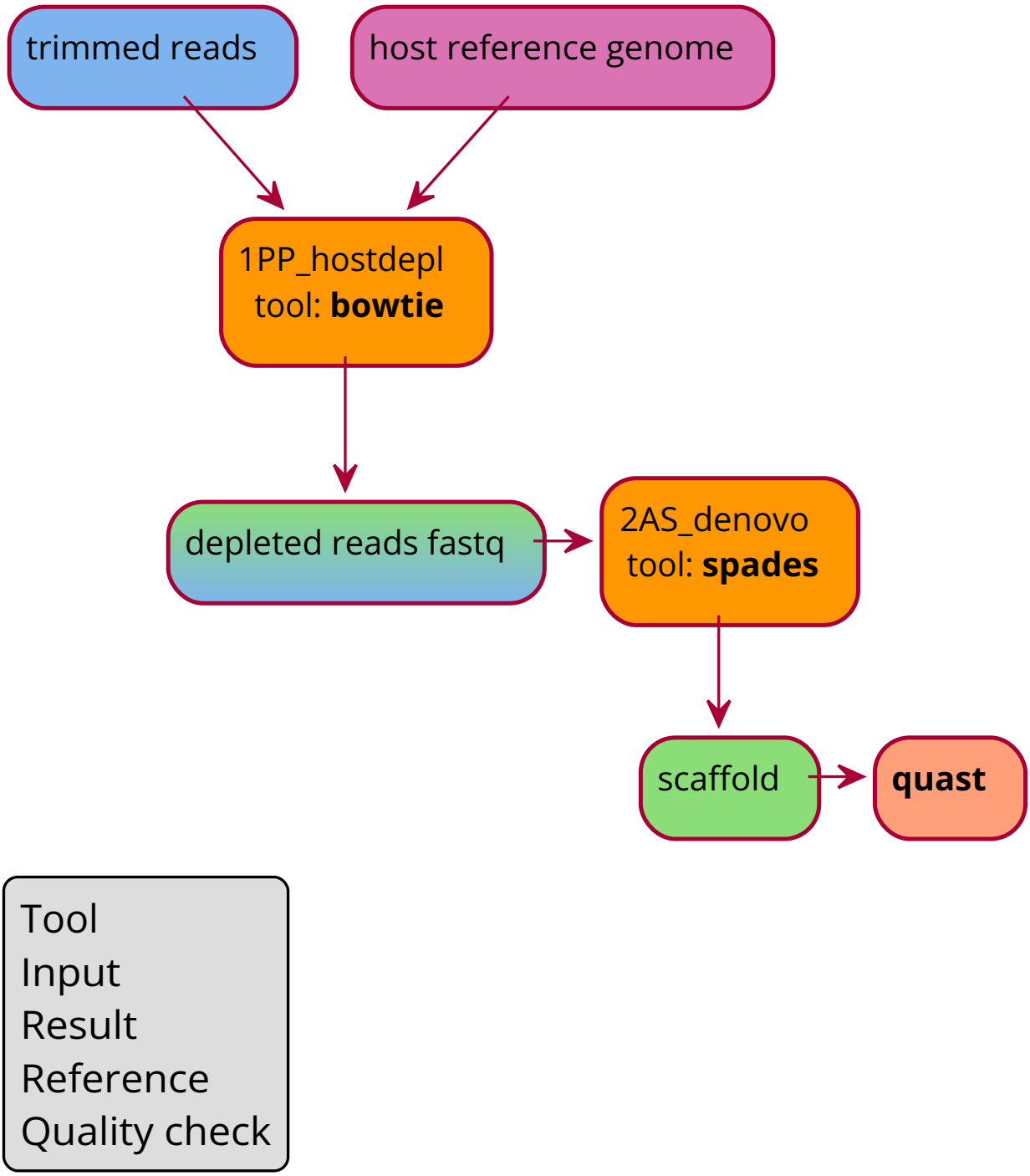
I files prodotti dalla pipeline "Covid Emergency" saranno gli stessi prodotti dalle analisi che la compongono, organizzati con la stessa gerarchia. Come riferimento si rimanda quindi alle relative sezioni delle singole analisi:

- Output del mapping eseguito con Snippy
- Output dell'assegnazione del lineage eseguito con Pangolin

Pipeline Deplezione & *de novo*

Introduzione

La pipeline **Deplezione & *de novo*** esegue consecutivamente la deplezione delle *reads* dell'ospite (1PP_hostdepl) e il *de novo assembly* (2AS_denovo).



Lancia Pipeline Deplezione & *de novo*

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Deplezione & *de novo*** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

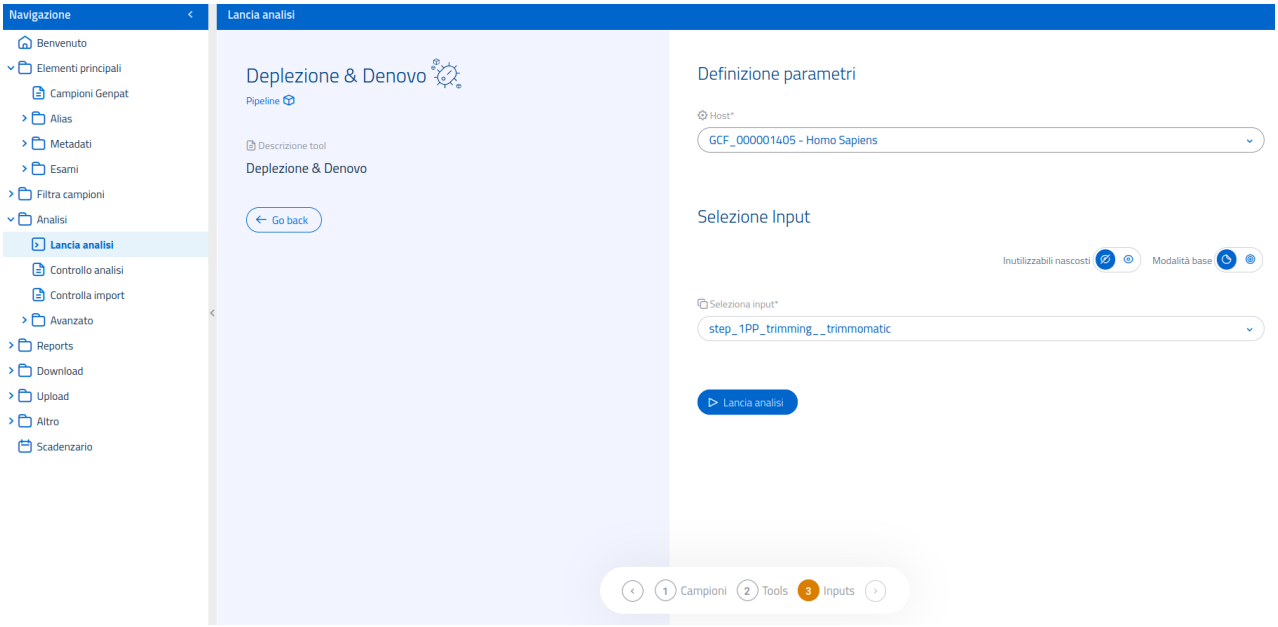
La pipeline esegue le 2 analisi seguenti:

1. Deplezione delle sequenze dell'ospite con **bowtie** (1PP_hostdepl__bowtie);
2. *de novo* assembly con **Spades** (2AS_denovo__spades) e controllo qualità con **Quast**.

Gli input utilizzabili sono pertanto gli stessi disponibili per l'esecuzione della deplezione dell'ospite tramite bowtie, ovvero il genoma di riferimento dell'ospite (nella sezione di definizione parametri, in alto) e le reads in formato fastq:

- step_1PP_hostdepl
- step_1PP_downsampling
- step_1PP_trimming

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

I files prodotti dalla pipeline "Deplezione & *de novo*" saranno gli stessi prodotti dalle analisi che la compongono, organizzati con la stessa gerarchia. Come riferimento si rimanda quindi alle relative sezioni delle singole analisi:

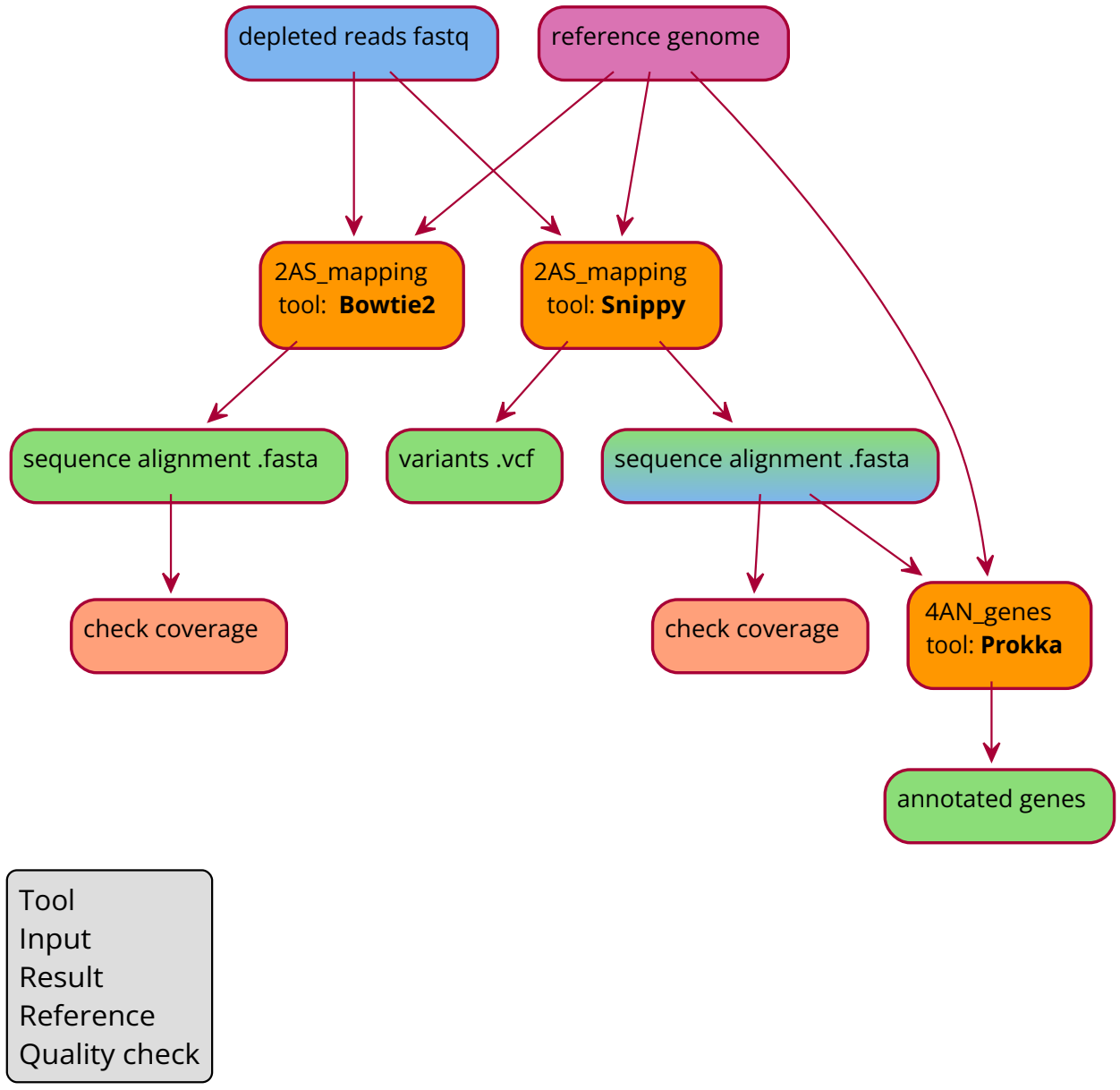
- Output della deplezione delle reads dell'ospite eseguita con bowtie
- Output del *de novo* assembly eseguito con spades e del *quality check* di Quast

Pipeline Draft del Genoma

Introduzione

La pipeline **Draft del Genoma** è parte della pipeline **Vdraft**, disponibile anche separatamente. Essa esegue il mapping delle sequenze (2AS_mapping) sia con Bowtie2 che con Snippy, seguito dall'annotazione del genoma (4AN_genes), tramite Prokka.

L'annotazione del genoma verrà eseguita solo sui risultati di Snippy e solo se come reference è stato usato un file in formato genebank (.gb o .gbk); a sua volta, Snippy potrà essere eseguito solo se il campione scelto possiede l'accertamento 4AN_genes associato (Prokka).



Lancia Pipeline Draft del Genoma

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Draft del Genoma** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

La pipeline **Draft del Genoma** comprende le analisi seguenti:

1. Mapping eseguito con **Bowtie2** (2AS_mapping__bowtie);
2. Mapping eseguito con **Snippy** (2AS_mapping__snippy);
3. Annotazione del genoma eseguito con **Prokka** (4AN_genes__prokka).

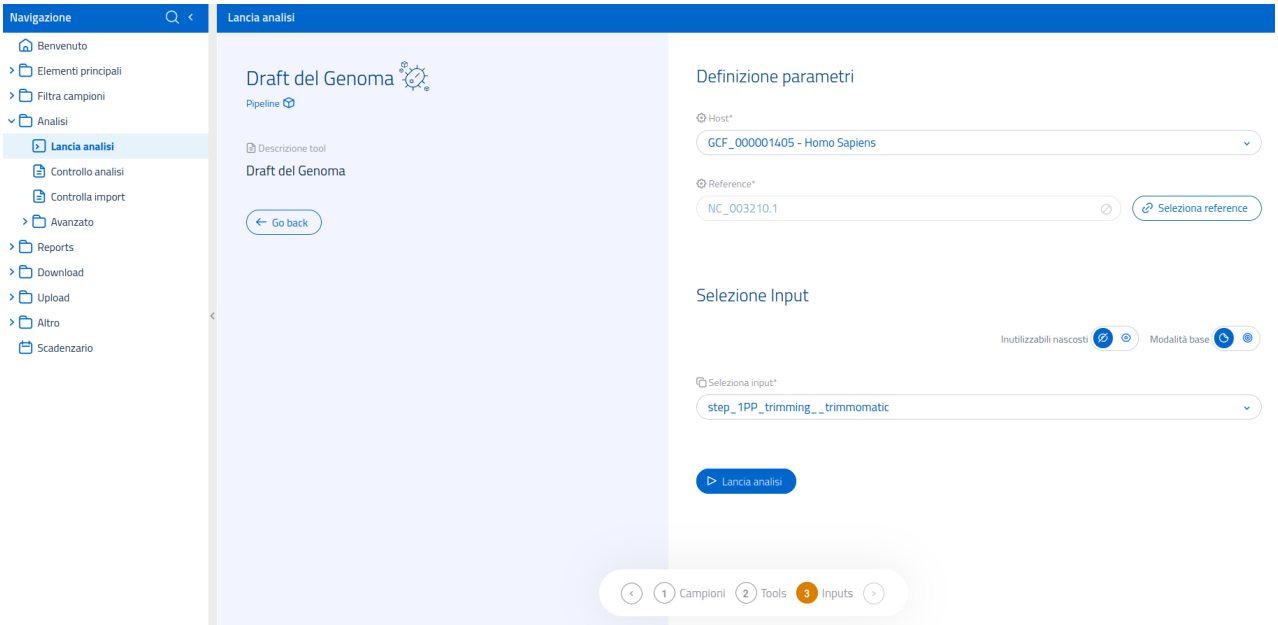
Entambe le esecuzioni del mapping comprendono inoltre il calcolo del *coverage*.

Gli input utilizzabili sono i risultati delle analisi di preprocessamento:

- *reads* deplete (step_1PP_hostdepl)
- *downsampled reads* (step_1PP_downsampling)
- *reads* trimmate (step_1PP_trimming)
- *reads* filtrate (step_1PP_filtering)

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

Sarà anche necessario specificare il genoma dell'organismo ospite e un genoma reference tra quelli disponibili, che verrà usato sia per i mapping che per l'annotazione funzionale.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

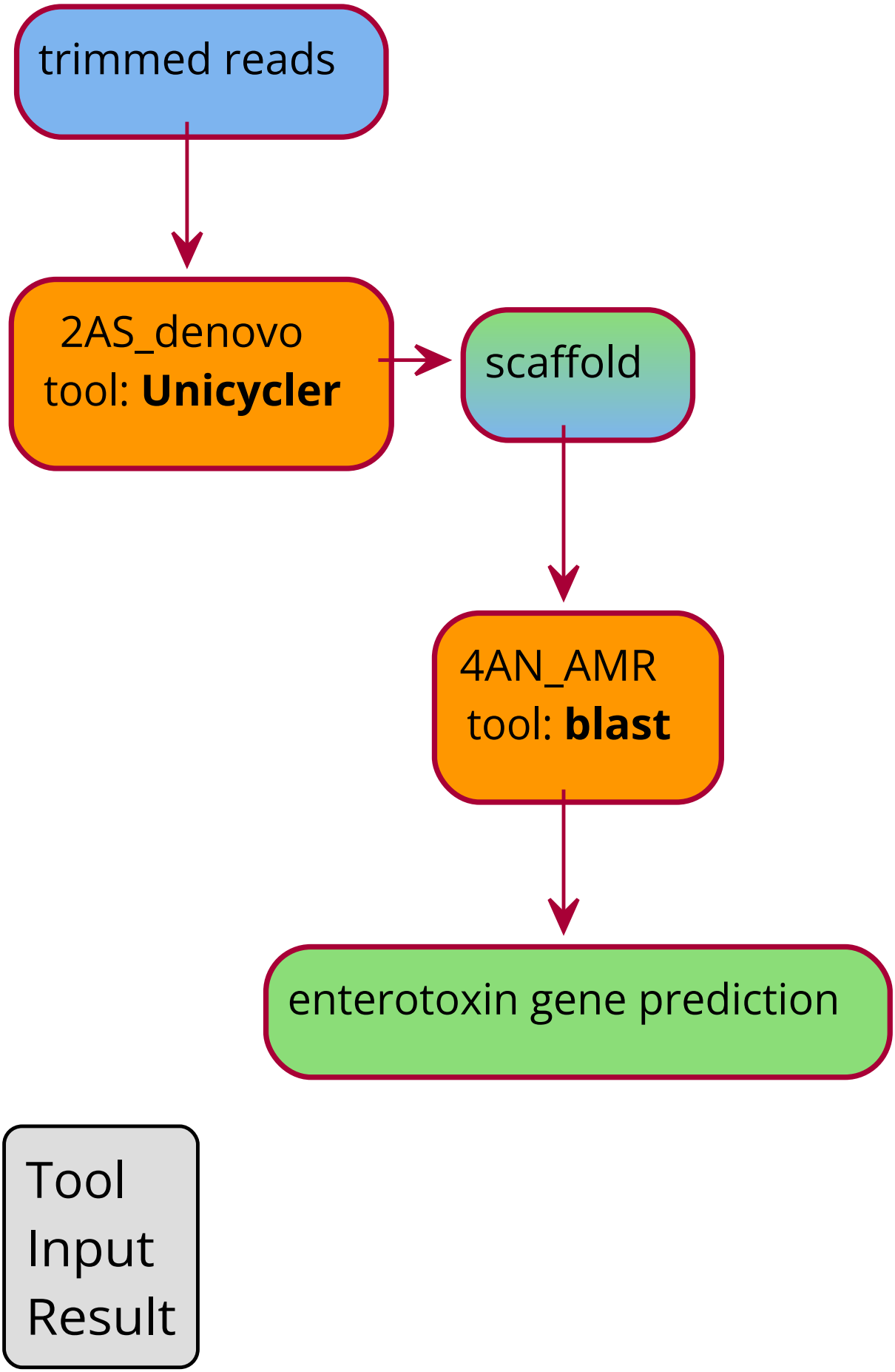
I files prodotti dalla pipeline "Draft del Genoma" saranno gli stessi prodotti dalle analisi che la compongono, organizzati con la stessa gerarchia. Come riferimento si rimanda quindi alle relative sezioni delle singole analisi:

- Output del mapping
- Output dell'annotazione funzionale del genoma

Pipeline Enterotoxin *S. aureus* Finder

Introduzione

La pipeline **Enterotoxin *Staphylococcus aureus* Finder** permette di identificare la presenza del gene codificante per enterotossina in *S. aureus*, eseguendo prima il *de novo* assembly (2AS_denovo), seguito dall'analisi di tipizzazione 4AN_AMR.


trimmed reads
2AS_denovo
tool: Unicycler
scaffold
4AN_AMR
tool: blast
enterotoxin gene prediction
Tool
Input
Result

Lancia Pipeline Enterotoxin *S. aureus* Finder

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Enterotoxin *S. aureus* Finder** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

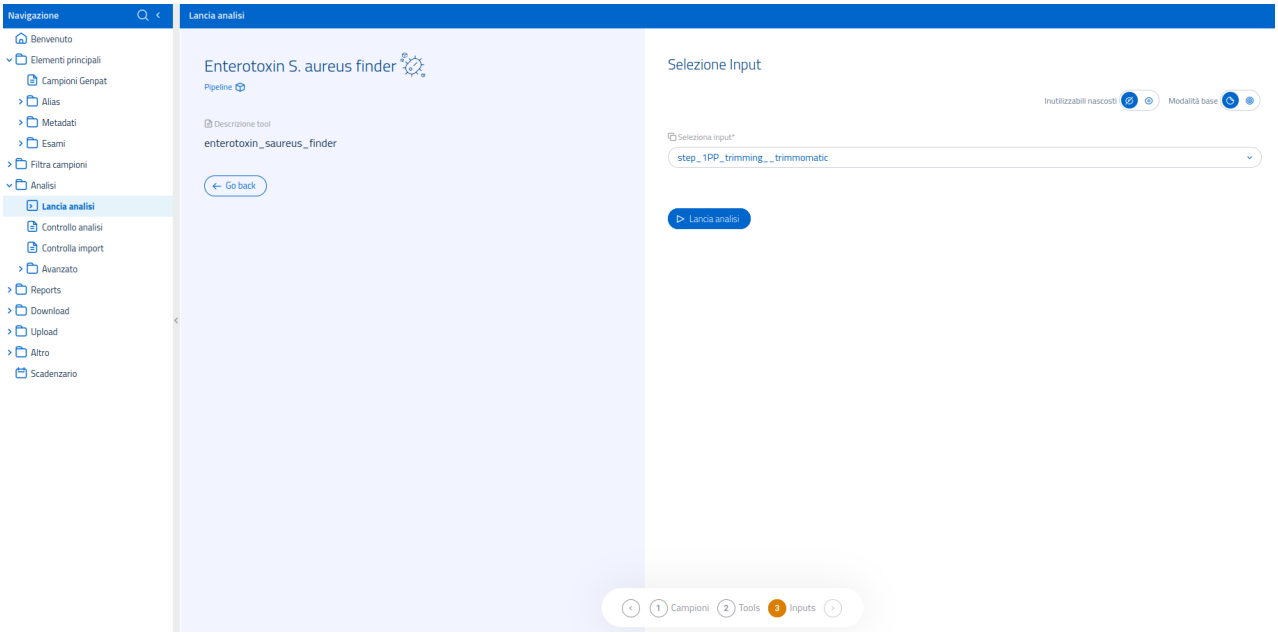
La pipeline **Enterotoxin *S. aureus* Finder** esegue:

1. *de novo* assembly con **Unicycler** (2AS_denovo__unicycler)
2. predizione della presenza del gene codificante l'enterotossina con **blast** (4AN_AMR__blast)

Gli input utilizzabili sono le *reads* trimmate derivanti da un'analisi di preprocessamento:

- *reads* deplete (step_1PP_hostdepl)
- *downsampled reads* (step_1PP_downsampling)
- *reads* trimmate (step_1PP_trimming)
- *reads* filtrate (step_1PP_filtering)

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

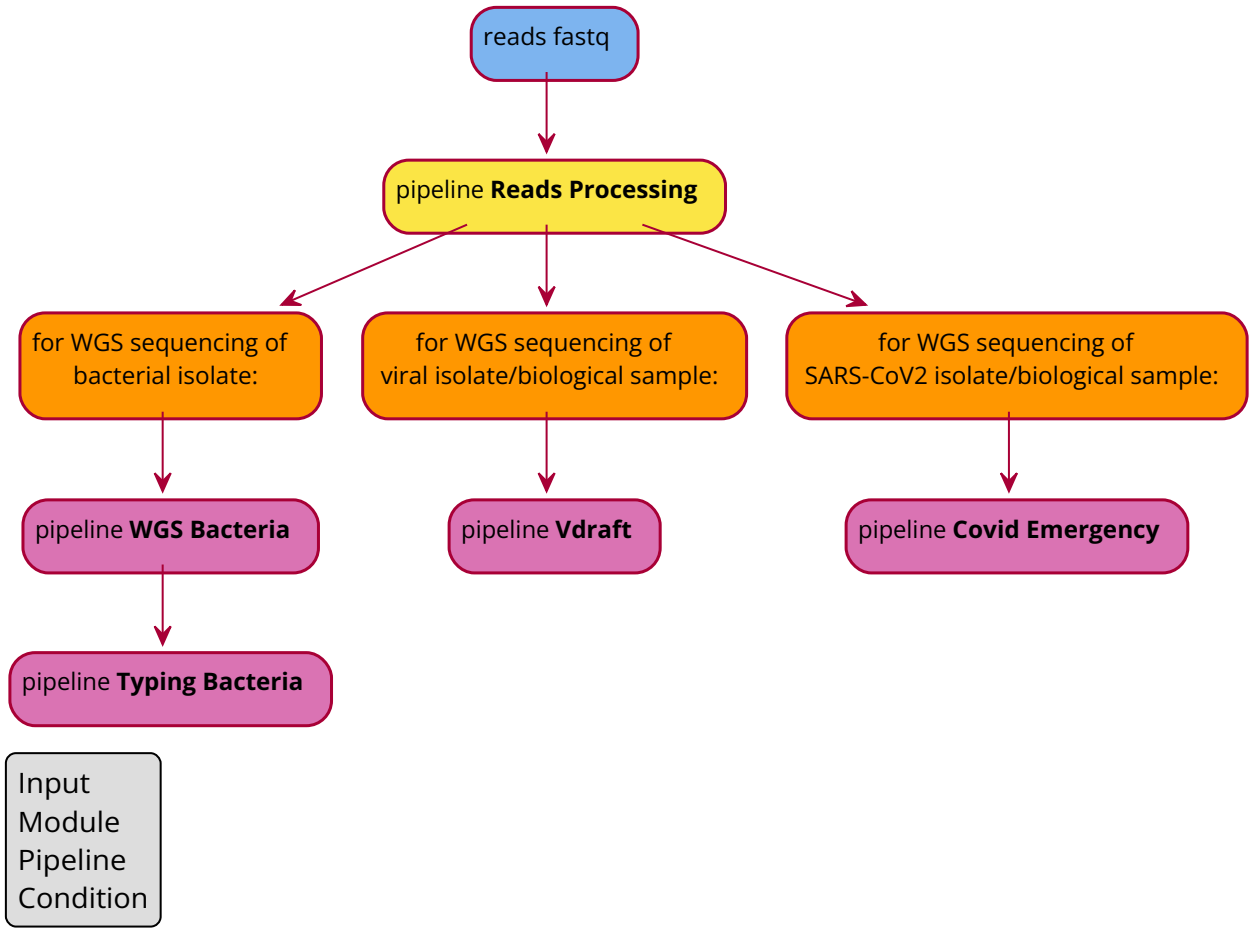
I files prodotti dalla pipeline "Enterotoxin *S. aureus* Finder" saranno gli stessi prodotti dalle analisi che la compongono, organizzati con la stessa gerarchia. Come riferimento si rimanda quindi alla relativa sezione delle singole analisi:

- Output del *de novo* assembly eseguito con Unicycler
- Output della predizione della presenza del gene d'interesse con blast

Pipeline NgsManager

Introduzione

La pipeline NgsManager permette la gestione delle *reads* e l'esecuzione di pipelines specifiche in base alla tipologia di campione.



Le pipelines eseguite da NgsManager sono le seguenti:

- Pipeline Processamento Raw Reads - I risultati di 3TX_class__kraken nella pipeline Reads Processing determinano l'identificazione di specie, in base alla quale vengono eseguite le analisi:
- Per batteri:

  Pipeline WGS su Batteri (esecuzione del *de novo* assembly con Shovill: step_2AS_denovo__shovill)

  Pipeline Typing su batteri
- Per virus:

  Pipeline Vdraft (esecuzione del *de novo* assembly con Abricate: step_2AS_denovo__abricate + modulo Draft del Genoma)
- Per SARS-CoV2:

  Pipeline Covid Emergency

(Per informazioni sulle singole analisi si consultino le pagine corrispondenti, raggiungibili cliccando sugli elementi sopra elencati).

Lancia Pipeline NgsManager

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **NgsManager** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

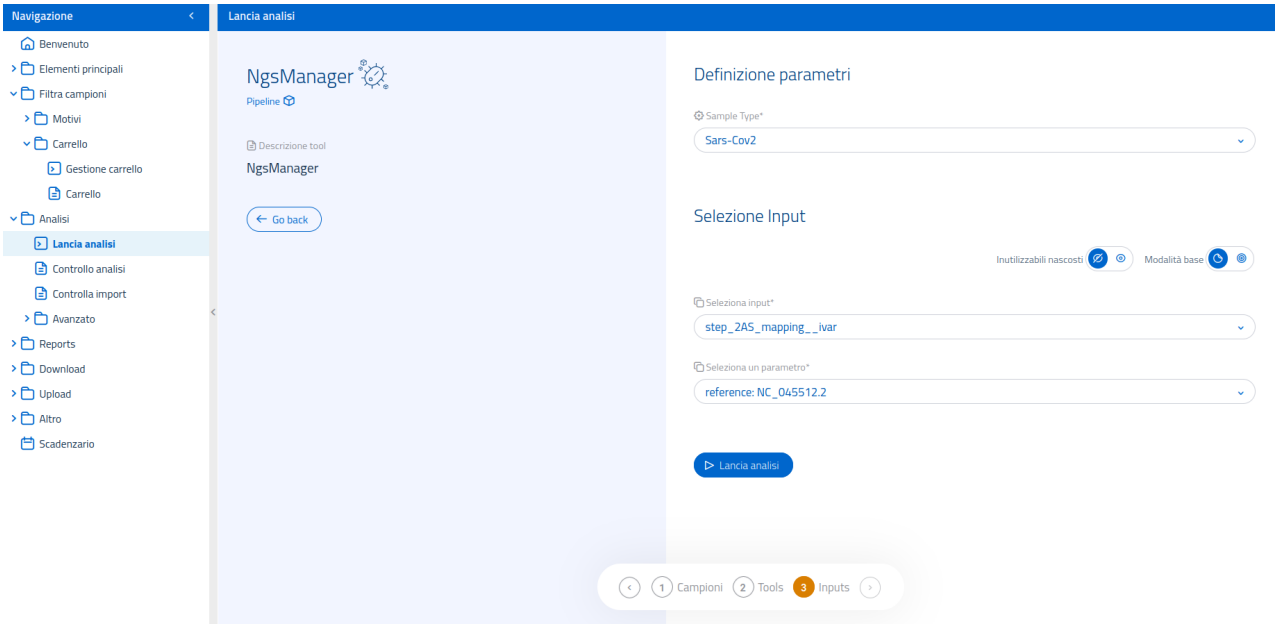
La selezione delle pipelines di NgsManager, in caso di lancio manuale, avviene scegliendo la tipologia di campione tra le 3 disponibili nella sezione di Definizione Parametri.

Per quanto riguarda l'input, NgsManager accetta:

- le *raw reads* (input step_0SQ_rawreads)
- il fasta del mapping (step_2AS_mapping)

Nel caso venga usato come input 2AS_mapping, sarà anche necessario un genoma reference, selezionato automaticamente dal sistema in base a quello usato per il mapping. La pipeline Processamento Raw Reads viene eseguita indistintamente per tutti i tipi di campione, ma solo se l'input selezionato è costituito dalle *raw reads*.

L'interfaccia per la selezione dell'input mette anche a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

I files prodotti da "NgsManager" dipendono dalle analisi che compongono le pipelines eseguite; essi sono organizzati con la stessa gerarchia usata per le analisi lanciate singolarmente. Come riferimento si rimanda quindi alle relative sezioni delle pipelines incluse in NgsManager, dove sarà possibile reperire tutte le informazioni riguardo le singole analisi e i link alle rispettive tabelle dei risultati.

- Risultati della pipeline "Processamento Raw Reads"
- Risultati della pipeline per batteri:
  - Risultati della pipeline "Typing su batteri"
- Risultati della pipeline per SARS-CoV2 "Covid Emergency"
- Risultati della pipeline per virus "Vdraft" (Draft del Genoma)
- Risultati della pipeline "WGS su Batteri" (2AS_denovo_shovill)

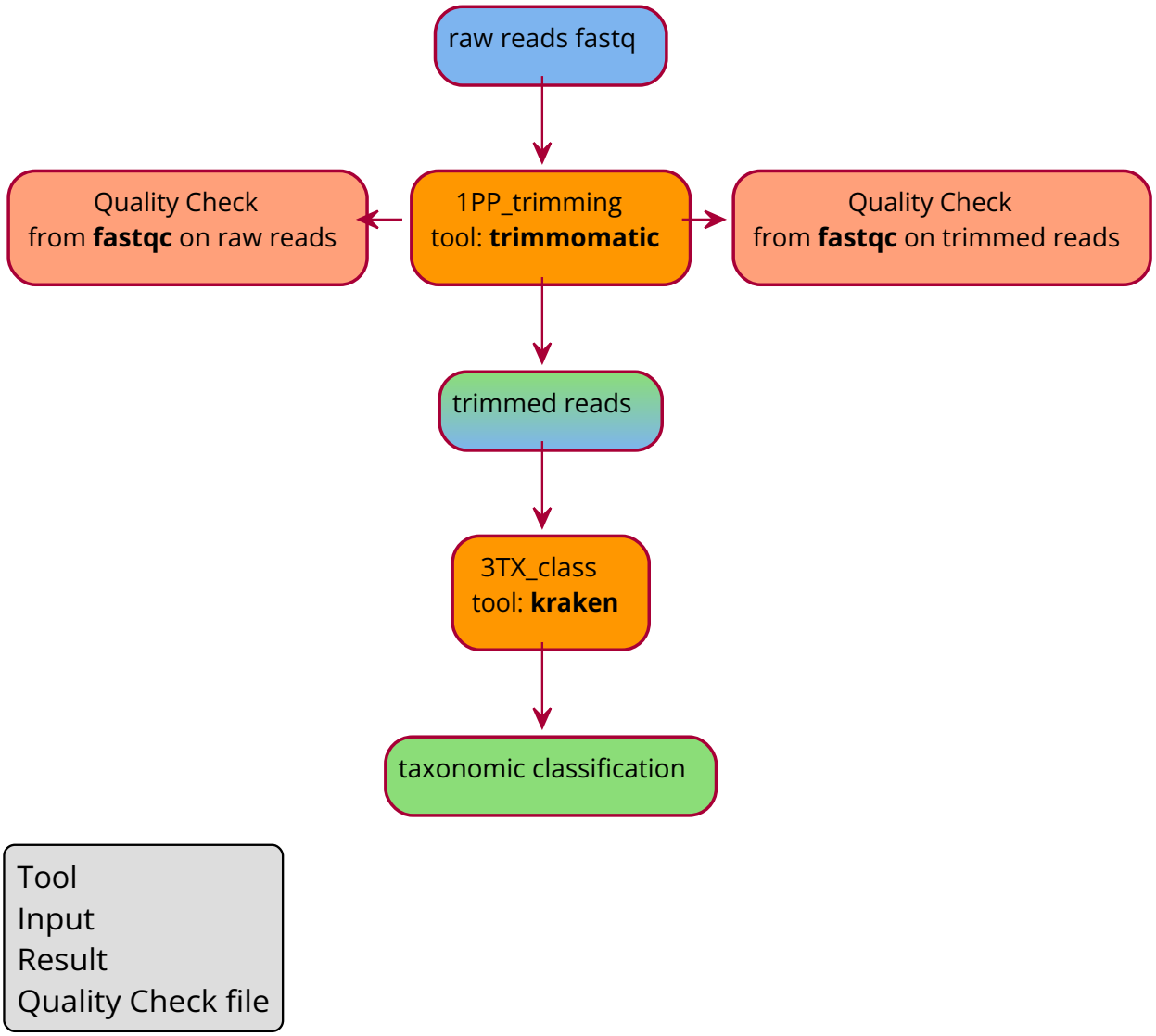
Pipeline Processamento Raw Reads

Introduzione

La pipeline **Processamento Raw Reads** esegue il preprocessamento delle *reads* (**1PP**), seguito dalla classificazione tassonomica.

Le *reads* prodotte dal sequenziatore vengono prima sottoposte ad un controllo della qualità con fastQC ed a trimming con trimmomatic (1PP_trimming); la successiva analisi tassonomica e l'**assegnazione della specie calcolata** vengono invece effettuate con Kraken (3TX_class).

> Si ricorda che nell'analisi 1PP_trimming il controllo qualità viene eseguito con **fastQC** sia sulle *raw reads* che sulle *reads* trimmate ottenute.



La pipeline esegue le 2 analisi seguenti:

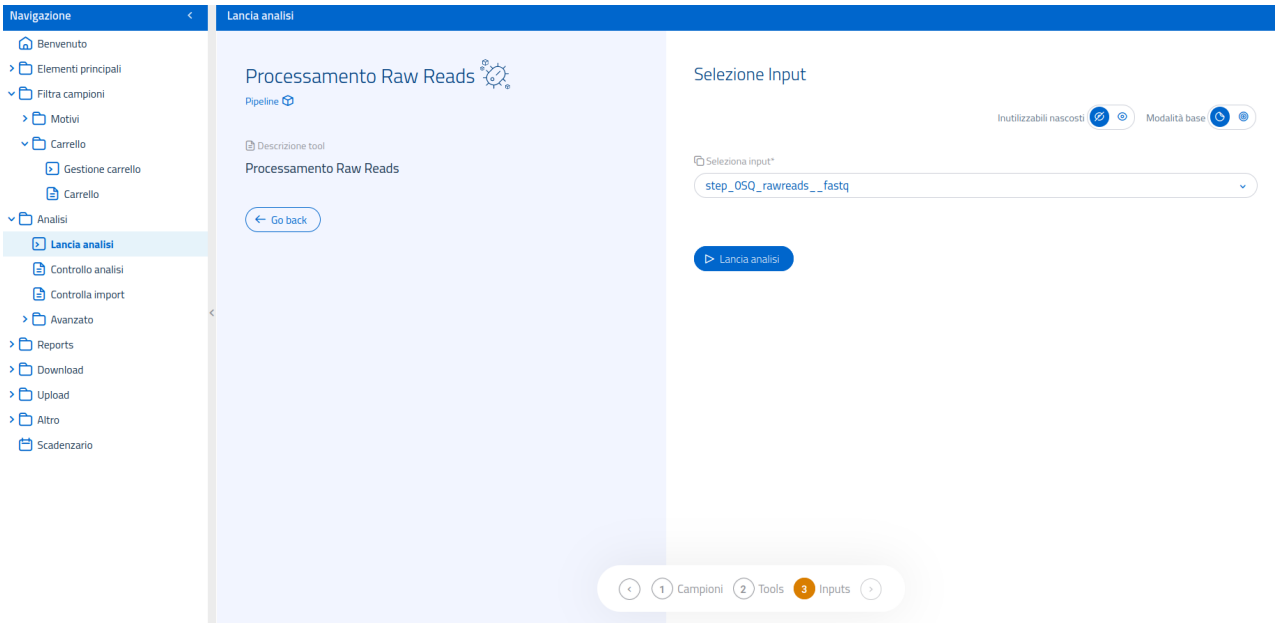
1. Trimming delle *raw reads* per eliminare le basi con bassa qualità con **trimmomatic** (1PP_trimming__trimmomatic) e **fastQC**;
2. Classificazione tassonomica dei batteri o virus individuati e controllo contaminazione con **Kraken** (3TX_class__kraken) e **Quast**. La classificazione con **Kraken** è resposabile dell'assegnazione della **specie calcolata**, sulla base della quale vengono automaticamente selezionate le pipelines a valle, se la Pipeline Processamento Raw Reads viene eseguita da NGSmanager.

Lancia Pipeline Processamento Raw Reads

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Processamento Raw Reads** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

Gli input utilizzabili sono le *raw reads* in formato fastq (interne o importate).

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

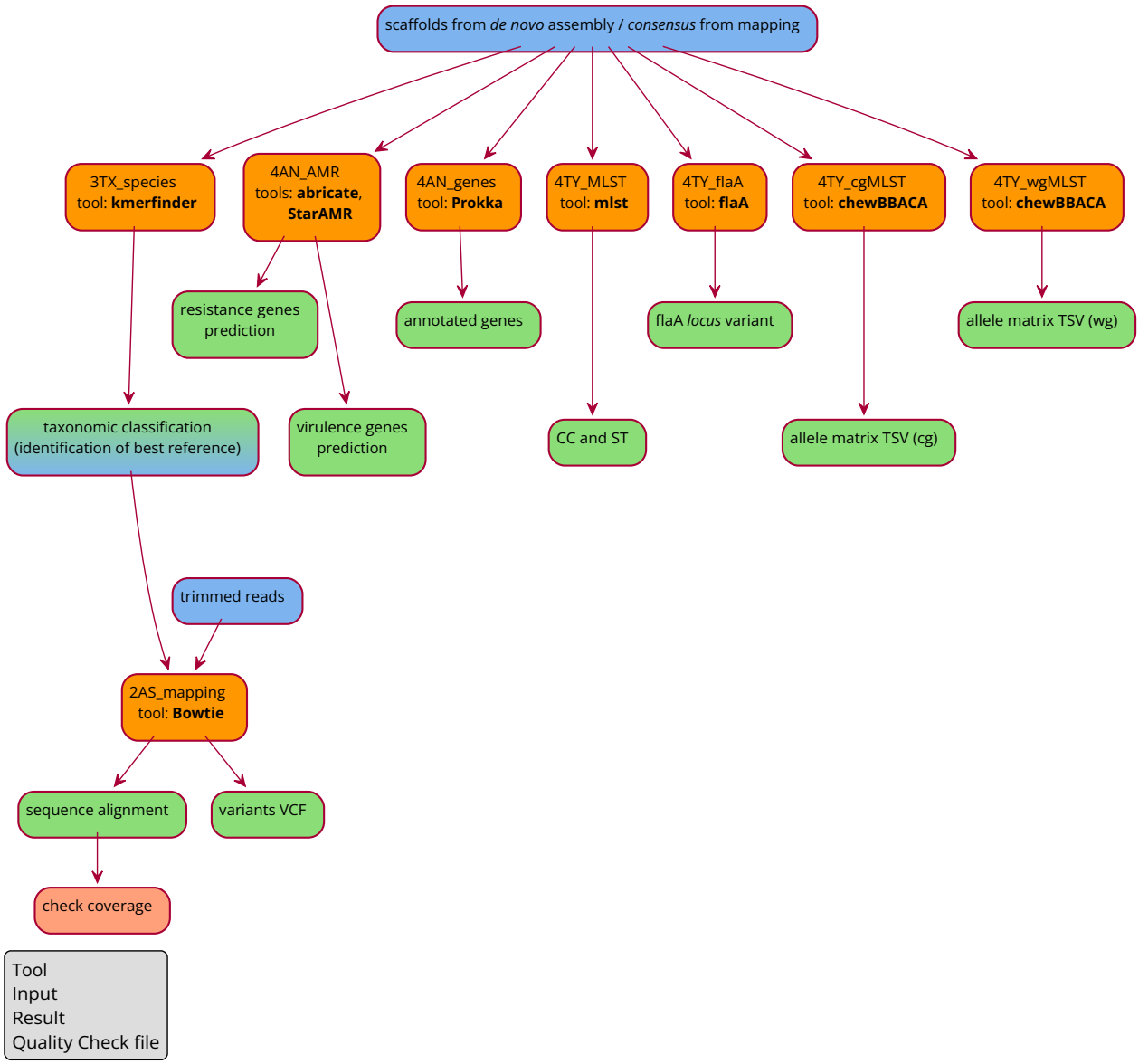
I files prodotti dalla pipeline "Processamento Raw Reads" saranno gli stessi prodotti dalle analisi che la compongono, organizzati con la stessa gerarchia. Come riferimento si rimanda quindi alle relative sezioni delle singole analisi:

- Output del trimming delle *reads* eseguito con trimmomatic e *quality check* di fastQC
- Output della classificazione tassonomica eseguita con Kraken e del *quality check* di Quast

Pipeline Typing su Batteri

Introduzione

La pipeline **Typing su Batteri** esegue in un unico lancio tutte le analisi di tipizzazione a cui vengono solitamente sottoposti i campioni di batteri.



La pipeline **Typing su Batteri** esegue le analisi seguenti:

1. 3TX_species: calcolo della specie con **kmerfinder** per assegnare il "best reference" da usare per il mapping a valle.
   Quando eseguito da NGSmanager gli inputs sono gli scaffolds del *de novo* da Shovill (pipeline WGS su Batteri).
2. 4AN_AMR: identificazione di geni correlati ad antibiotico-resistenza e virulenza.
3. 4TY_flaA: determinazione della variante del locus flaA (solo per *Campylobacter*).
4. 4TY_MLST: Multi-Locus Sequence Typing *in silico* con software MLST.
5. 4TY_cgMLST: core-genome MLST con software ChewBBACA.
6. 4TY_wgMLST: whole-genome MLST con software ChewBBACA.
7. 4AN_genes: annotazione funzionale del genoma con **Prokka**.

Sulla base dei risultati di kmerfinder (3TX_species) viene inoltre eseguita:

- 2AS_mapping con Bowtie: il reference usato è quello identificato come "best reference" da kmerfinder, tra i genomi presenti nel database statico disponibile (kmerfinder).

Il software eseguito per le analisi **4TY_cgMLST** e **4TY_wgMLST**, quando lanciate tramite pipeline NGSmanager, è sempre **ChewBBACA**.

4TY_MLST_mlst viene eseguito se è disponibile lo schema per il microrganismo (la lista di schemi è consultabile nella guida ufficiale di mlst su GitHub); **4TY_flaA** viene invece eseguito solo su *Campylobacter*.

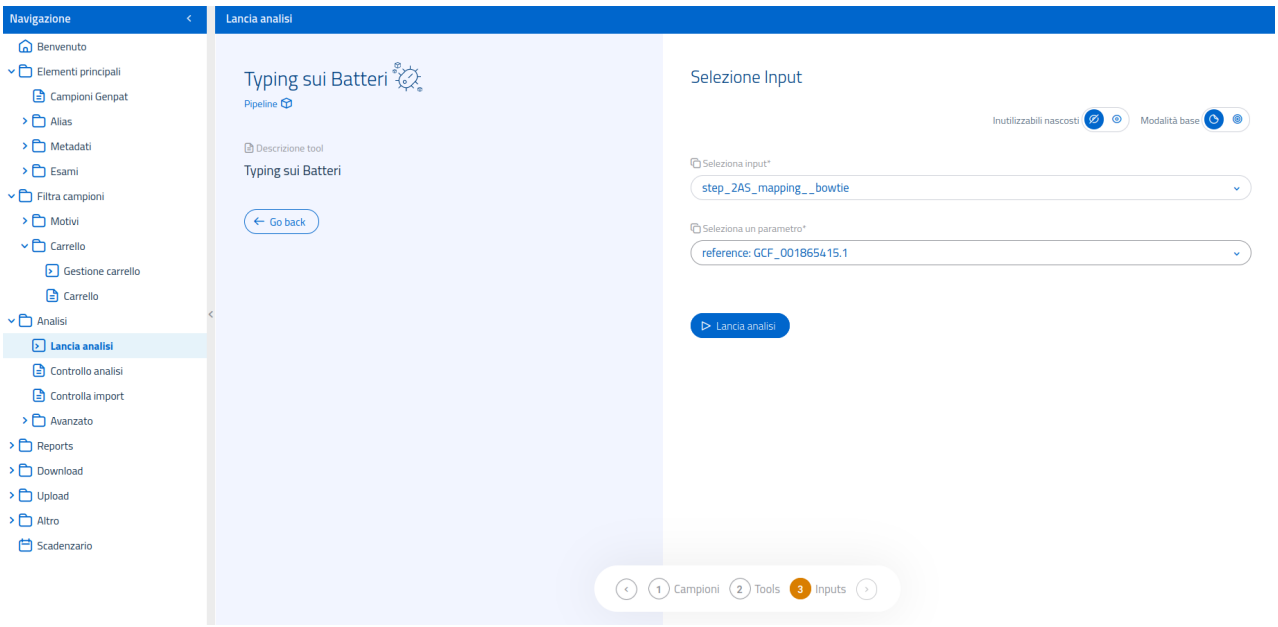
Lancia Pipeline Typing su Batteri

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Typing su Batteri** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

Gli input utilizzabili sono le sequenze *consensus* (formato .fasta) prodotte dal mapping (2AS_mapping) o gli *scaffolds* del *de novo* assembly (2AS_denovo).

Nel caso vengano usati come input le sequenze *consensus* del mapping, sarà anche necessario specificare un genoma reference tra quelli disponibili.

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

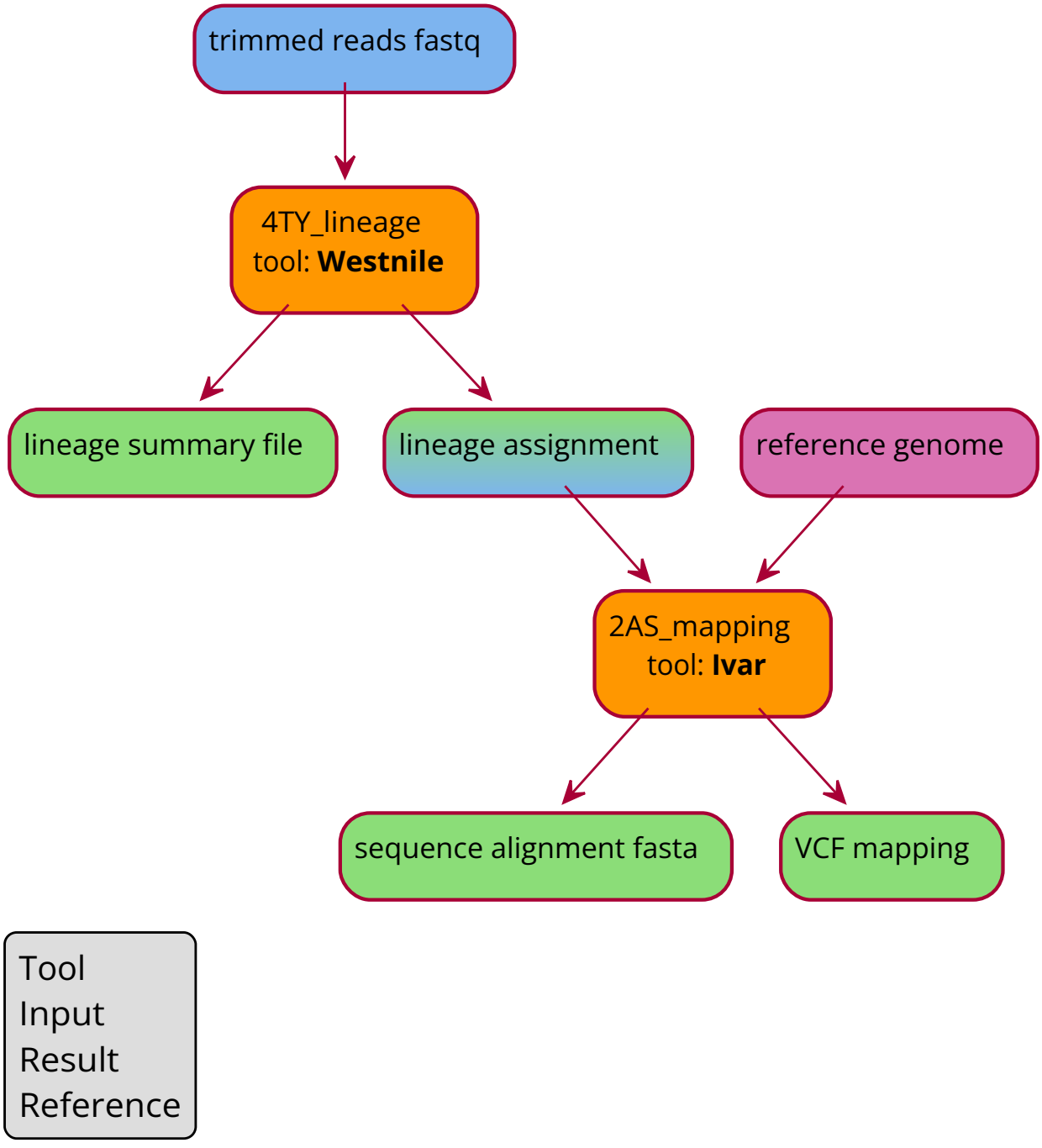
I files prodotti dalla pipeline "Typing su Batteri" saranno gli stessi prodotti dalle analisi che la compongono, organizzati con la stessa gerarchia. Come riferimento si rimanda quindi alle relative sezioni delle singole analisi:

- Output dell'assegnazione della specie
- Output della predizione di antibiotico-resistenza e virulenza
- Output per la determinazione della variante del *locus* flaA
- Output del Multi-*Locus* Sequence Typing (MLST)
- Output del core-genome MLST (cgMLST)
- Output del whole-genome MLST (wgMLST)
- Output dell'annotazione del genoma
- Output del mapping sui risultati di kmerfinder

Pipeline WNV (West Nile Virus)

Introduzione

La pipeline "WNV - Lineage Calculation and Mapping" calcola il *lineage* per West Nile Virus e ne effettua il mapping contro il *reference* corrispondente.



Lancia Pipeline WNV

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **WNV** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

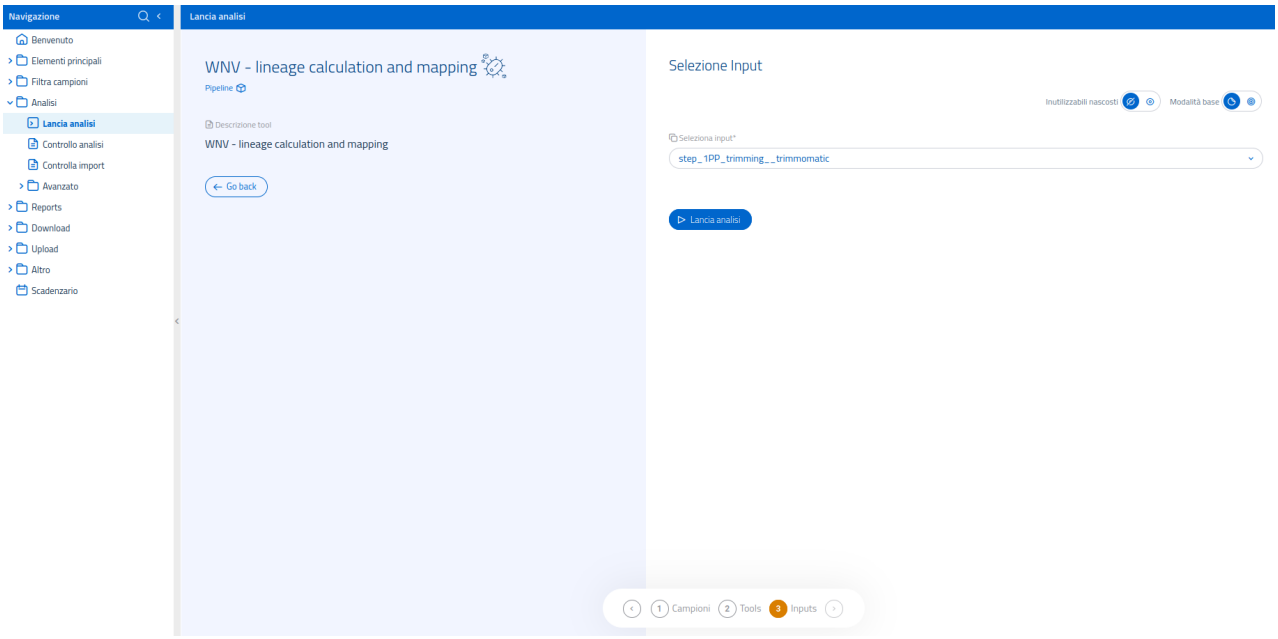
La pipeline **WNV** esegue le analisi:

1. Calcolo del lineage di West Nile Virus con **Westnile** (4TY_lineage__westnile)
2. Mapping con **Ivar** (2AS_mapping__ivar) di tutti i segmenti di genoma da virus segmentati forniti, usando più *references*

Gli input utilizzabili sono i risultati delle analisi di preprocessamento:

- *reads* deplete (step_1PP_hostdepl)
- *downsampled reads* (step_1PP_downsampling)
- *reads* trimmate (step_1PP_trimming)
- *reads* filtrate (step_1PP_filtering)

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Il reference usato per la fase di mapping sarà il genoma base associato al lineage assegnato da **Westnile**, secondo le corrispondenze riportate nella tabella sottostante:

lineage assegnato	reference
lin. 1	FJ483548.1
lin. 2	HQ537483.1
lin. 3	AY765264.1
lin. 4	AY277251.1
lin. 5	GQ851605.1
lin. 6	GU047875.1
lin. 7	KY703855.1
lin. 8	KY703856.1
lin. 9	KJ831223.1

Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

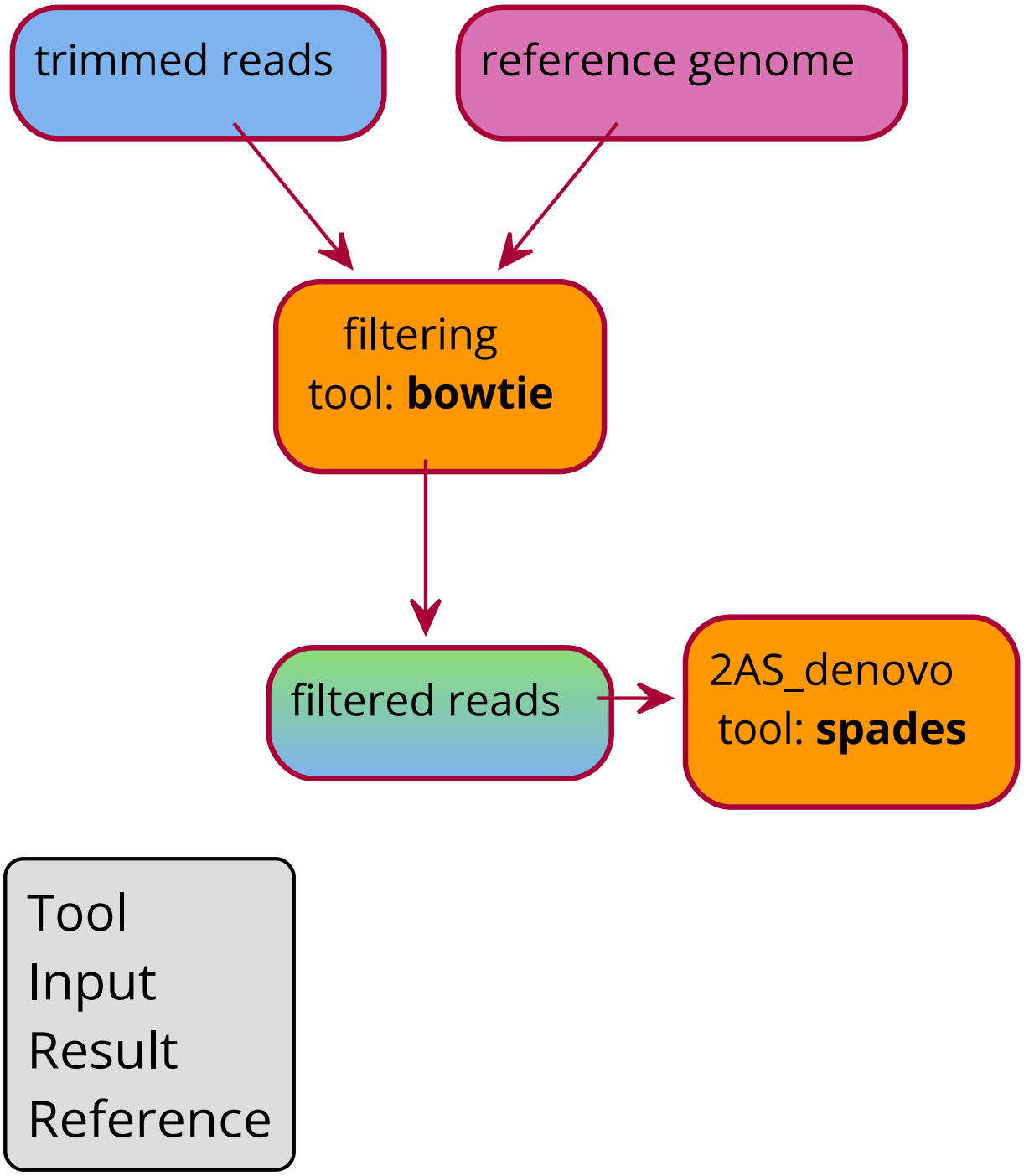
I files prodotti dalla pipeline "WNV" saranno gli stessi prodotti dalle analisi che la compongono, organizzati con la stessa gerarchia. Come riferimento si rimanda quindi alle relative sezioni delle singole analisi.

- Output dell'assegnazione del lineage
- Output del mapping

Filtering & *de novo*

Introduzione

La pipeline **Filtering & *de novo*** elimina le reads che non mappano sul reference selezionato (1PP_filtering) ed esegue il *de novo* assembly (2AS_denovo).



Lancia Pipeline Filtering & *de novo*

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Filtering & *de novo*** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

La pipeline **Filtering & *de novo*** esegue:

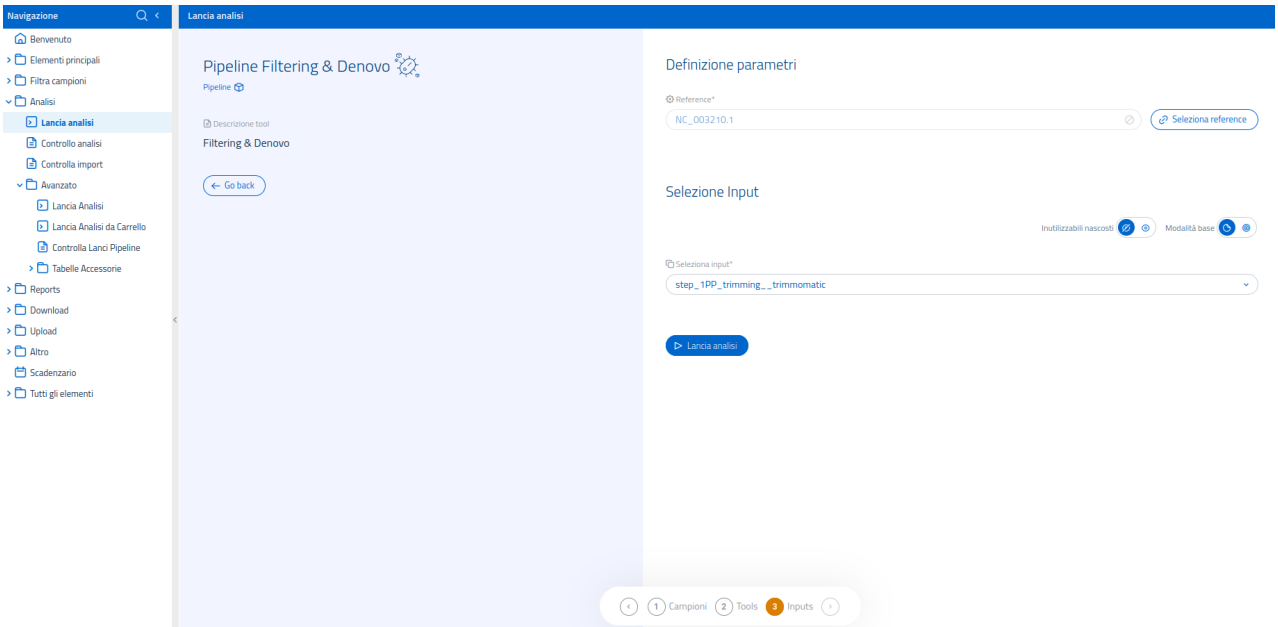
1. rimozione delle *reads* che non mappano con il *reference* tramite **Bowtie** (1PP_filtering__bowtie)
2. *de novo* assembly con **SPAdes** (2AS_denovo__spades)

Gli input utilizzabili sono costituiti dalle *reads* trimmate derivanti da un'analisi di preprocessamento:

- *reads* deplete (step_1PP_hostdepl)
- *downsampled reads* (step_1PP_downsampling)
- *reads* trimmate (step_1PP_trimming)
- *reads* filtrate (step_1PP_filtering)
- *reads* generate da fasta (step_1PP_generated)

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

Nella sezione dedicata alla selezione dei parametri sarà anche necessario specificare un genoma reference, indispensabile per effettuare la prima fase della pipeline: *filtering* (1PP_filtering).



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

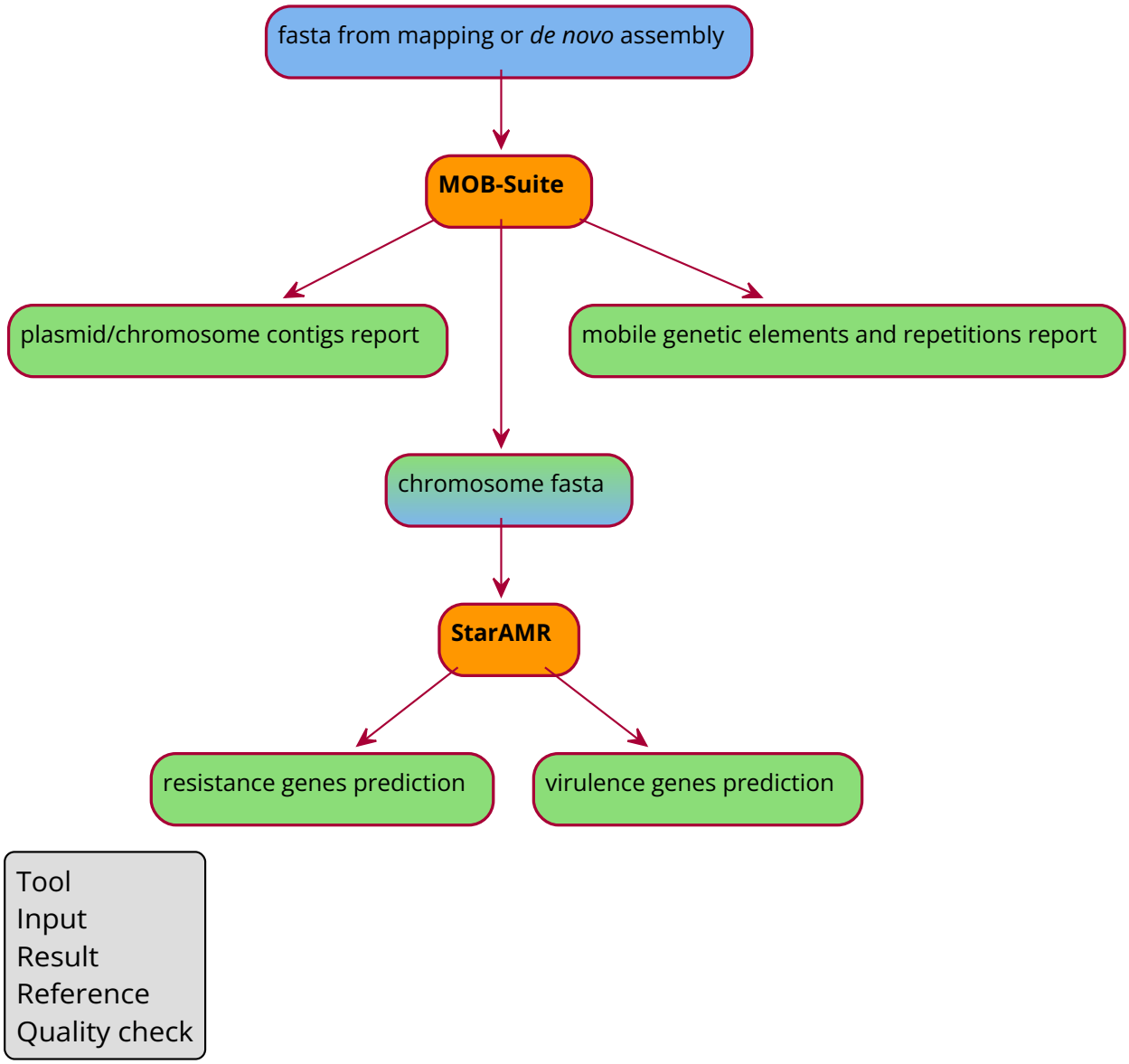
I files prodotti dalla pipeline "Filtering & *de novo*" saranno gli stessi prodotti dalle analisi che la compongono, organizzati con la stessa gerarchia. Come riferimento si rimanda quindi alla relativa sezione delle singole analisi:

- Output del filtraggio delle *reads* con Bowtie
- Output del *de novo* assembly con SPAdes

Pipeline Plasmids AMR

Introduzione

La pipeline **Plasmids AMR** permette la tipizzazione dei plasmidi, eseguendo 4TY_plasmid con MOB-Suite e successivamente 4AN_AMR con StarAMR, sul fasta creato da 4TY_plasmid.



Lancia Pipeline Plasmids AMR

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Plasmids AMR** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

La pipeline **Plasmids AMR** comprende le analisi seguenti:

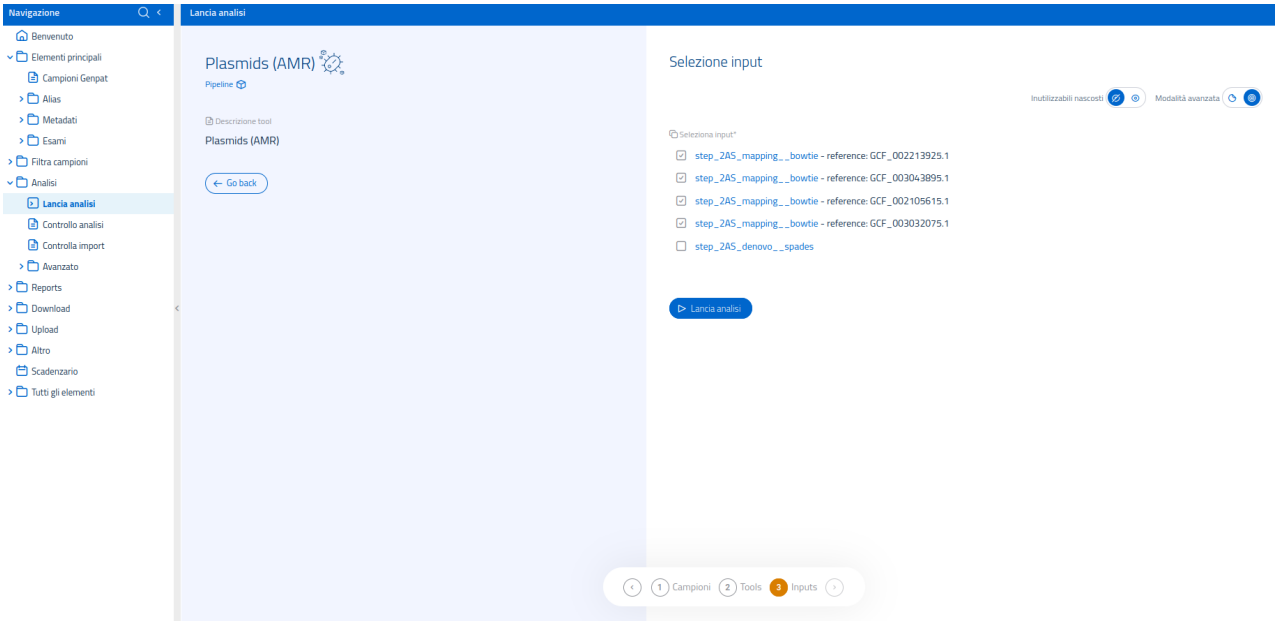
1. Identificazione dei plasmidi eseguito con **MOB-Suite** (4TY_plasmid__mobsuite);
2. Predizione dei geni correlati ad antibiotico-resistenza eseguito con **Snippy** (4AN_AMR);

Gli input utilizzabili sono gli stessi della prima analisi eseguita (4TY_plasmid), ovvero gli output delle analisi di ricostruzione della sequenza:

- fasta da *de novo* assembly (step_2AS_denovo)
- fasta da mapping (step_2AS_mapping)
- fasta da assembly ibrido (step_2AS_hybrid)
- fasta da *de novo* assembly per campioni di metagenomica (step_2MG_denovo)
- *reads* importate (step_2AS_import)

Nel caso di input proveniente da mapping, sarà anche necessario fornire il reference usato per il mapping (campo precompilato).

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

I files prodotti dalla pipeline "Plasmids AMR" saranno gli stessi prodotti dalle analisi che la compongono, organizzati con la stessa gerarchia. Come riferimento si rimanda quindi alle relative sezioni delle singole analisi:

- Output della caratterizzazione dei plasmidi
- Output della predizione di antibiotico-resistenza

Pipeline nf-core/ampliseq

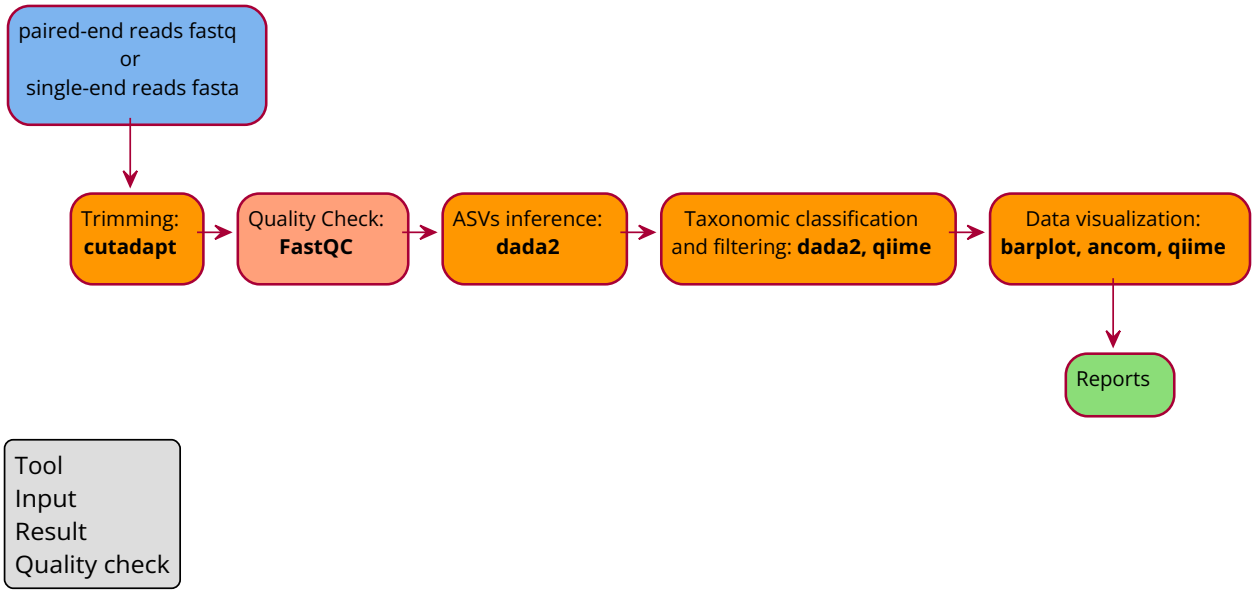
Introduzione

La pipeline **nf-core/ampliseq** disponibile in GenPat proviene della community nf-core, dedita alla costruzione e condivisione di pipeline bioinformatiche gestite con Nextflow. Essa effettua il sequenziamento ed il *denoising* degli ampliconi, inoltre può disporre di diversi databases per l'assegnazione tassonomica (tra cui databases per sequenze 16S, ITS, CO1 and 18S). La pipeline supporta reads paired-end Illumina, single-end Illumina, PacBio o IonTorrent.

La versione della pipeline nf-core/ampliseq disponibile in GenPat è la **2.7.1**.

Pagina ufficiale della versione 2.7.1 di nf-core/ampliseq sulla community nf-core: https://nf-co.re/ampliseq/2.7.1

Pagina GitHub di nf-core/ampliseq: https://github.com/nf-core/ampliseq



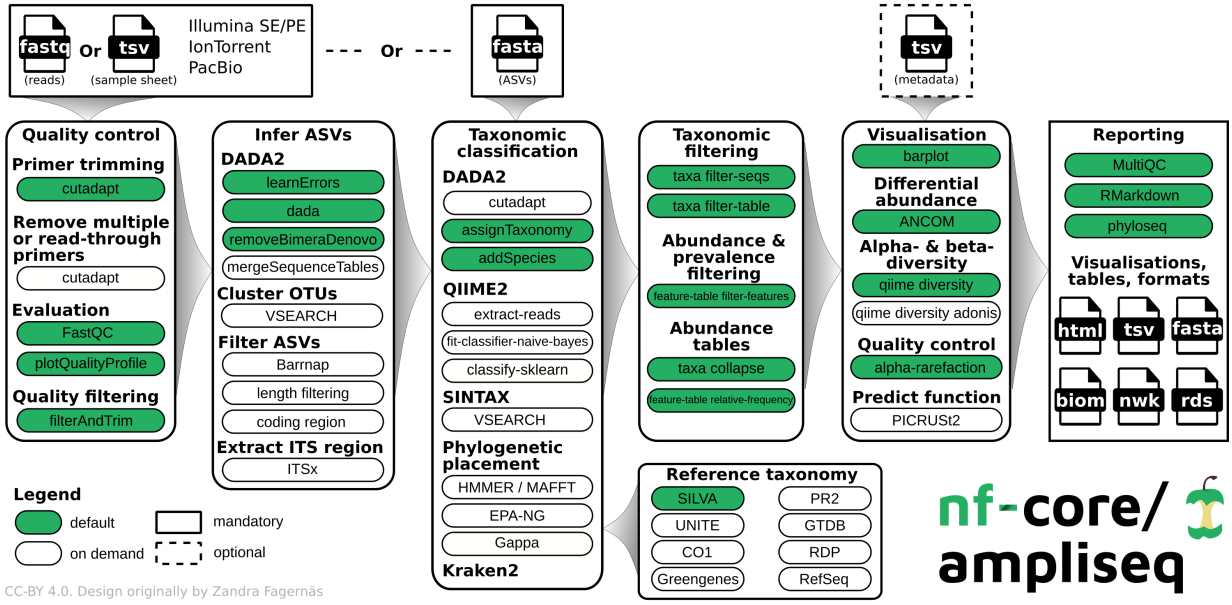


Immagine di proprietà della community nf-core: https://nf-co.re/ampliseq/2.7.1

Le analisi di default di nf-core/ampliseq (in verde nell'immagine sovrastante) sono quelle che vengono eseguite in GenPat. Alcune delle analisi "on demand" possono essere aggiunte sfruttando i parametri addizionali nel lancia analisi, inoltre alcune analisi di default possono essere saltate, sempre sfruttando i parametri addizionali. Si invita a consultare le sezioni "parameters" e "parameters: skipping specific steps" della guida ufficiale di nf-core/ampliseq per sfruttare al meglio i parametri addizionali ed il campo di testo per i parametri *shell-like*.

Lancia Pipeline nf-core/ampliseq

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **nf-core/ampliseq** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

Le fasi della pipeline sono riassunte nello schema all'inizio della pagina.

Gli input utilizzabili sono le reads paired-end Illumina in formato fastq o le reads single-end da apparecchiature Illumina, IonTorrent o PacBio in formato fasta:

- step_1PP_trimming
- step_1PP_hostdepl
- step_1PP_downsampling
- step_1PP_filtering

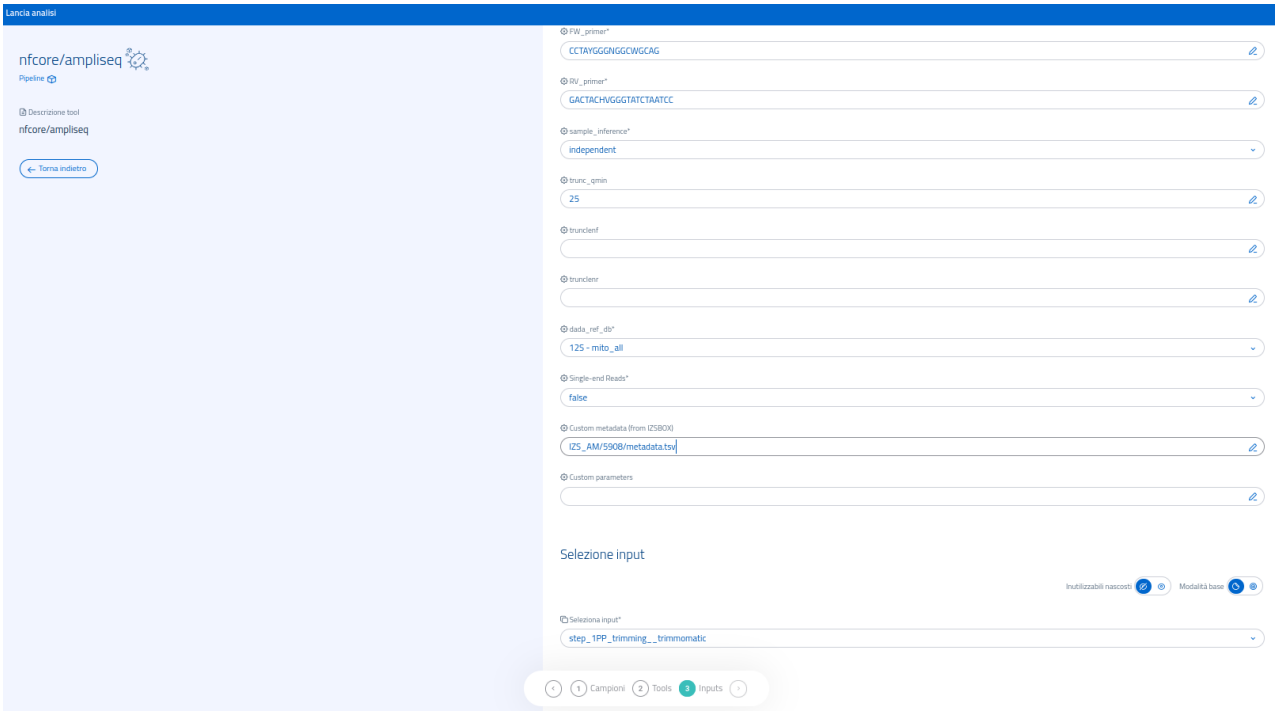
L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.

Parametri addizionali

Di seguito sono elencati i parametri addizionali disponibili nella sezione apposita del sistema di lancio:

- `FW_primer` e `RV_primer`: campi di testo in cui incollare la sequenza dei primers *forward* e *reverse*, rispettivamente. Necessario per il trimming dei primers.
- `sample_inference`: modalità di inferenza per il calcolo delle ASVs.
  - `independent`: modalità "risparmio", più leggera e veloce ma con minore sensibilità;
  - `pooled`: modalità "performance", con maggiore sensibilità ma più pesante da eseguire;
  - `pseudo`: "pseudo-pooled". Modalità bilanciata;
- `trunclenf`, `trunclenr` e `trunc_qmin`: la fase di denoising richiede reads della stessa lunghezza. I parametri per il troncamento sono campi di testo in cui specificare manualmente la posizione nucleotidica da usare come valore di *cutoff* per la read *forward* (`trunclenf`) e per la read *reverse* (`trunclenr`). Le sequenze più corte vengono troncate e le quelle più corte vengono scartate. `trunc_qmin` permette di specificare un valore minimo di qualità della chiamata del nucleotide: se `trunclenf` e `trunclenr` non vengono specificati, i valori di *cutoff* vengono determinati automaticamente, troncando quando il valore mediano del quality score scende sotto la soglia specificata.
- `dada_ref_db`: selezione del database usato da dada2 per l'assegnazione tassonomica (in GenPat sono attualmente disponibili il database di rRNA 12S "mito-all" e di rRNA 16S "silva").
- `Reads`: selezione della tipologia di reads (*paired-end*/ *single-end*).
- `Custom metadata (from IZSBOX)`: permette di indicare un file di metadati presente nel proprio IZSbox (incollare il percorso del file).
- `Custom parameters`: gli altri parametri di nf-core/ampliseq sono disponibili come modificatori da scrivere nel campo di testo come comandi da terminale. La guida ufficiale contiene tutte le informazioni necessarie.

> Si invita a consultare le descrizioni dei parametri direttamente nella guida ufficiale nf-core/ampliseq: https://nf-co.re/ampliseq/2.7.1/parameters.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

La scheda riassuntiva della pipeline completata con successo permette di accedere alla cartella di output, in modo da recuperare i files tabulari e le immagini dei grafici, ma mette a disposizione anche links diretti per la visualizzazione delle immagini e dei report.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. La cartella principale contiene le directories `data` e `results`. Nella cartella `data` è possibile trovare i files delle reads usate. All'interno della directory `results` sono invece presenti i risultati, organizzati in sotto-cartelle corrispondenti ai softwares lanciati dalla pipeline.

La tabella in basso presenta la lista delle principali cartelle ed il loro contenuto, insieme ad alcune informazioni utili.

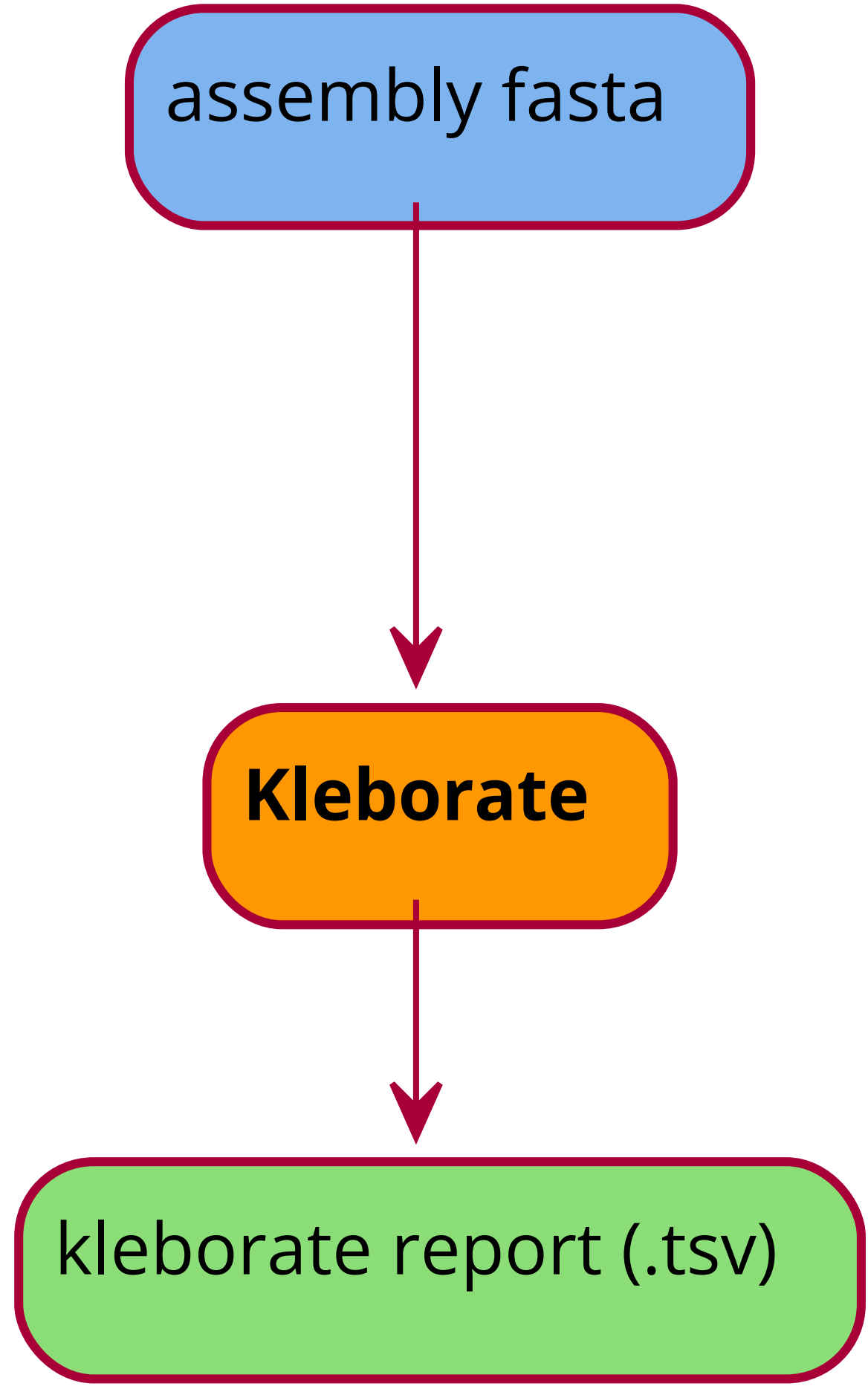
Cartella	Contenuto
`results/`	sotto-cartelle dei softwares e tabella riassuntiva (`overall_summary.tsv`)
`results/cutadapt`	logs di cutadapt (`DSXXXXXXXX.trimmed.cutadapt.log`) e tabella riassuntiva (`cutadapt_summary.tsv`)
`results/dada2`	tabelle tsv delle statistiche di dada2, degli ASVs trovati e fasta delle sequenze degli ASVs (`ASV_seqs.fasta`)
`results/dada2/QC`	grafici del quality check in formato pdf
`results/input`	file dei metadati (`metadata.tsv`)
`results/multiqc`	report di multiQC in formato html, cartelle dei dati di output e dei grafici di multiQC
`results/qiime`	cartelle dei dati prodotti da qiime: `abundance_tables` , `alpha-rarefaction` , `barplot` , `diversity` , `phylogenetic_tree` , `rel_abundance_tables` , `representative_sequences`
`results/summary_report`	grafici in formato immagine svg e report in formato html

Pipeline Klebsiella - Kleborate

Introduzione

La pipeline Klebsiella lancia il software Kleborate per lo screening dei genomi di *Klebsiella pneumoniae* e del complesso di specie di *Klebsiella pneumoniae* (*K. pneumoniae* species complex, KpSC). Kleborate analizza il genoma per ottenere dati riguardo: - Specie; - MLST di KpSC; - *loci* di virulenza associati a ICEKp: yersiniabactin (ybt), colibactin (clb), salmochelin (iro), hypermucoidy (rmp); - *loci* di virulenza associati a plasmidi: salmochelin (iro), aerobactin (iuc), hypermucoidy (rmp, rmpA2); - determinanti AMR: geni acquisiti, SNPs, geni troncati e β-lattamasi intrinseche, antigeni K (capsule) e O (LPS); - sierotipo, tramite alleli wzi e Kaptive.

Pagina GitHub di Kleborate: https://github.com/klebgenomics/Kleborate

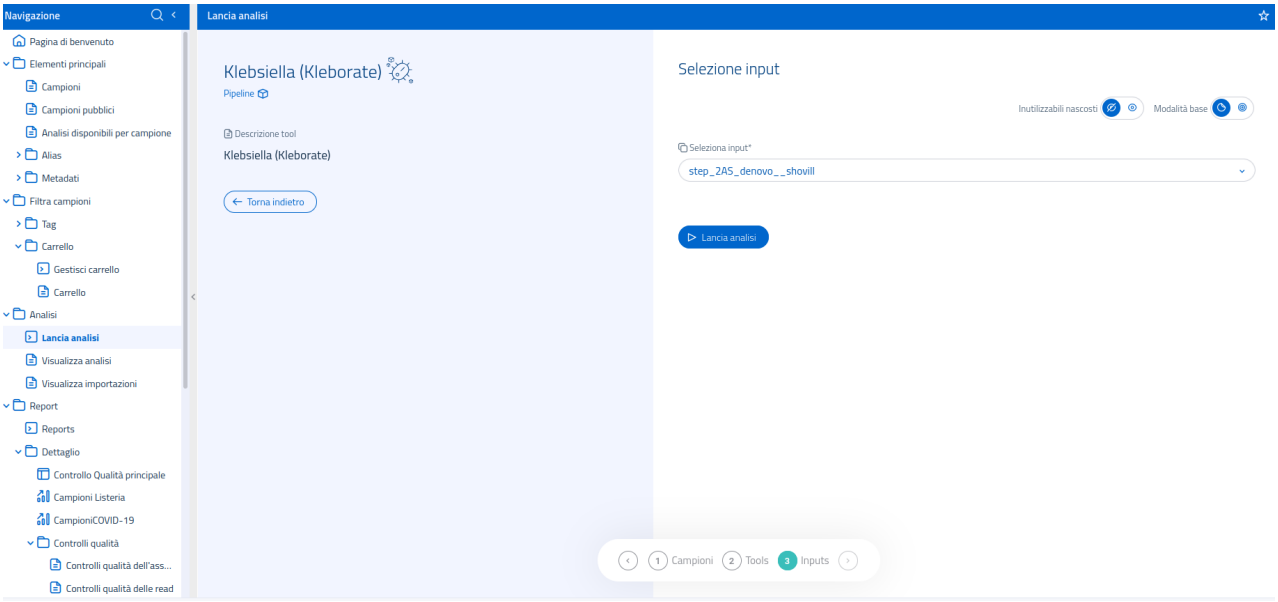

assembly fasta
Kleborate
kleborate report (.tsv)

Lancia Pipeline Kleborate

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Klebsiella (Kleborate)**, sarà possibile lanciare l'analisi usando come input l'assemblaggio *de novo*:

- step_2AS_denovo
- step_2MG_denovo
- step_2AS_hybrid

L'interfaccia per la selezione dell'input mette inoltre a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

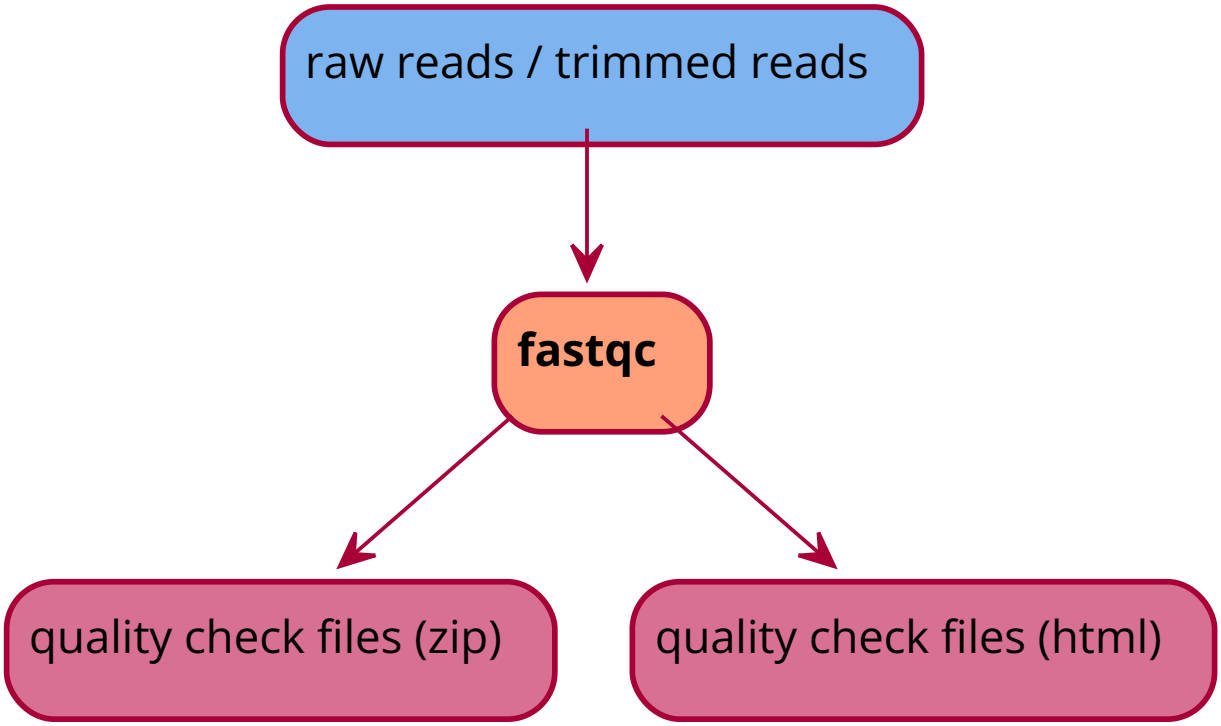
La pipeline "Kleborate" produce un unico output: una tabella tsv con tutti i dati di tipizzazione elaborati dal software.

File	Descrizione	Posizione
`*_kleborate.tsv`	tabella tsv contenente tutti i risultati della tipizzazione	cartella result

QC fastqc

Introduzione

QC fastqc fa parte della sottocategoria di analisi per il controllo qualità (*Quality Check*, QC); essa esegue il software **fastqc** sulle *raw reads* o sulle *reads* trimmate, fornendo le relative statistiche di qualità.



Input
Quality Check tool
Quality Check result

Lancia fastqc

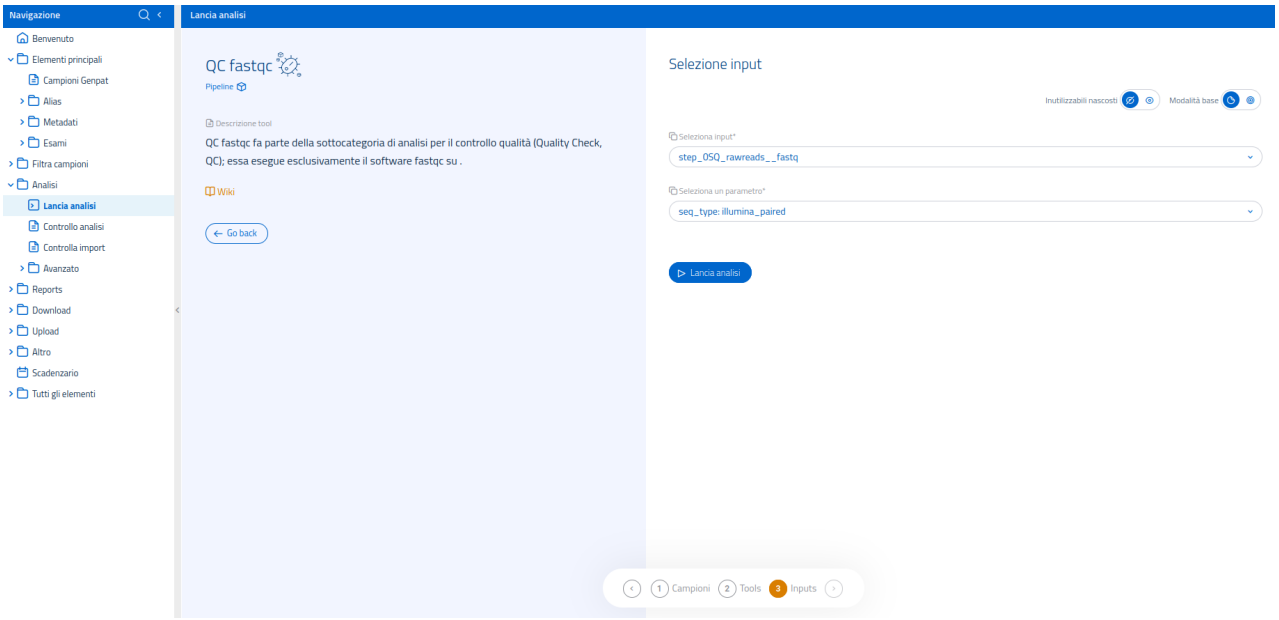
Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **QC fastqc** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

Gli input utilizzabili sono le *raw reads* in formato fastq (interne o esterne) o le *reads* trimmate (o provenienti da analisi di preprocessamento):

- *raw reads* (step_0SQ_rawreads)
- *reads* trimmate (step_1PP_trimming)
- *reads* filtrate (step_1PP_filtering)
- *reads* fastq generate da fasta (step_1PP_generated)
- *reads* deplete (step_1PP_hostdepl)
- *reads* sottoposte a *downsampling* (step_1PP_downsampling)

Nota: fastqc può essere eseguito su *reads paired* o *unpaired*. Nel primo caso ogni risultato di fastqc sarà prodotto per ognuna delle reads e sarà contrassegnato dai suffissi "R1" ed "R2"

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

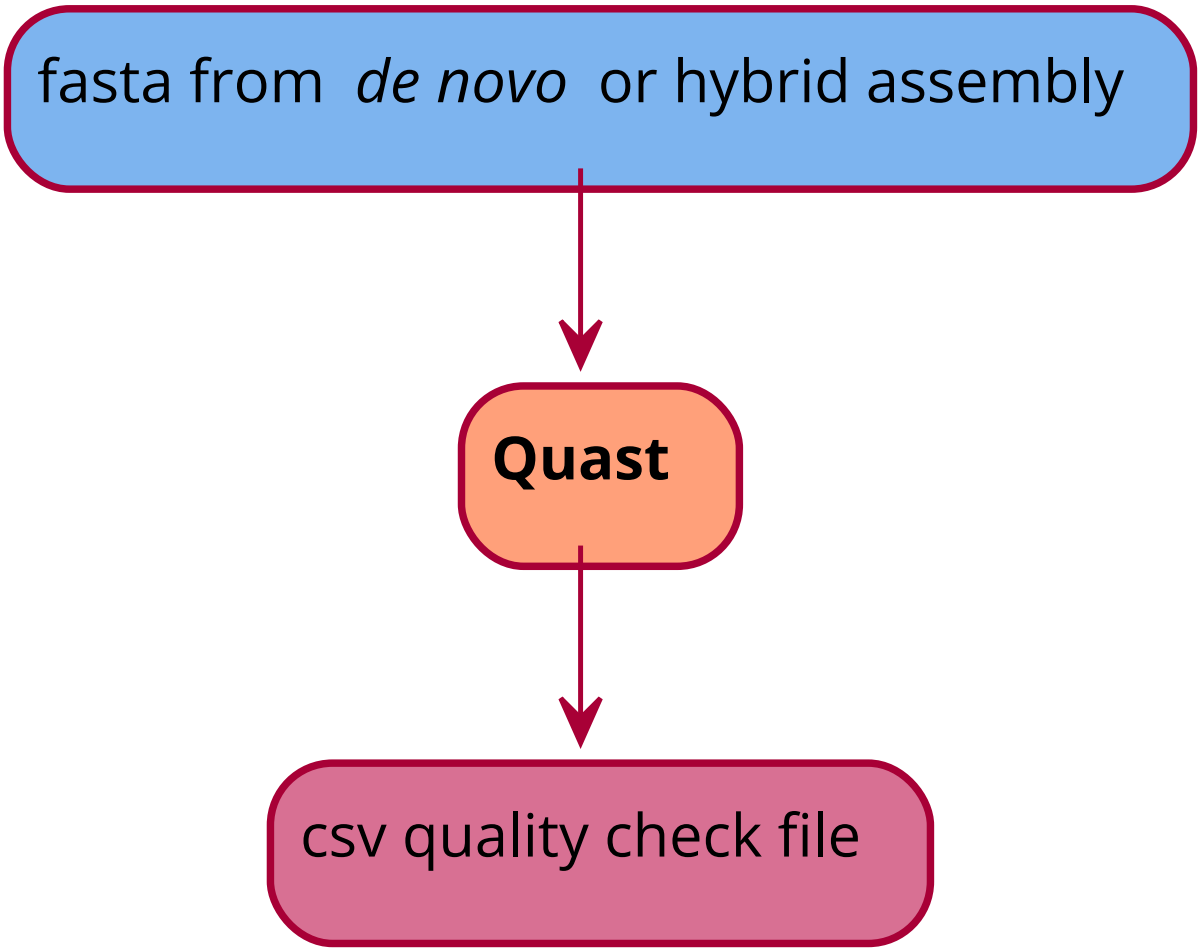
I files prodotti dalla pipeline "QC fastqc" saranno gli stessi prodotti dallo stesso tool all'interno delle analisi che lo usano come software per il controllo qualità. Per consultare i risultati di fastqc si invita quindi a fare riferimento alla tabella presente al seguente link agli outputs di 1PP_trimming:

- *Quality check* di fastQC

QC Quast

Introduzione

QC Quast fa parte della sottocategoria di analisi per il controllo qualità (*Quality Check*, QC); essa esegue il software **Quast** sui file dell'assembly (o assembly ibrido).



Input
Quality Check tool
Quality Check result

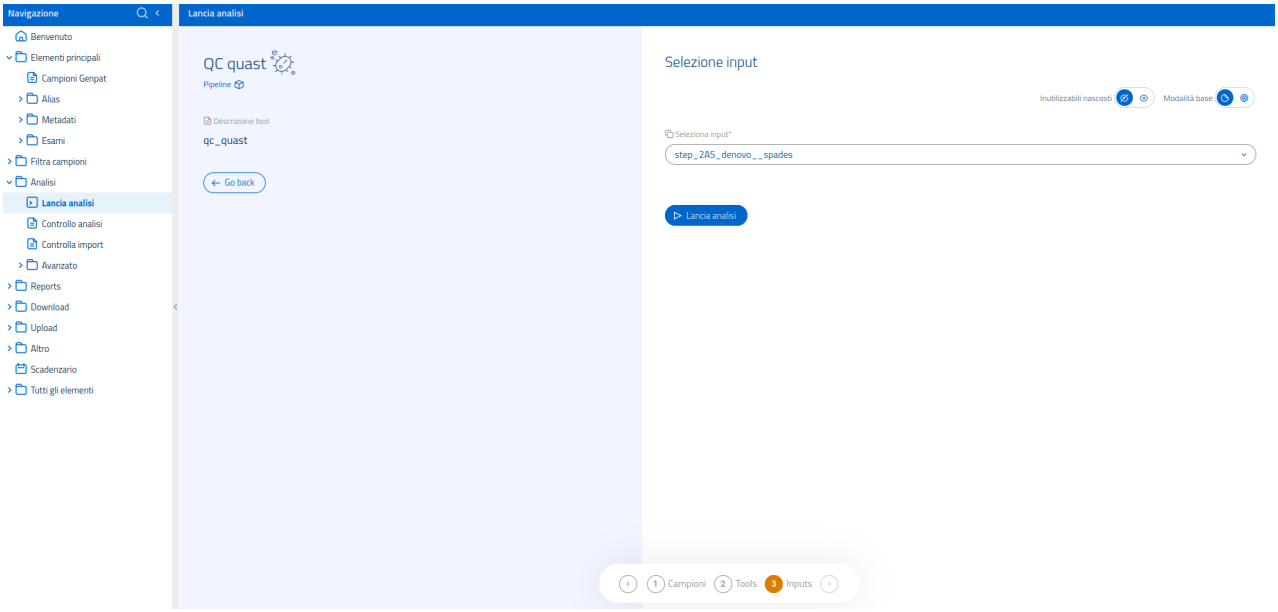
Lancia Quast

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **QC Quast** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

Gli input utilizzabili sono i files di assembly:

- *de novo* assembly (step_2AS_denovo)
- assembly ibrido (step_2AS_hybrid)
- *de novo* assembly da campioni di metagenomica (step_2MG_denovo)

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

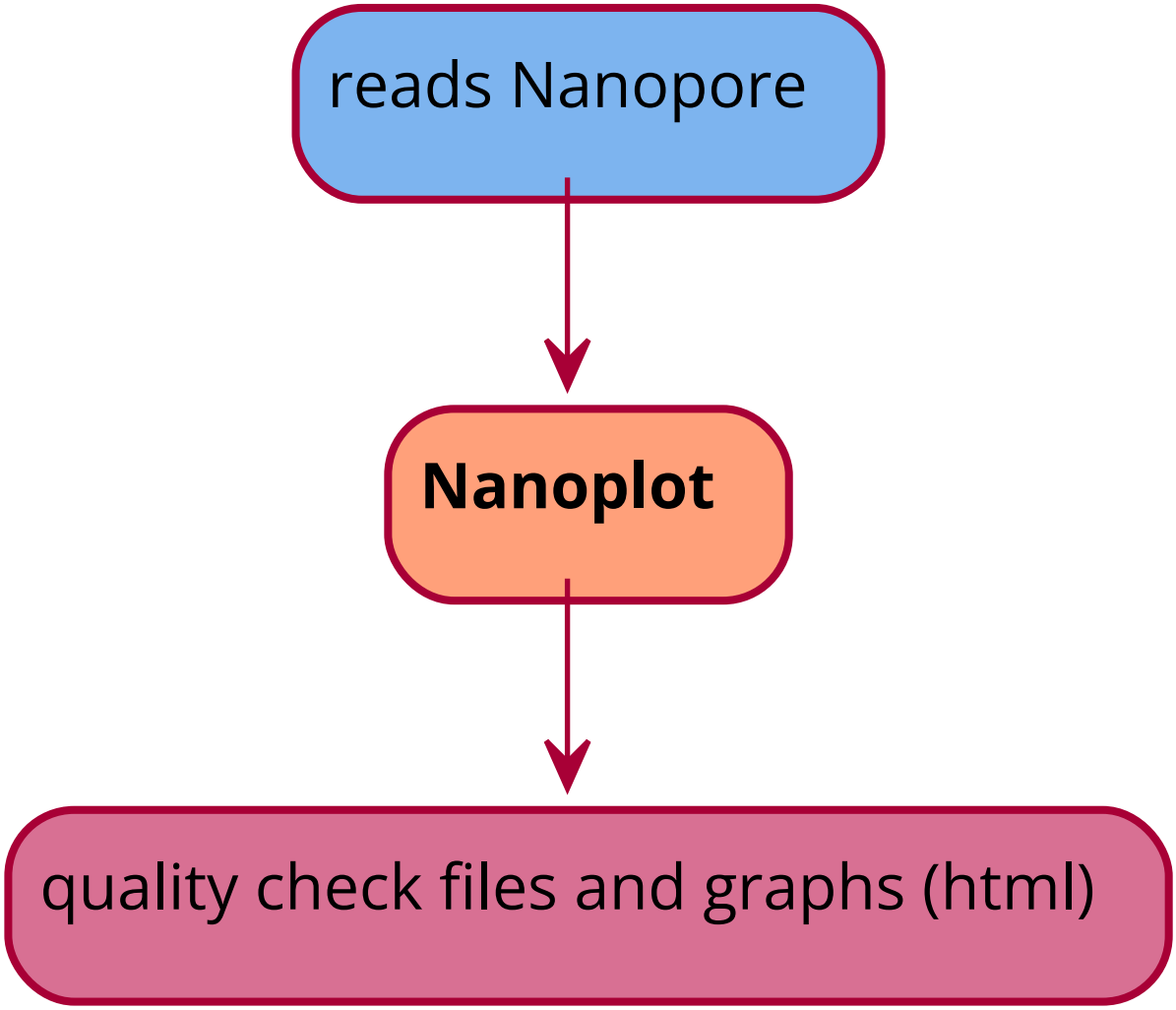
I files prodotti dalla pipeline "QC Quast" saranno gli stessi prodotti dallo stesso tool all'interno delle analisi che lo usano come software per il controllo qualità. Per consultare i risultati di Quast si invita quindi a fare riferimento alla tabella presente al seguente link agli outputs di 2AS_denovo:

- *Quality check* di Quast

QC Nanoplot

Introduzione

QC Nanoplot fa parte della sottocategoria di analisi per il controllo qualità (*Quality Check*, QC); essa esegue il software **Nanoplot** sulle *raw reads* da apparati Nanopore.



Input
Quality Check tool
Quality Check result

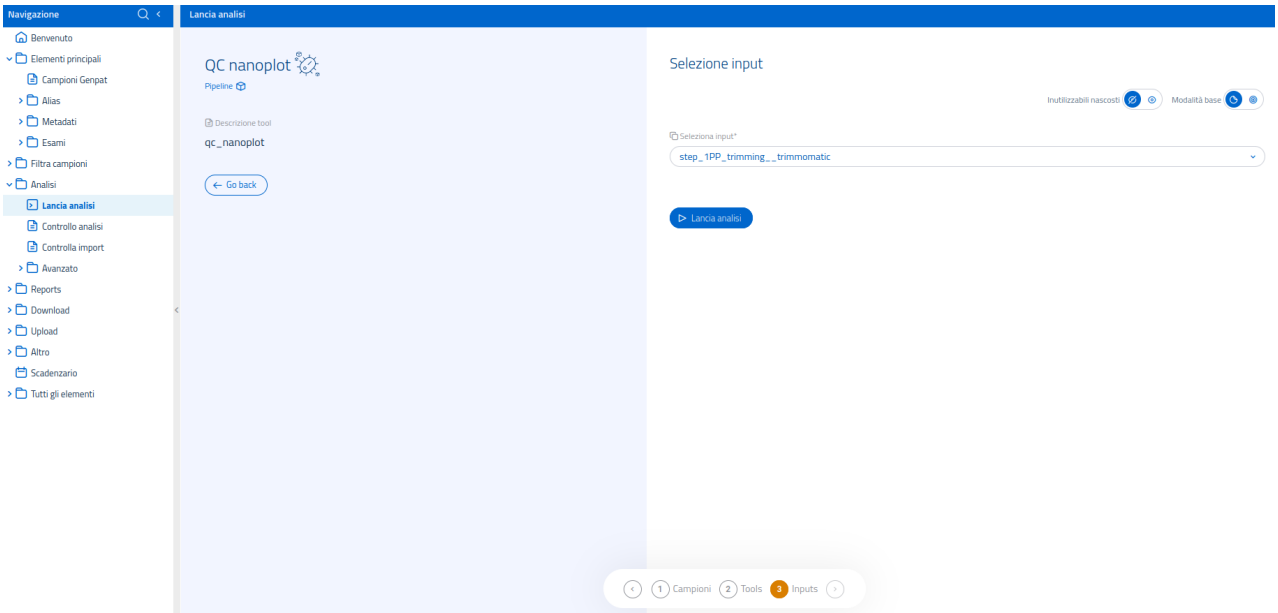
Lancia Nanoplot

Nel sistema di lancio analisi, è possibile usare il filtro in alto per visualizzare esclusivamente le pipelines. Una volta selezionata la pipeline **Nanoplot** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma.

Gli input utilizzabili sono le *raw reads* in formato fastq (interne o esterne) o le *reads* trimmate (o provenienti da analisi di preprocessamento), **purchè provengano da apparati Nanopore**:

- *raw reads* (step_0SQ_rawreads)
- *reads* trimmate (step_1PP_trimming)
- *reads* filtrate (step_1PP_filtering)
- *reads* fastq generate da fasta (step_1PP_generated)
- *reads* deplete delle *reads* dell'ospite (step_1PP_hostdepl)
- *reads* sottoposte a *downsampling* (step_1PP_downsampling)

L'interfaccia per la selezione dell'input mette a disposizione la modalità di selezione input avanzata, per permettere l'utilizzo di input processati da metodi diversi, usati a monte nel flusso di analisi.



Una volta lanciata la pipeline, la pagina genererà un link alla sezione Controllo analisi, per permettere di visualizzare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo la pipeline, sia al termine dell'esecuzione.

Risultati

> Per consultare la guida sul download dei files dalla piattaforma si faccia riferimento all'apposita pagina.

La cartella di output può essere esplorata cliccando sul link prodotto dalla pagina di download o sul link presente all'interno della scheda dell'analisi. All'interno della directory "Results" sarà possibile trovare le 2 cartelle:

- **meta**: ("metadata") in cui vengono salvati i files di log e di configurazione del processo eseguito.
- **result**: in cui sono salvati i files con i risultati prodotti dall'analisi.

La tabella seguente riporta l'elenco dei files prodotti da Nanoplot, insieme ad alcune informazioni utili.

File	Descrizione	Posizione
DSXXXXXX-DTXXXXXX_ID_LengthvsQualityScatterPlot_dot.html	scatterplot della lunghezza delle reads VS qualità media delle reads	cartella result
DSXXXXXX-DTXXXXXX_ID_LengthvsQualityScatterPlot_kde.html	scatterplot della lunghezza delle reads VS qualità media delle reads	cartella result
DSXXXXXX-DTXXXXXX_ID_NanoPlot-report.html	tabella delle statistiche	cartella result
DSXXXXXX-DTXXXXXX_ID_NanoStats.txt	tabella delle statistiche in formato di testo	cartella result
DSXXXXXX-DTXXXXXX_ID_Non_weightedHistogramReadlength.html	istogramma delle lunghezze delle reads	cartella result
DSXXXXXX-DTXXXXXX_ID_Non_weightedLogTransformed_HistogramReadlength.html	istogramma delle lunghezze delle reads dopo trasformazione logaritmica	cartella result
DSXXXXXX-DTXXXXXX_ID_WeightedHistogramReadlength.html	istogramma pesato delle lunghezze delle reads	cartella result
DSXXXXXX-DTXXXXXX_ID_WeightedLogTransformed_HistogramReadlength.html	istogramma pesato delle lunghezze delle reads dopo trasformazione logaritmica	cartella result
DSXXXXXX-DTXXXXXX_ID_Yield_By_Length.html	curva della resa cumulativa rispetto alla lunghezza delle reads	cartella result

Nota: quasi tutte le tabelle e i grafici prodotti da Nanoplot sono in formato html. Si consiglia di utilizzare un web browser per la visualizzazione.

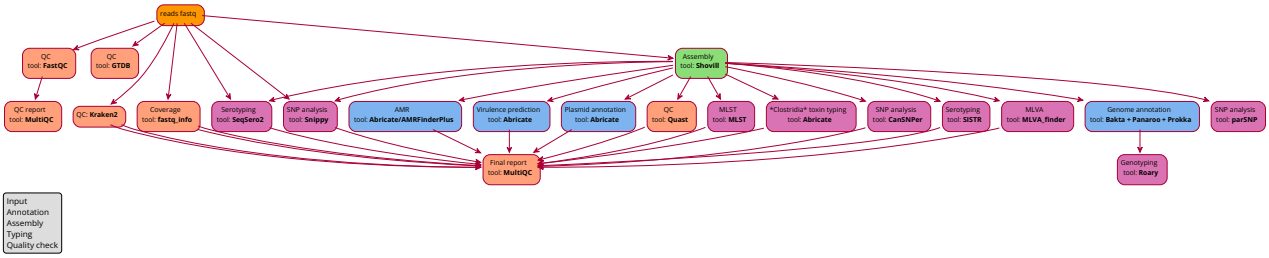
IAEA

WGSBAC Pipeline - IAEA

Introduzione

La pipeline WGSBAC effettua genotipizzazione e caratterizzazione degli isolati batterici, utilizzando come input i files fastq.gz da *whole-genome-sequencing* (WGS). Nella piattaforma sono implementati i workflows per *Salmonella* e *Brucella*. La pagina GitLab di WGSBAC permette di consultare una lista completa delle analisi disponibili per la pipeline.

- **Official WGSBAC page:** https://gitlab.com/FLI_Bioinfo/WGSBAC
- **Authors:** Jörg Linde (@joerg.linde) and Mostafa Abdel-Glil (@Mostafa.Abdel-Glil)



Lanciare la pipeline

La **Pipeline IAEA WGSBAC** è disponibile nella sezione Lancia Analisi della piattaforma come 2 pulsanti distinti: ono per *Brucella* ed uno per *Salmonella*.

Altri microorganismi che possono essere analizzati con la pipeline originale **non sono supportati dall'installazione nella presente piattaforma**.

La pipeline WGSBAC esegue una suite di tools bioinformatici suddivisibili in:

1. Quality Check (QC) delle reads e trimming
2. Ricostruzione della sequenza
3. QC dell'assemblaggio e controllo di contaminazione
4. Annotazione funzionale del genoma
5. Tipizzazione
6. Hierarchical clustering

Una lista dettagliata di tutti i softwares eseguibili dalla pipeline è presente nella pagina GitLab ufficiale di WGSBAC.

Dopo la selezione della versione della **Pipeline WGSBAC** per il microorganismo desiderato, sarà encessario scegliere i parametri dell'analisi.

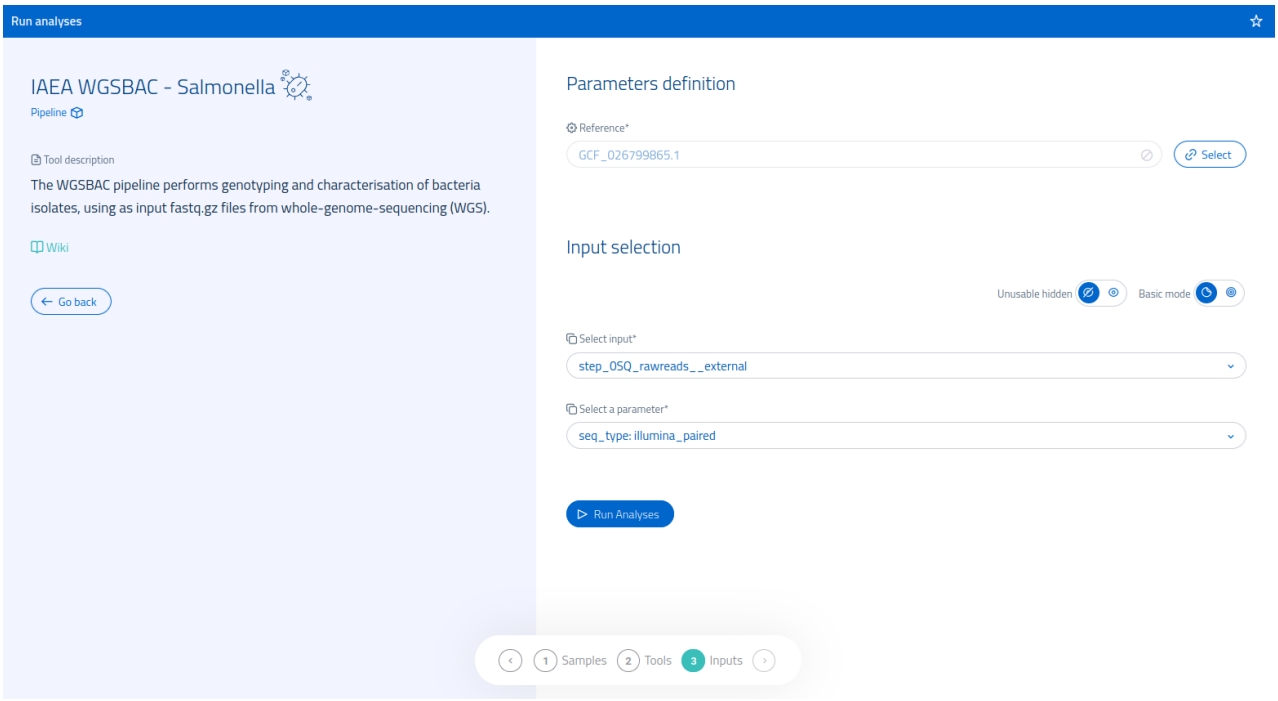
Gli input necessari sono reads fastq compresse (`.fastq.gz`), che possono essere fornite come:

- **step_0SQ_rawreads**
- **step_1PP_trimming**

> Si noti che l'opzione **step_1PP_trimming** come input viene mantenuta solo per comodità dell'utente, infatti la pipeline WGSBAC effettua di default il trimming con **fastp**.

Per poter eseguire alcune delle analisi della pipeline, è anche necessario specificare un genoma reference per il microorganismo selezionato.

> **Nota importante:** si ricordi che **la pipeline WGSBAC può essere utilizzata solo su reads Illumina paired end**. Non sono attualmente supportate altre tecnologie di sequenziamento.



Dopo aver confermato il lancio dell'analisi, la pagina genererà un link alla sezione Controlla Analisi, dove è possibile visualizzare lo stato di esecuzione dell'analisi. Il sistema manderà delle notifiche una volta terminata l'esecuzione.

Outputs

I files di output della pipeline WGSBAC includono gli outputs di tutte le analisi singole lanciate nel workflow. In basso sono elencate le cartelle dei risultati. Per una lista completa si rimanda alla pagina GitLab ufficiale WGSBAC: (https://gitlab.com/FLI_Bioinfo/WGSBAC).

Le cartelle dei risultati sono in `work > 0 > XX > XXXXXXX` (dove XX e XXXXXXX rappresentano dei nomi variabili per le cartelle: sono presenti 2 cartelle XX, una delle quali comprende i files di input, mentre la seconda la cartella `wgsbac_results`).

- Per l'esecuzione su campioni di *Brucella*:

Analysis - Tool/DB	**Results Directory**
contamination check - confindr	`wgsbac_results > confindr`
coverage calculation	`wgsbac_results > coverage`
trimming and QC - fastp	`wgsbac_results > fastp`
decompressed fastq files	`wgsbac_results > fastq`
raw reads quality check - fastqc	`wgsbac_results > fastqc`
reference fasta	`wgsbac_results > finalAssembly`
input fastq.gz files	`wgsbac_results > input`
contamination check - kraken2	`wgsbac_results > kraken`
Multi Lucus Sequence Typing - mlst	`wgsbac_results > mlst`
QC - multiQC	`wgsbac_results > multiqc`
QC - fastQC	`wgsbac_results > multiqc_fastqc`
plasmid identification - plasmidfinder/platon	`wgsbac_results > plasmids`
assembly QC - Quast	`wgsbac_results > quast`
de novo assembly - Shovill	`wgsbac_results > rawassembly_res`
variant calling - Snippy	`wgsbac_results > snippy`
species identification - Sourmash	`wgsbac_results > sourmash`

- Per l'esecuzione su campioni di *Salmonella*:

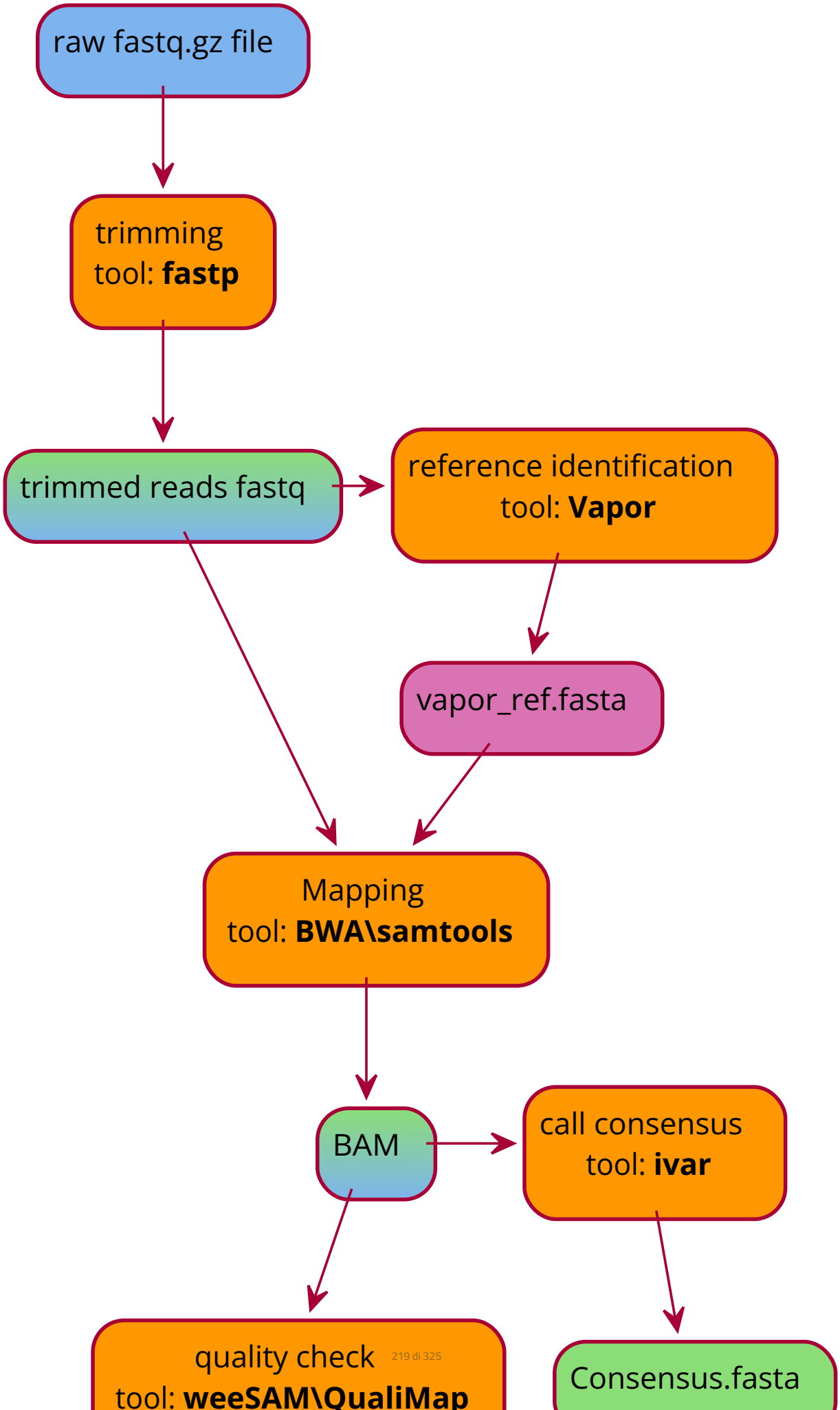
Analysis - Tool/DB	Results Directory
contamination check - confindr	`wgsbac_results > confindr`
coverage calculation	`wgsbac_results > coverage`
trimming and QC - fastp	`wgsbac_results > fastp`
decompressed fastq files	`wgsbac_results > fastq`
raw reads quality check - fastqc	`wgsbac_results > fastqc`
reference fasta	`wgsbac_results > finalAssembly`
input fastq.gz files	`wgsbac_results > input`
contamination check - kraken2	`wgsbac_results > kraken`
Multi Lucus Sequence Typing - mlst	`wgsbac_results > mlst`
QC - multiQC	`wgsbac_results > multiqc`
QC - fastQC	`wgsbac_results > multiqc_fastqc`
plasmid identification - plasmidfinder/platon	`wgsbac_results > plasmids`
assembly QC - Quast	`wgsbac_results > quast`
de novo assembly - Shovill	`wgsbac_results > rawassembly_res`
variant calling - Snippy	`wgsbac_results > snippy`
species identification - Sourmash	`wgsbac_results > sourmash`
virulence genes identification - SPI	`wgsbac_results > virulence`
Salmonella serotype prediction - seqSero	`seqSero`
Salmonella serovar prediction - sistr	`sistr`

LSDV Pipeline - IAEA

Introduction

This pipeline analyzes LSDV samples, automatically identifying the reference and reconstructing the consensus sequence.

- **Official tool page:** To be added
- **Authors:** Graham Freimanis and Richard Orton


raw fastq.gz file
trimming
tool: **fastp**
trimmed reads fastq
reference identification
tool: **Vapor**
vapor_ref.fasta
Mapping
tool: **BWA\samtools**
BAM
call consensus
tool: **ivar**
quality check
tool: **weeSAM\QualiMap**
Consensus.fasta

How to run

After selection of the **IAEA LSDV pipeline** in the Run Analysis page, a confirmation interface will allow to go on to the parameter selection.

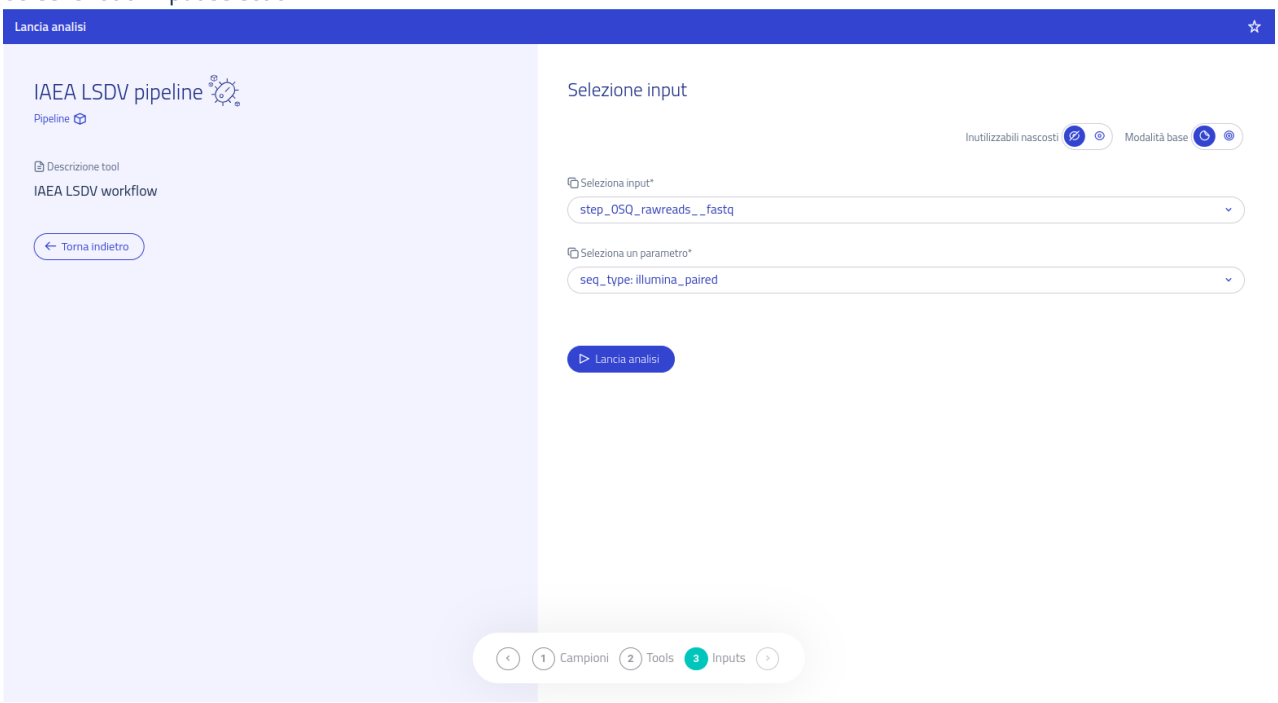
The **IAEA LSDV pipeline** pipeline consists of 3 main steps:

1. Reads trimming
2. Identification of the best reference
3. Mapping

The required input are raw fastq reads (`.fastq.gz`):

- **step_0SQ_rawreads**

Screenshot of input selection



After confirmation, the Run Analysis page will generate a link to the Check Analysis section. From there, users can visualize the execution status of the analysis. Automatic notifications will be sent after successfully running the pipeline and after execution has finished.

Outputs

The output directory is available at the link in the download page or at the link presente in the analysis' summary card, and will have the following structure: `results > YEAR > ID` . The last directory's suffix will be replaced with the name of the chosen tool. At that path there will be 2 directories:

- **meta**: ("metadata") contains log and configuration files.
- **result**: contains the analysis' output files.

All the results produced by all the pipeline tools can be found in the `result > LSDV-PIPELINE-RESULTS` folder.

Data visualization

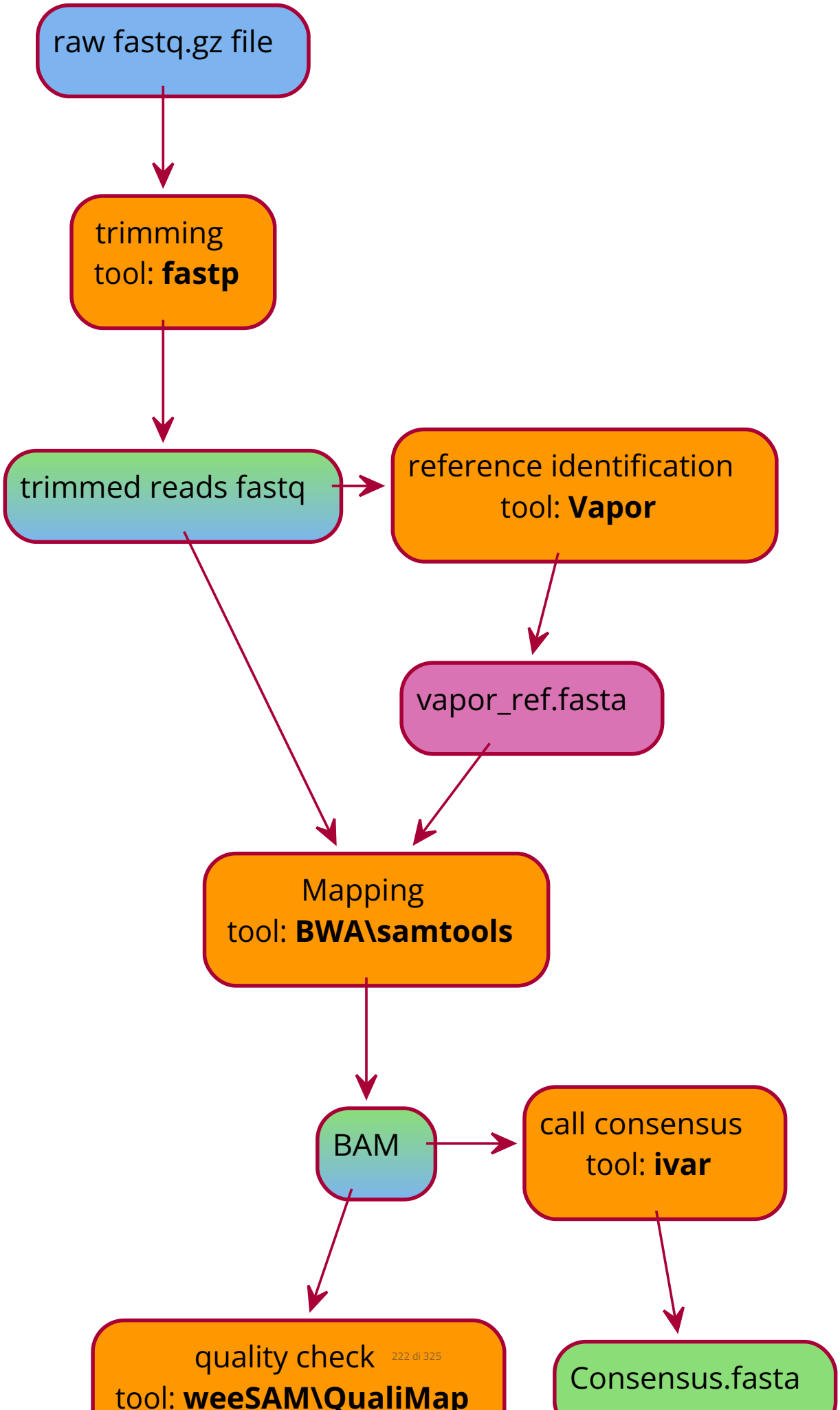
The pipeline produces an html webpage that can be opened in the browser

AIV Pipeline - IAEA

Introduzione

Questa pipeline elabora campioni di virus dell'influenza aviaria, determinando automaticamente la sequenza reference per ciascun segmento e ricostruendo la sequenza consensus corrispondente.

- **Pagina ufficiale del tool:** To be added
- **Autori:** Graham Freimanis e Richard Orton


raw fastq.gz file
trimming
tool: **fastp**
trimmed reads fastq
reference identification
tool: **Vapor**
vapor_ref.fasta
Mapping
tool: **BWA\samtools**
BAM
call consensus
tool: **ivar**
quality check
tool: **weeSAM\QualiMap**
Consensus.fasta

Come eseguire la pipeline

Una volta selezionata l'analisi **IAEA AIV pipeline** nella pagina dedicata al lancio di analisi, il sistema passerà ad un'interfaccia di conferma che permetterà di procedere con la selezione dei parametri.

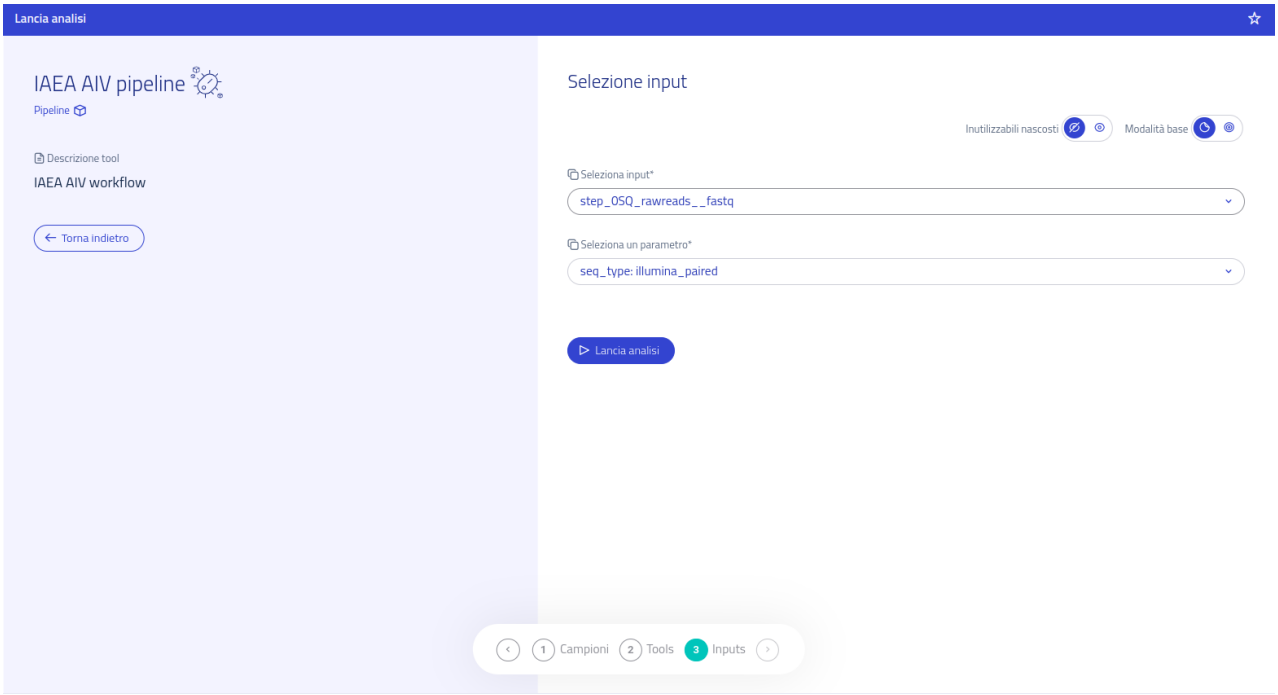
La **IAEA AIV pipeline** prevede 3 fasi principali:

1. *trimming* delle *read*;
2. identificazione del miglior reference;
3. calcolo del lineage e mapping.

Gli input richiesti sono *raw reads* in formato FASTQ (`.fastq.gz`):

- **step_0SQ_rawreads**

Screenshot della selezione degli input



Una volta lanciata l'analisi, la pagina genererà un link alla sezione Controllo analisi, per permettere di verificare lo stato del processo. L'utente verrà notificato dal sistema sia una volta lanciata con successo l'analisi, sia al termine dell'esecuzione.

Cartella dei risultati

La cartella dei risultati, **Result folder**, è accessibile cliccando sul link presente all'interno della scheda dell'analisi, nella sezione *Dati risultato*. All'interno della conseguente cartella `results` , è possibile trovare i files prodotti dalla pipeline "IAEA AIV" che saranno gli stessi prodotti dalle analisi che la compongono:

- *Trimming* con fastp
- Mapping con Ivar
- Calcolo del lineage con Vapor

Data visualization

La pipeline produce alcuni file di reportistica in formato .html che possono essere visualizzati attraverso il browser.

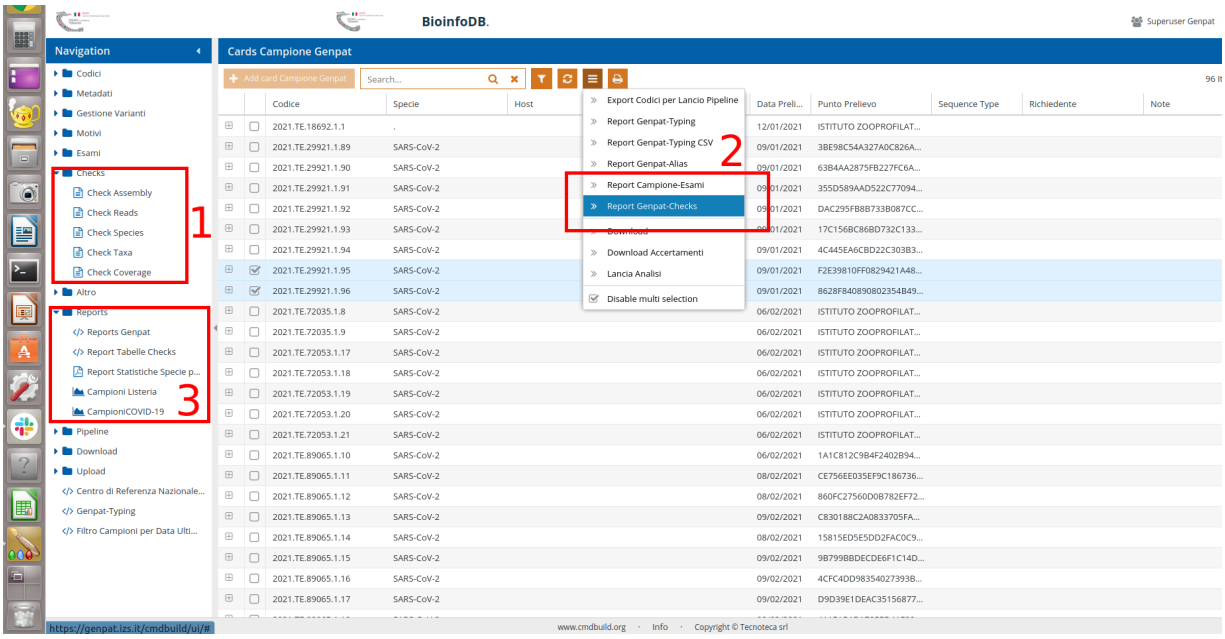
6. Report

6.1 Controlli qualità

6.1.1 Pagine dei controlli qualità

Dal menù di navigazione si può accedere ai vari Controlli Qualità disponibili **(1)**.

Per ogni controllo qualità, è possibile creare il report sia nella pagina dei report dal menù `Reports > Report Genpat` **(3)**, sia nel menù contestuale `Report Genpat-Checks` **(2)** presente nella pagina dei campioni.



I checks disponibili sono:

- Controlli qualità dell'Assembly
- Controlli qualità delle Read
- Controlli qualità delle Specie
- Controlli qualità dei Taxa
- Controlli qualità del Coverage
- Controlli qualità del cgMLST
- Controlli qualità di Kleborate
- Controlli qualità della classificazione tramite ML

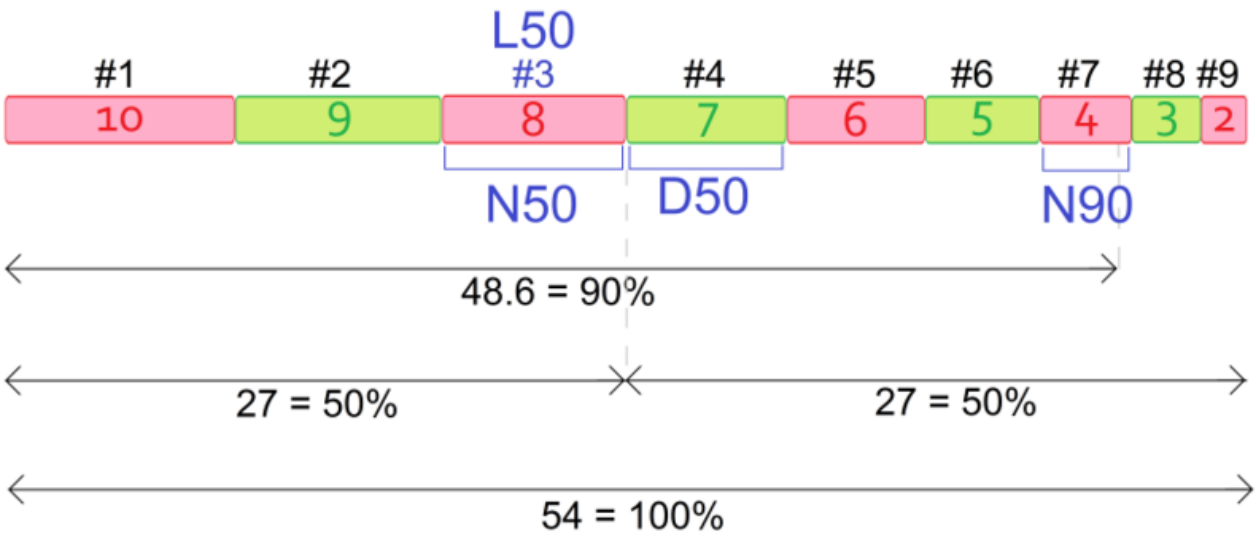
6.1.2 Controlli qualità dell'assemblaggio

La tabella *Controlli qualità dell'assemblaggio* fornisce informazioni sulla qualità degli assembly *de novo*.

Risultato	Campione	Contigs	Contig più lunga	Lunghezza Totale	N50	N75	L50	L75	NS PER 100KBP
150506-7112296-2AS_denovo-spades	DS7112296-1307-A01-2015-TE-5750-...	103	1293857	2983816	429669	233491	2	4	3.82
150506-7112297-2AS_denovo-spades	DS7112297-1308-A02-2015-TE-5750-...	87	1005536	2988668	505234	325286	2	4	4.28
150506-7112298-2AS_denovo-spades	DS7112298-1309-A03-2015-TE-5750-...	113	1475490	2920414	1475490	362614	1	3	10.34
150506-7112299-2AS_denovo-spades	DS7112299-1310-A04-2015-TE-5750-...	69	505234	2979887	321926	125746	4	8	1.04
150506-7112300-2AS_denovo-spades	DS7112300-1311-A05-2015-TE-5750-...	115	1049858	3071777	260884	228525	3	6	2.8
150506-7112301-2AS_denovo-spades	DS7112301-1312-A06-2015-TE-5750-...	114	915026	2999530	422072	236921	3	5	10.57
150506-7112302-2AS_denovo-spades	DS7112302-1313-A07-2015-TE-5750-...	78	592118	3021548	548065	321341	3	5	9.86
150506-7112303-2AS_denovo-spades	DS7112303-1314-A08-2015-TE-5750-...	137	645829	2909870	420668	237410	3	6	0.41
150506-7112304-2AS_denovo-spades	DS7112304-1315-A09-2015-TE-5750-...	262	1005536	3033457	325660	255199	3	5	7.25
150506-7112306-2AS_denovo-spades	DS7112306-1317-A11-2015-TE-5750-...	285	579753	3036676	311559	255330	4	6	0.56
150506-7112307-2AS_denovo-spades	DS7112307-1318-A12-2015-TE-5750-...	183	1212082	2994204	348982	247289	2	5	3.31
150506-7112308-2AS_denovo-spades	DS7112308-1319-B01-2015-TE-5750-...	115	653546	3039649	496496	265261	3	5	0.49
150506-7112309-2AS_denovo-spades	DS7112309-1320-B02-2015-TE-5750-...	105	623910	2986552	477699	430590	3	5	7.43
150506-7112310-2AS_denovo-spades	DS7112310-1321-B03-2015-TE-5750-...	135	1470539	2906110	1470539	267403	1	3	0.31
150506-7112311-2AS_denovo-spades	DS7112311-1322-B04-2015-TE-5750-...	83	505411	2982722	325995	128767	4	7	4.19
150506-7112312-2AS_denovo-spades	DS7112312-1323-B05-2015-TE-5750-...	75	1469945	2890604	1469945	517151	1	3	2.18
150506-7112313-2AS_denovo-spades	DS7112313-1324-B06-2015-TE-5750-...	82	702631	2987285	422194	217278	3	6	0.77
150506-7112314-2AS_denovo-spades	DS7112314-1325-B07-2015-TE-5750-...	167	802687	3080413	419303	148309	3	6	4.38
150506-7112315-2AS_denovo-spades	DS7112315-1326-B08-2015-TE-5750-...	136	1490203	2942819	1490203	320550	1	3	3.23
150506-7112316-2AS_denovo-spades	DS7112316-1327-B09-2015-TE-5750-...	230	1223820	3125447	518072	257936	2	5	6.46
150506-7112317-2AS_denovo-spades	DS7112317-1328-B10-2015-TE-5750-...	95	910561	2989218	495585	224922	3	5	3.91
150506-7112318-2AS_denovo-spades	DS7112318-1329-B11-2015-TE-5750-...	232	1005536	3026163	488004	325286	3	4	3.87
150506-7112319-2AS_denovo-spades	DS7112319-1330-B12-2015-TE-5750-...	245	1211536	3005985	372384	253969	2	5	4.06
150506-7112320-2AS_denovo-spades	DS7112320-1331-C01-2015-TE-5750-...	85	1006251	2991942	502233	233474	2	5	7.32

Le colonne dei metadati, oltre a **Codice** e **Descrizione** che si riferiscono all'identificativo del risultato dell'analisi, sono: * **Risultato**: codice del risultato (RIS_CD) al quale le statistiche si riferiscono; * **Campione Genpat**: codice GENPAT_CD del campione; * **Contig**: numero totale di contigs ottenuto dall'assemblaggio; * **Contig più lungo**: lunghezza del contig più lungo ottenuto dall'assemblaggio; * **Lunghezza Totale**: numero di nucleotidi totali che costituiscono la sequenza ricostruita dall'assemblaggio; * **N50**: lunghezza del contig con il quale si raggiunge il 50% della sequenza, ordinando i contig per lunghezza in maniera decrescente; * **N75**: lunghezza del contig con il quale si raggiunge il 75% della sequenza, ordinando i contig per lunghezza in maniera decrescente; * **L50**: numero di contig, ordinati in maniera decrescente per lunghezza, necessario per ottenere il 50% della sequenza; * **L75**: numero di contig, ordinati in maniera decrescente per lunghezza, necessario per ottenere il 75% della sequenza; * **NS per 100Kbp**: numero di basi incerte (N) per 100Kbp nella sequenza assemblata.

N50 e statistiche associate: il valore N50 rappresenta la lunghezza del contig più corto al 50% della lunghezza totale del genoma, se questi vengono ordinati dal più lungo al più corto (N50 su Wikipedia). Analogamente il N75 rappresenta la lunghezza del contig più corto necessario per raggiungere il 75% della lunghezza totale della sequenza (sempre considerando i contigs ordinati in modo decrescente per lunghezza).


L50
#1
#2
#3
#4
#5
#6
#7
#8
#9
10
9
8
7
6
5
4
3
2
N50
D50
N90
48.6 = 90%
27 = 50%
27 = 50%
54 = 100%

6.1.3 Controlli qualità delle read

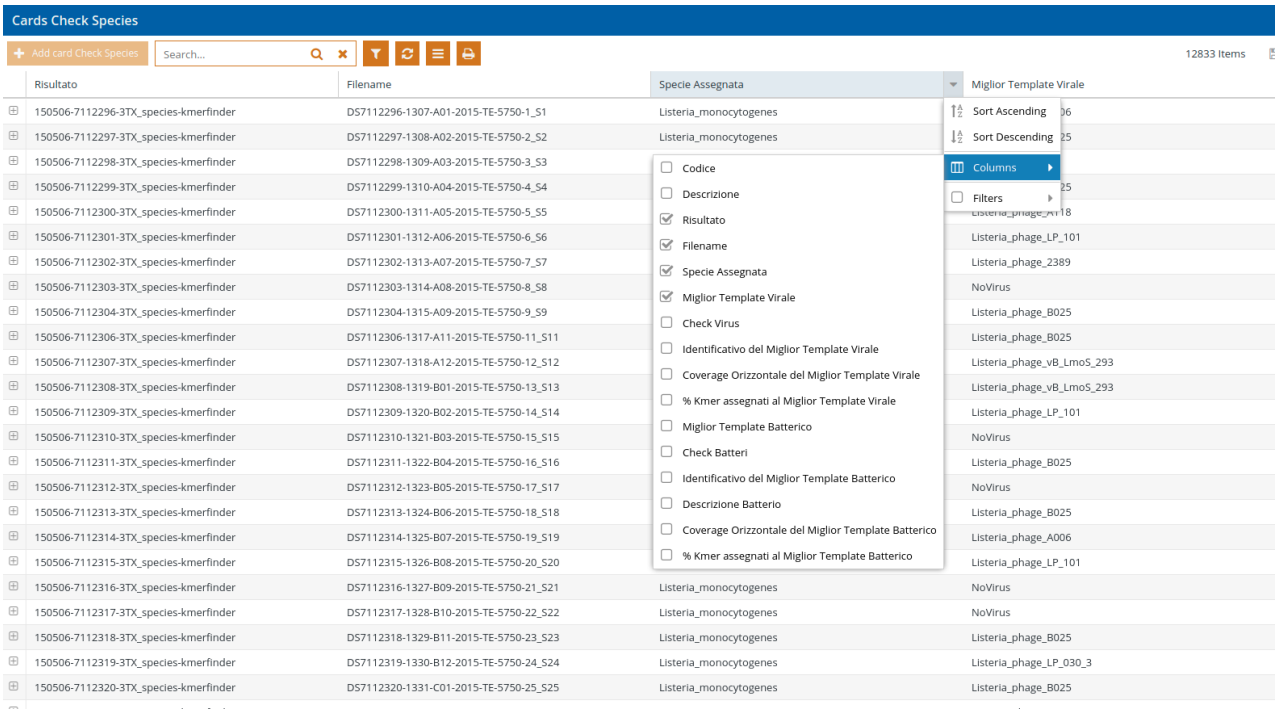
La tabella *Controlli qualità delle read* fornisce le informazioni e le statistiche di qualità relative alle read grezze e alle read trimmate.

Risultato	Campione	# Totale Raw Reads	% Reads Con Overlap ...	Q30 Coppie Reads Dop...	Qualità Media Coppie ...	Lunghezza Media Copp...
150506-7112296-0SQ_rawreads-fastq	DS7112296-1307-A01-2015-TE-5750-1_S1	6546860	37.82	94.14	34.15	133.12
150506-7112297-0SQ_rawreads-fastq	DS7112297-1308-A02-2015-TE-5750-2_S2	4064516	29.52	93.79	33.89	139.91
150506-7112298-0SQ_rawreads-fastq	DS7112298-1309-A03-2015-TE-5750-3_S3	4274866	24.41	93.34	33.67	139.1
150506-7112299-0SQ_rawreads-fastq	DS7112299-1310-A04-2015-TE-5750-4_S4	2977708	22.66	93.17	33.64	138.59
150506-7112300-0SQ_rawreads-fastq	DS7112300-1311-A05-2015-TE-5750-5_S5	4902212	31.36	94.14	33.94	139.06
150506-7112301-0SQ_rawreads-fastq	DS7112301-1312-A06-2015-TE-5750-6_S6	3351392	18.31	92.66	33.44	138.58
150506-7112302-0SQ_rawreads-fastq	DS7112302-1313-A07-2015-TE-5750-7_S7	3446192	22.65	93.1	33.64	139.05
150506-7112303-0SQ_rawreads-fastq	DS7112303-1314-A08-2015-TE-5750-8_S8	6078304	23.97	93.12	33.69	139.58
150506-7112304-0SQ_rawreads-fastq	DS7112304-1315-A09-2015-TE-5750-9_S9	8181924	30.1	92.94	33.74	139.14
150506-7112306-0SQ_rawreads-fastq	DS7112306-1317-A11-2015-TE-5750-11_S11	13296920	39.02	94.29	34.17	134.24
150506-7112307-0SQ_rawreads-fastq	DS7112307-1318-A12-2015-TE-5750-12_S12	9331576	37.63	94.19	34.13	135.31
150506-7112308-0SQ_rawreads-fastq	DS7112308-1319-B01-2015-TE-5750-13_S13	5461232	31.66	93.56	33.74	138.2
150506-7112309-0SQ_rawreads-fastq	DS7112309-1320-B02-2015-TE-5750-14_S14	4780072	31.36	94.01	33.84	138.8
150506-7112310-0SQ_rawreads-fastq	DS7112310-1321-B03-2015-TE-5750-15_S15	6121972	27.37	93.55	33.62	139.05
150506-7112311-0SQ_rawreads-fastq	DS7112311-1322-B04-2015-TE-5750-16_S16	3311792	18.17	92.9	33.37	138.27
150506-7112312-0SQ_rawreads-fastq	DS7112312-1323-B05-2015-TE-5750-17_S17	3476948	22.49	93.42	33.51	139.81
150506-7112313-0SQ_rawreads-fastq	DS7112313-1324-B06-2015-TE-5750-18_S18	3562558	15.97	92.54	33.23	138.75
150506-7112314-0SQ_rawreads-fastq	DS7112314-1325-B07-2015-TE-5750-19_S19	6355226	27.28	93.48	33.65	139.57
150506-7112315-0SQ_rawreads-fastq	DS7112315-1326-B08-2015-TE-5750-20_S20	5704952	21.67	93.2	33.5	139.59
150506-7112316-0SQ_rawreads-fastq	DS7112316-1327-B09-2015-TE-5750-21_S21	11212730	36.25	93.71	33.85	136.04
150506-7112317-0SQ_rawreads-fastq	DS7112317-1328-B10-2015-TE-5750-22_S22	3788216	24.14	93.08	33.53	138.62
150506-7112318-0SQ_rawreads-fastq	DS7112318-1329-B11-2015-TE-5750-23_S23	9549672	34.07	93.91	33.85	137.62
150506-7112319-0SQ_rawreads-fastq	DS7112319-1330-B12-2015-TE-5750-24_S24	9214680	33.11	93.99	33.85	137.95
150506-7112320-0SQ_rawreads-fastq	DS7112320-1331-C01-2015-TE-5750-25_S25	4796620	33.95	92.52	33.59	136.32

Le colonne dei metadati, oltre a **Codice** e **Descrizione** che si riferiscono all'identificativo del risultato dell'analisi, sono: * **Risultato**: codice del risultato (RIS_CD); * **Campione Genpat**: codice del campione (GENPAT_CD); * **Numero totale di read grezze**: numero totale delle read prima del trimming; * **Numero totale di megabasi grezze**: numero totale di nucleotidi nella sequenza grezza (in Mbp); * **% di coppie di read con overlap dopo il trimming**: percentuale di sovrapposizione, parziale o totale, tra le coppie di read dopo il trimming; * **Coppie di read con Q30 dopo il trimming**: coppie di read con qualità media (Phred score, Q) pari o superiore a 30 dopo il trimming; * **Qualità media delle coppie di read dopo il trimming**: qualità media delle coppie di read dopo il trimming; * **Lunghezza media delle coppie di read dopo il trimming**: lunghezza media delle coppie di read dopo il trimming; * **Numero totale di megabasi nelle coppie di read dopo il trimming**: numero di basi sequenziate nelle coppie di read rimaste dopo il trimming (in Mbp); * **Numero totale di coppie di read dopo il trimming**: numero totale delle coppie di read (paired-end) rimaste dopo il trimming; * **Numero totale di read non accoppiate dopo il trimming**: numero totale delle read (paired-end) singole rimaste dopo il trimming; * **Numero totale di read scartate dopo il trimming**: numero di read scartate dal processo di trimming; * **Numero totale di read trimmate**: numero di read dopo il trimming; * **Qualità media delle read grezze**: qualità media delle read prima del trimming (read grezze); * **Read grezze con Q30**: percentuale di read grezze con qualità media (Phred score, Q) pari o superiore a 30; * **Lunghezza media delle read grezze**; * **Attenzione**: codice dei controlli con risultati inferiori ai valori-soglia; * **Fallito**: codice del controllo fallito.

6.1.4 Controlli qualità della specie

La tabella *Controlli qualità della specie* fornisce le informazioni relative ai controlli di specie effettuati con (tipizzazione *in silico*).

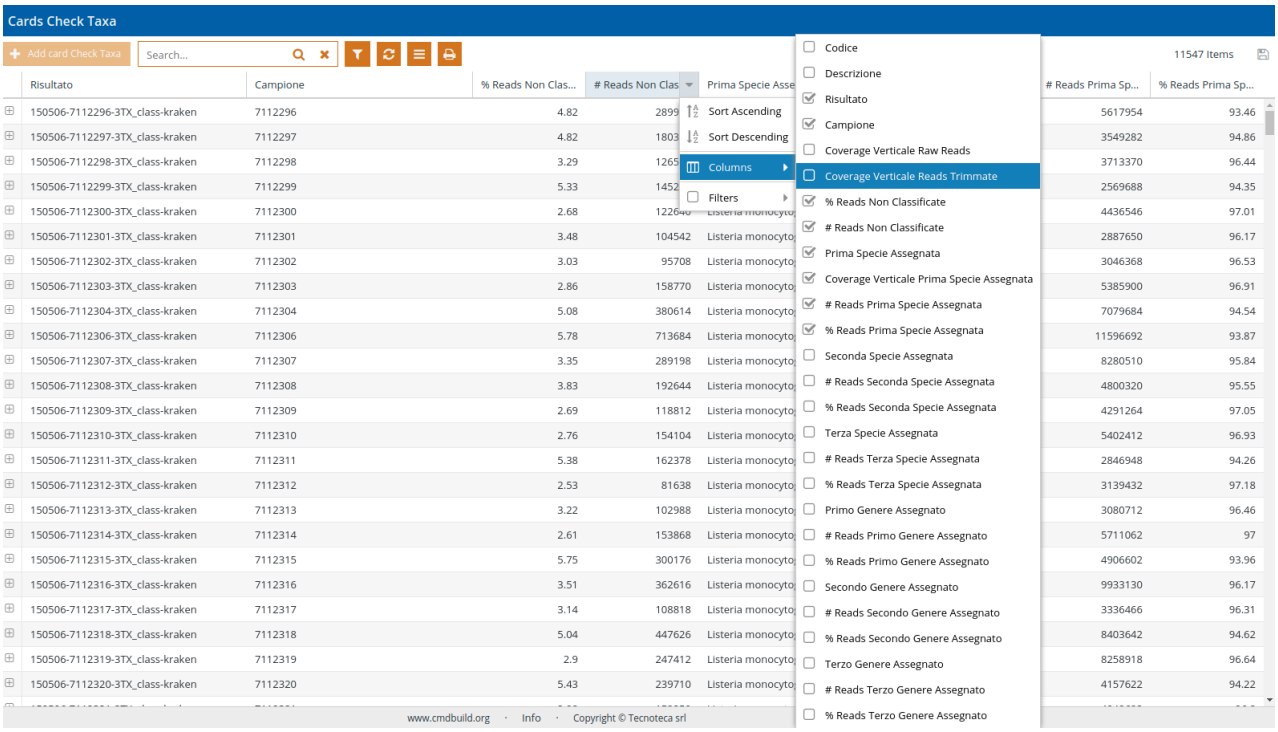


Le colonne dei metadati, oltre a **Codice** e **Descrizione** che si riferiscono all'identificativo del risultato dell'analisi, sono:

- **Risultato**: codice del risultato (RIS_CD);
- **Campione Genpat**: codice GENPAT_CD del campione;
- **Specie Assegnata**: la specie batterica assegnata da KmerFinder;
- **Miglior template batterico**: miglior template batterico assegnato da KmerFinder;
- **Verifica della presenza di specie batteriche**: controllo della presenza di sequenze di origine batterica;
- **Identificativo del miglior template batterico**: identificativo del miglior reference batterico nel database GenPat;
- **Descrizione del batterio**: nome e tipo del batterio;
- **Coverage orizzontale del miglior template batterico**: percentuale della sequenza ottenuta utilizzando, come reference, il miglior template batterico individuato;
- **% di K-mer assegnati al miglior template batterico**: percentuale di k-mer assegnati al miglior template batterico .

6.1.5 Controlli qualità dei taxa

La tabella *Controlli qualità dei taxa* fornisce informazioni relative alle possibili contaminazioni presenti nel campione, valutate attraverso specifiche analisi *in silico*.



Le colonne dei metadati, oltre a **Codice** e **Descrizione** che si riferiscono all'identificativo del risultato dell'analisi, sono:

- **Risultato**: codice del risultato (RIS_CD) al quale le statistiche si riferiscono;
- **Campione Genpat**: codice GENPAT_CD del campione;
- **Coverage verticale delle read grezze**: coverage verticale ipotetico delle read grezze;
- **Coverage verticale delle read trimmate**: coverage verticale ipotetico delle read trimmate;
- **% di read non classificate**: percentuale di read non assegnate ad alcun microorganismo;
- **Numero di read non classificate**: numero di read non assegnate ad alcun microorganismo;
- **Prima specie assegnata**: prima specie identificata (la più presente);
- **Coverage verticale della prima specie assegnata**: coverage verticale del genoma della prima specie assegnata (considerando le read trimmate);
- **Numero di read assegnate alla prima specie**: numero di read assegnate alla prima specie identificata;
- **% di read assegnate alla prima specie**: percentuale di read assegnate alla prima specie identificata;
- **Seconda specie assegnata**: seconda specie identificata;
- **Numero di read assegnate alla seconda specie**: numero di read assegnate alla seconda specie identificata;
- **% di read assegnate alla seconda specie**: percentuale di read assegnate alla seconda specie identificata;
- **Terza specie assegnata**: terza specie identificata;
- **Numero di read assegnate alla terza specie**: numero di read assegnate alla terza specie identificata;
- **% di read assegnate alla terza specie**: percentuale di read assegnate alla terza specie identificata;
- **Primo genus assegnato**: primo genus identificato;
- **Numero di read assegnate al primo genus**: numero di read assegnate al primo genus identificato;
- **% di read assegnate al primo genus**: percentuale di read assegnate al primo genus identificato;
- **Secondo genus assegnato**: secondo genus identificato;
- **Numero di read assegnate al secondo genus**: numero di read assegnate al secondo genus identificato;
- **% di read assegnate al secondo genus**: percentuale di read assegnate al secondo genus identificato;
- **Terzo genus assegnato**: terzo genus assegnato;
- **Numero di read assegnate al terzo genus**: numero di read assegnate al terzo genus identificato;
- **% di read assegnate al terzo genus**: percentuale di read assegnate al terzo genus identificato.

6.1.6 Controlli qualità del coverage

La tabella *Controlli qualità del coverage* fornisce informazioni sulla qualità dei mapping.

Risultato Esame	Campione Genpat	DS	Coverage Verticale	Coverage Orizzon...	Nr. Reads Mappate	Min Coverage Ver...	Max Coverage Ver...
210216-11277177-2AS_mapping-bowtie	2021.TE.29921.1.89	11277177	970.499	0.0790891	16646	1	5450
210216-11277178-2AS_mapping-bowtie	2021.TE.29921.1.90	11277178	4846.52	0.999933	1022439	2	8000
210216-11277179-2AS_mapping-bowtie	2021.TE.29921.1.91	11277179	3463.8	0.999699	709880	1	7984
210216-11277180-2AS_mapping-bowtie	2021.TE.29921.1.92	11277180	5001.54	0.999866	1079673	1	8001
210216-11277181-2AS_mapping-bowtie	2021.TE.29921.1.93	11277181	5151.71	0.998997	1130544	1	8000
210216-11277182-2AS_mapping-bowtie	2021.TE.29921.1.94	11277182	4690.7	0.999933	982487	1	7999
210216-11277183-2AS_mapping-bowtie	2021.TE.29921.1.95	11277183	1821.76	0.982878	367602	1	5468
210216-11277184-2AS_mapping-bowtie	2021.TE.29921.1.96	11277184	4153.99	0.999933	861140	1	7999
210216-11467945-2AS_mapping-bowtie	2021.TE.72035.1.8	11467945	2444.87	0.983848	495740	1	7997
210216-11467946-2AS_mapping-bowtie	2021.TE.72035.1.9	11467946	4242.48	0.983915	869857	1	7999
210216-11468081-2AS_mapping-bowtie	2021.TE.72053.1.17	11468081	3354.51	0.982577	678061	0	7999
210216-11468082-2AS_mapping-bowtie	2021.TE.72053.1.18	11468082	4256.92	0.995184	886836	1	8002
210216-11468083-2AS_mapping-bowtie	2021.TE.72053.1.19	11468083	2211.78	0.967696	455238	1	7999
210216-11468084-2AS_mapping-bowtie	2021.TE.72053.1.20	11468084	2876.38	0.982778	610874	1	8000
210216-11468085-2AS_mapping-bowtie	2021.TE.72053.1.21	11468085	4263.93	0.996355	883503	1	7999
210216-11470165-2AS_mapping-bowtie	2021.TE.89065.1.2	11470165	566.457	0.987426	114667	1	2237
210216-11470166-2AS_mapping-bowtie	2021.TE.89065.1.3	11470166	627.792	0.98241	126458	1	2422
210216-11470167-2AS_mapping-bowtie	2021.TE.89065.1.4	11470167	668.549	0.98134	134479	0	2573
210216-11470168-2AS_mapping-bowtie	2021.TE.89065.1.5	11470168	770.229	0.990402	156382	0	2874
210216-11470169-2AS_mapping-bowtie	2021.TE.89065.1.6	11470169	751.727	0.981106	151320	0	2893
210216-11470170-2AS_mapping-bowtie	2021.TE.89065.1.7	11470170	802.643	0.985119	162151	0	3097
210216-11470171-2AS_mapping-bowtie	2021.TE.89065.1.8	11470171	863.58	0.982343	173907	0	3116
210216-11470172-2AS_mapping-bowtie	2021.TE.89065.1.9	11470172	1436.4	0.990904	291729	0	5040
210216-11470173-2AS_mapping-bowtie	2021.TE.89065.1.10	11470173	792.833	0.995954	161731	0	2977

La tabella, oltre alle colonne **Codice** e **Descrizione** che si riferiscono all'identificativo del risultato dell'analisi, è composta dalle seguenti colonne di metadati: * **Risultato analisi**: codice del risultato dell'analisi (RIS_CD); * **Campione Genpat**: codice del campione (GENPAT_CD); * **Campione DS**: identificatico univoco interno del campione sequenziato; * **Note**: codice della sequenza reference utilizzata; * **Coverage verticale**: (o *depth of coverage*) è il numero di volte in cui il genoma viene coperto dalle read mappate, considerando la media di tutte le posizioni (*coverage verticale medio*); * **Coverage orizzontale**: (o *breadth of coverage*) è la percentuale di genoma coperta dal mapping delle read rispetto alla lunghezza del genoma reference; * **Numero di read mappate**: numero totale di read mappate sul reference; * **Coverage verticale minimo**: numero minimo di volte in cui una singola posizione viene coperta dalle read mappate; * **Coverage verticale massimo**: numero massimo di volte in cui una singola posizione viene coperta dalle read mappate; * **%IUPAC**: percentuale di caratteri IUPAC all'interno della sequenza ricostruita/assemblata; * **%Ns**: percentuale di caratteri IUPAC "N" all'interno della sequenza ricostruita/assemblata; * **Lunghezza consensus**: lunghezza della sequenza consensus ottenuta.

Link utili:

- Coverage su Wikipedia
- Coverage su sequencing.com
- Coverage su illumina.com

6.1.7 Controlli qualità del cgMLST

La tabella dei *Controlli qualità del cgMLST* fornisce informazioni sulla qualità del core genome Multi Locus Sequence Typing.

Le colonne dei metadati riportano il **Codice** dell'analisi e il codice del **Campione** a cui si riferiscono i dati. Le altre colonne elencano le statistiche aggregate degli alleli chiamati da **chewBBACA** (https://github.com/B-UMMI/chewBBACA): * **% di chiamati**: percentuale degli alleli chiamati da chewBBACA rispetto al totale dei *loci* nello schema; * **Numero chiamati**: numero degli alleli chiamati da chewBBACA; * **Annotati**, **Nuovi**, **Non trovati**, **Scartati**: ognuna di queste colonne riporta il numero di alleli in base a come sono stati chiamati o classificati da chewBBACA; - Annotati: varianti alleliche già presenti nello schema; - Nuovi: nuovi alleli, non presenti nello schema, ai quali sono stati assegnati dei nuovi codici; - Non trovati: (*Locus* Not Found - LNF), i *loci* che non sono stati trovati nel core genome del campione; - Scartati: *loci* per i quali la chiamata allelica non è sufficientemente affidabile.

Nota: il numero di *loci* dipende dallo schema di chewBBACA **per lo specifico microrganismo**.

6.1.8 Controlli qualità di Kleborate

La tabella dei *Controlli qualità Kleborate* riporta i risultati della pipeline Kleborate per lo screening di *Klebsiella pneumoniae*.

I dati sono organizzati per tipologia. Il numero di colonne è elevato, pertanto si consiglia di visualizzare i dati aggiuntivi aprendo la scheda dell'item di interesse o facendo uso della gestione delle colonne. Vi ricordiamo che l'apertura di un singolo item permetterà la visualizzazione di tutti i metadati disponibili.

- La tipologia **Controllo qualità di specie** comprende colonne come **Specie** e **Match di specie** che riportano, rispettivamente, la specie identificata da Kleborate e il grado di associazione;
- la tipologia **Controllo qualità dell'assembly** comprende le colonne dei dati di qualità riguardanti l'assemblaggio, come la **Lunghezza totale**, il **Numero di contigs**, il **Contig più lungo** ed **N50** (vedi la pagina sul controllo qualità dell'assemblaggio);
- la tipologia **MLST** comprende colonne con informazioni derivanti dal Multi *Locus* Sequence Typing, come il Sequence Type (ST);
- la tipologia **Resistenze** comprende le colonne con i dati riguardanti i geni di resistenza del campione a vari composti;
- la tipologia **Virulenze** comprende le colonne con i dati riguardanti i geni di virulenza;
- la tipologia **Dati di base** contiene colonne con informazioni aggiuntive e dati riguardanti altri geni codificanti per proteine di interesse.

6.1.9 Controlli qualità della classificazione tramite ML

La tabella del *controllo qualità della classificazione tramite ML* contiene il report principale della predizione binaria della sorgente alimentare per campioni di *Listeria* (tramite 4TY_ML meat classifier).

La tabella è costituita dai seguenti campi: * **Campione**: codice del campione Genpat; * **Matrice Genpat**: matrice di origine; * **Predizione**: risultato della predizione binaria della sorgente alimentare ('MEAT'/'NOT_MEAT'); * **Probabilità**: valore percentuale della classificazione che ha determinato la predizione.

7. Download

7.1 Procedura di Download

Dalla sezione **Download** del menù laterale, l'utente può accedere alle funzionalità omonime.

La pagina *Download* condivide interfaccia e funzionamento con quella per il lancio di analisi; ciò implica, dunque, che il metodo standard per scaricare dei file dalla piattaforma preveda l'uso di una raccolta di codici già presenti nel *carrello* o in un *tag* attivo. La scelta tra tag o carrello avviene prima di poter proseguire ed è necessaria per specificare i campioni a cui sono associati i file da scaricare.

Il sistema di download permette di: * aggiungere campioni al carrello oppure attivare un tag, tramite apposite scorciatoie; * usare la funzione di **Selezione Rapida** dei campioni che consente di riempire velocemente il carrello usando un file in formato CSV oppure elencando i codici nell'apposito campo; * scegliere tra due categorie di download: - **download Single sample**, che permette di scegliere liberamente i risultati di tutte le analisi lanciate con qualsiasi tool; - **download Preset**, che fornisce delle scorciatoie per un veloce download dei risultati di analisi specifiche; * filtrare i file da scaricare.

Il video che segue fornisce una breve guida all'esecuzione del download:

Esecuzione del download

> I risultati delle analisi potranno essere scaricati unicamente se precedentemente **importati**.

Analogamente al lancio di un'analisi, il carrello viene attivato per proseguire con il download; inoltre sono anche disponibili le funzionalità di **Selezione Rapida** e **Carica RIS_CD**.

La **Selezione Rapida** consente di popolare rapidamente il carrello con campioni di interesse. **Carica RIS_CD** permette di utilizzare il codice univoco di una o più analisi lanciate, dando facile accesso a risultati di analisi meno recenti.

> Ulteriori informazioni sull'uso del RIS_CD per il lancio analisi e per le funzioni di download sono disponibili nella sezione della guida "Esportazione del RIS_CD".

Il lancio delle analisi genera un link alla scheda del processo dell'analisi richiesta e, in determinati casi, da tale scheda è possibile accedere ad alcuni files e opzioni di visualizzazione o download dei dati. Per ulteriori informazioni sul flusso di lavoro, è possibile consultare le seguenti pagine Wiki:

- Fasi del sistema di lancio analisi;
- Controllo analisi.

7.1.1 Preset

Le procedure per i download "**Single sample**" e "**Preset**" sono identiche. I download **Preset** sono scorciatoie che:

- permettono di evitare di specificare tutti i parametri necessari;
- sono disponibili solo per una selezione di combinazioni (le più usate) tra le analisi possibili.

La funzionalità **Preset** è pensata per facilitare il flusso di lavoro per gli utenti che eseguono di routine il download di file specifici: i parametri sono già preimpostati, il che permette di scorrere più rapidamente e con meno click tra le interfacce.

7.1.2 Visualizzazione dei download

Nella sezione **Visualizza download** del menù laterale è possibile consultare le schede dei processi di download lanciati.

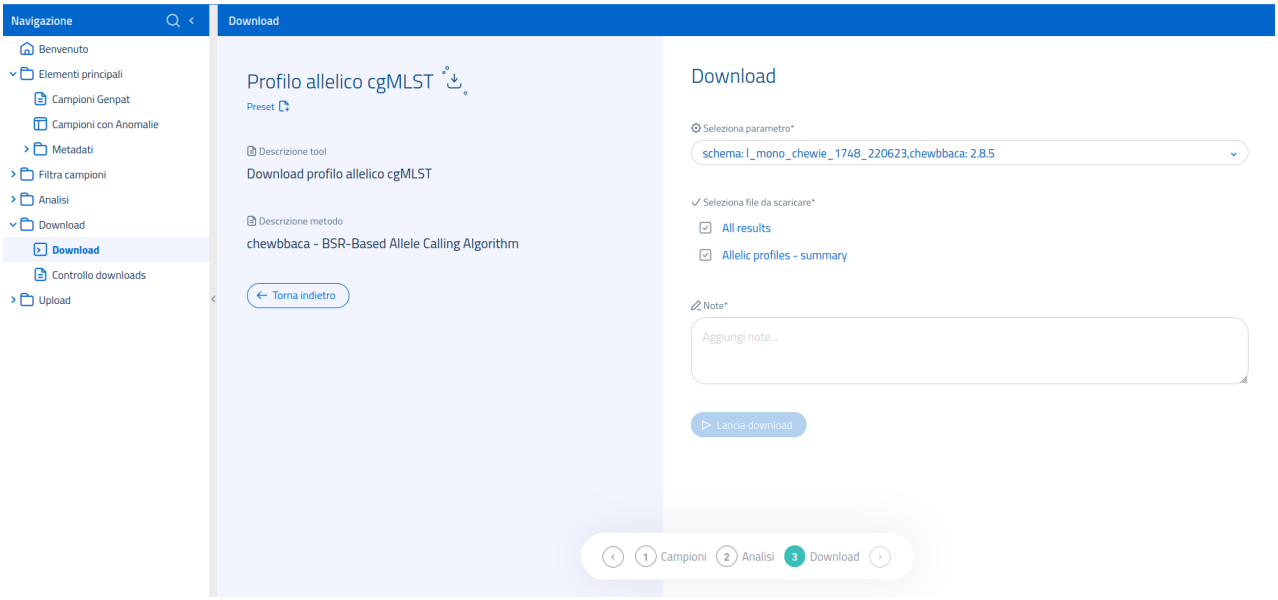
La pagina **Visualizza Downloads** conserva la cronologia dei download dell'utente. Ogni elemento (scheda) contiene informazioni utili riguardo il processo effettuato, tra cui identificativi, links e comando `wget` per poter ripetere il download.

Analogamente agli elementi nelle altre tabelle in piattaforma, ogni scheda è espandibile e contiene dettagli addizionali.

Note per l'utente

Sia le funzionalità di download *Preset* che la selezione dei files da scaricare sono orientate a svilupparsi su misura delle necessità degli utenti. Nel caso sussista il bisogno di aggiungere un download **Preset** o un **filtro** sui files da scaricare per un'analisi, gli utenti possono farne richiesta al reparto Bioinformatica.

Nota per gli utenti con profilo IZS: unicamente per gli utenti con profilo di tipo "IZS" (ad esempio "IZS Listeria"), l'ultima fase per il lancio del download di profili allelici e di *de novo* assembly presenterà un campo "Note", come mostrato nell'immagine in basso. Il tasto per il download sarà disabilitato e tali utenti dovranno scrivere una breve nota per comunicare il motivo del download. Una volta aggiunta la nota obbligatoria, il tasto "*Lancia download*" diventerà cliccabile.



7.2 Download Massivo per Utenti Windows

In questa pagina sono elencate alcune soluzioni per sopperire all'impossibilità, per gli utenti Windows, di eseguire nativamente il comando `wget` generato dalla piattaforma per il download di tutti i files.

7.2.1 Metodo 1: wget.exe

Il metodo riportato in basso consiste in un espediente per l'esecuzione del comando UNIX `wget`, in modo da scaricare tutti i files di un'analisi in una singola azione.

1. Scaricare il programma "wget.exe" per Windows tramite il link https://eternallybored.org/misc/wget/1.21.4/64/wget.exe;
2. Salvare (o spostare) il file wget.exe scaricato nella propria cartella "Download";
3. Aprire la command line Windows (**cmd**, disponibile nel menù Start) e spostarsi nella cartella "Download" (scrivere il comando in basso e premere "Invio"):

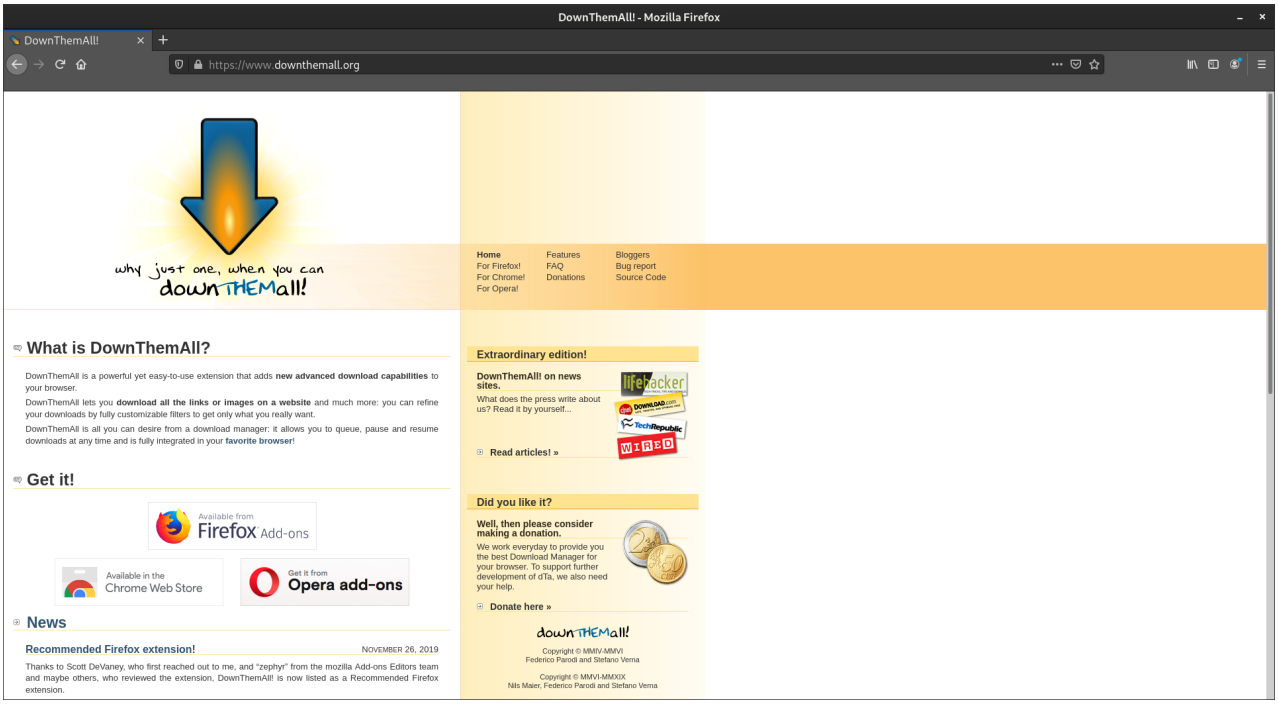
   ```
   cd Download
   ```

4. Copiare il comando `wget` generato dalla pagina di download della piattaforma nel Command Prompt e lanciarlo;
5. Al termine della procedura di download, i files si troveranno in una cartella con il nome fornito dal comando lanciato (ad esempio "f0086d8f-948d-4753-ad85-5171f1a5dda8"), situata nella cartella "Download".

7.2.2 Metodo 2: Estensione per Browser "DownThemAll"

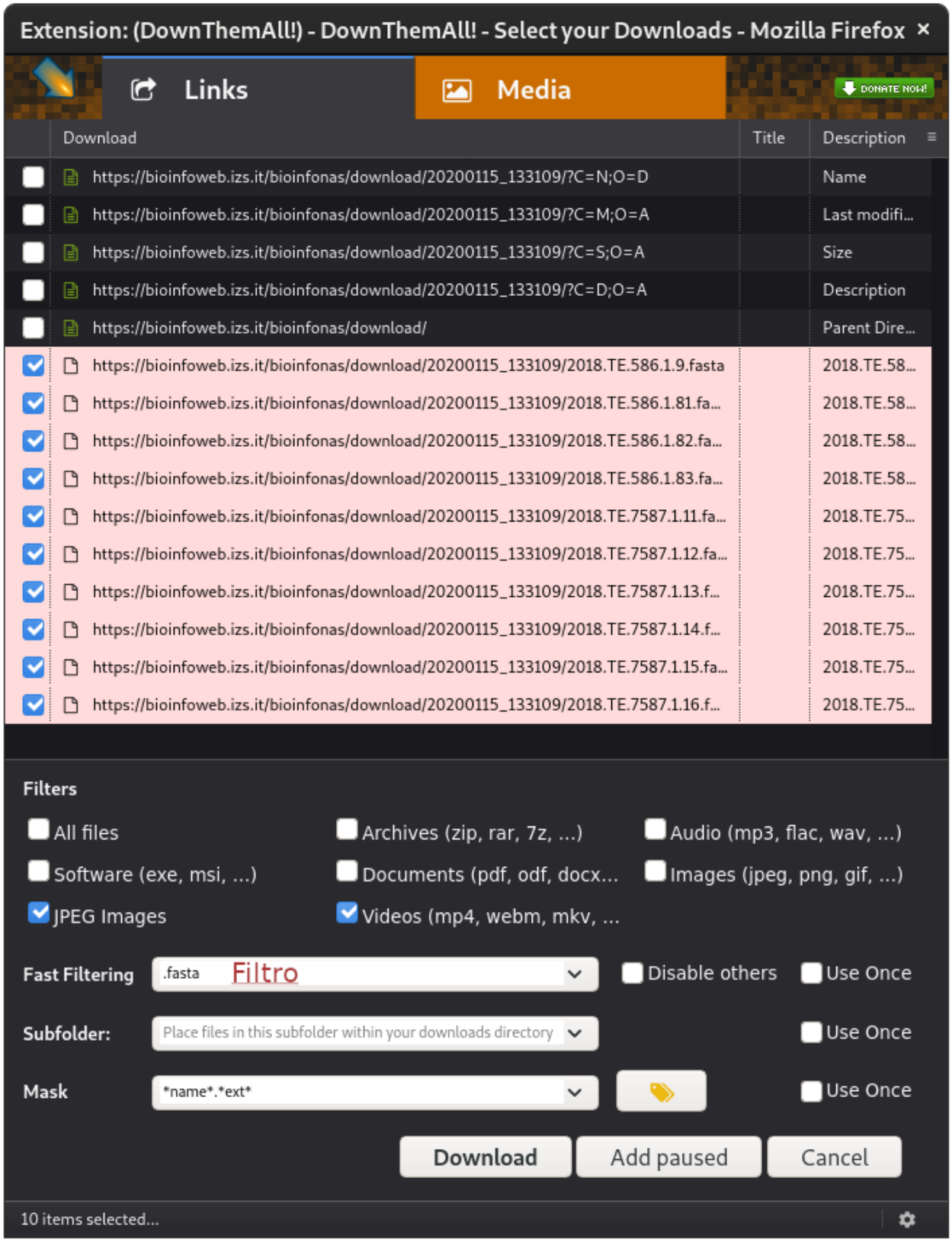
L'estensione per web browser "DownThemAll!" può essere aggiunta a Firefox, Chrome ed Edge e permette il download di più files da una pagina web.

1. "DownThemAll!" è disponibile al seguente indirizzo: https://www.downthemall.org/;
2. Nella sezione "Get it!" è possibile scegliere la versione per il proprio browser e installarla:



1. Una volta installata l'estensione, quando si esplora una cartella dei risultati sarà possibile scaricare i files elencati nella pagina, usando la funzione "DownloadThemAll!" (raggiungibile cliccando sull'icona del software, in genere in alto a destra nella barra delle funzioni del browser):

1. La finestra che si aprirà permette di filtrare o selezionare i files prima di effettuarne il download tramite l'omonimo pulsante:


Extension: (DownThemAll!) - DownThemAll! - Select your Downloads - Mozilla Firefox
Links
Media
DONATE NOW!
Download
Title
Description
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/?C=N;O=D
Name
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/?C=M;O=A
Last modifi...
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/?C=S;O=A
Size
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/?C=D;O=A
Description
https://bioinfoweb.izs.it/bioinfonas/download/
Parent Dire...
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/2018.TE.586.1.9.fasta
2018.TE.58...
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/2018.TE.586.1.81.fa...
2018.TE.58...
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/2018.TE.586.1.82.fa...
2018.TE.58...
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/2018.TE.586.1.83.fa...
2018.TE.58...
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/2018.TE.7587.1.11.fa...
2018.TE.75...
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/2018.TE.7587.1.12.fa...
2018.TE.75...
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/2018.TE.7587.1.13.f...
2018.TE.75...
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/2018.TE.7587.1.14.f...
2018.TE.75...
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/2018.TE.7587.1.15.fa...
2018.TE.75...
https://bioinfoweb.izs.it/bioinfonas/download/20200115_133109/2018.TE.7587.1.16.f...
2018.TE.75...
Filters
All files
Archives (zip, rar, 7z, ...)
Audio (mp3, flac, wav, ...)
Software (exe, msi, ...)
Documents (pdf, odf, docx...
Images (jpeg, png, gif, ...)
JPEG Images
Videos (mp4, webm, mkv, ...
Fast Filtering
.fasta
Filtro
Disable others
Use Once
Subfolder:
Place files in this subfolder within your downloads directory
Use Once
Mask
name.*ext*
Use Once
Download
Add paused
Cancel
10 items selected...

8. Upload

8.1 Introduzione

La piattaforma mette a disposizione due funzioni principali per caricare le sequenze:

- una funzione di **upload** che consente di caricare **proprie sequenze**;
- ed una funzione che consente di importare sequenze direttamente da **NCBI**.

Per quanto riguarda il caricamento di proprie sequenze, viene richiesta la compilazione e validazione di un file di **metadati** insieme al caricamento di un file compresso in formato `.zip`, contenente `.fasta` o `.fastq` a cui fanno riferimento i metadati validati. Per il caricamento da NCBI è sufficiente essere a conoscenza degli **Accession numbers**.

8.1.1 Argomenti

8.2 Upload

La procedura di upload permette di caricare in piattaforma delle sequenze dal proprio computer. Essa consiste di tre fasi:

1. **preparazione e caricamento dei metadati**;
2. **caricamento delle sequenze**.
3. **processamento dei file**.

Le prime due fasi richiedono delle azioni da parte dell'utente, mentre la terza avviene automaticamente al completamento del caricamento dei file.

Tramite la pagina di **Upload** (`Uploads > Upload` del menù di navigazione laterale), è possibile dunque compilare e validare **metadati** e caricare file `.fasta` o `.fastq`, compressi in un unico file in formato `.zip` .

8.2.1 Preparazione e caricamento dei metadati

Nella pagina di **Upload** (`Uploads > Upload` del menù di navigazione laterale), è possibile aggiungere delle schede contenenti i metadati dei campioni, tramite due diversi metodi:

- Aggiungi da form
- Aggiungi da file

I metadati caricati verranno validati dal sistema, che effettuerà anche un controllo sui metadati minimi obbligatori: lo stato di validazione sarà visibile nella colonna omonima.

Aggiungi metadati da form

Il metodo **Aggiungi da form** consente di creare manualmente ed in modo rapido una scheda da associare ad un singolo campione. La scheda creata dovrà essere riempita dall'utente con i metadati del campione a cui fa riferimento, inserendo almeno i metadati obbligatori.

> **Tip!** Consulta il **glossario dei metadati** per capire i tipi di metadati richiesti e relative descrizioni. Troverai anche una guida per la compilazione del modello.

Il video seguente dà una rapida dimostrazione di come compilare una scheda con i metadati di un singolo campione, insieme ad alcune utili funzionalità della gestione dei metadati tramite scheda.

upload-add-from-form.mp4

> **Tip!** Le schede dei metadati possono essere modificate da interfaccia, a prescindere dal metodo originariamente usato per caricare i metadati sulla piattaforma.

Aggiunta punto prelievo

Per quanto rigurda il campo `sampling_point`, se si hanno i permessi necessari è disponibile una funzionalità che permette di aggiungere un punto prelievo da sorgenti esterne, quando non presente nella piattaforma. Per maggiori dettagli è possibile visitare la sezione dedicata di questo wiki: Punto Prelievo.

upload-add-sampling-point.mp4

Aggiungi metadati da file

Il metodo **Aggiungi da file** consente di caricare un file tabulare, contenente la lista di campioni di cui eseguire l'upload, insieme ai rispettivi metadati.

Il video seguente mostra come caricare un file di metadati per uno o più campioni.

upload-add-from-file.mp4

Modelli di files dei metadati

Qui sotto sono disponibili i modelli dei files dei metadati da scaricare:

- template.tsv
- template.xls

Una descrizione delle colonne dei metadati è disponibile nel glossario dei metadati.

> **Tip!** Se non conosci i codici identificativi dei metadati, puoi caricare un record manualmente per poi fare un export del template con le funzioni presenti nel menu contesutale ⋮ della tabella.

upload-template-tip.mp4

8.2.2 Caricamento delle sequenze

Per poter accedere al form di caricamento dei file è necessario che siano presenti dei **metadati validi** corrispondenti ai files che si intende caricare sulla piattaforma. Se tale condizione è rispettata, l'utente potrà **procedere con l'upload**.

Sarà sufficiente selezionare un file in formato `.zip` contenente le sequenze da caricare (tramite funzione drag-and-drop o cliccando nell'apposita area).

> Puoi trovare un approfondimento sulle regole di preparazione dei file di sequenze da caricare sul paragrafo files delle sequenze della sezione regole metadati e zip.

In aggiunta, l'utente può specificare se eseguire automaticamente la pipeline corrispondente alla tipologia di campione tramite il toggle **Processa upload**.

Il video seguente dimostra come usare le funzionalità di upload.

upload-files.mp4

Importante! Il caricamento dei file può richiedere del tempo. Chiudere la finestra o lasciare la pagina durante il processo di caricamento interrompe l'upload.

Sequenze accettate

Le sequenze accettate ricadono in una delle 4 seguenti tipologie:

- **Fasta**, per l'upload diretto di sequenze in formato **fasta**.
- **Raw reads Illumina**, per l'upload di raw reads **fastq** da apparati Illumina, di tipologia **short paired-end** reads.
- **Raw reads Iontorrent**, per l'upload di raw reads **fastq** da apparati Iontorrent, di tipologia **long single-end** read.
- **Raw reads Nanopore**, per l'upload di raw reads **fastq** da apparati Nanopore, di tipologia **long single-end** read.

La tipologia di file e la tecnologia o apparato di sequenziamento utilizzati per produrlo dovranno essere specificati nelle apposite colonne del file dei metadati, come descritto nel **glossario dei metadati**.

8.2.3 Processamento dei file

Al termine del caricamento dei files, il processo di upload non è ancora concluso ed è possibile controllare il suo stato in **Controllo upload**.

Quando tutti i processi relativi al caricamento dei files saranno stati completati sarai avvisato tramite una notifica.

8.3 Upload da NCBI

La pagina **Upload da NCBI** permette di caricare in piattaforma una o più sequenze presenti nei databases NCBI (National Center for Biotechnology Information). I databases da cui è possibile attingere sono rappresentati da 3 diverse tipologie di upload:

- **Assembly fasta**, per l'upload di files fasta dal database "Assembly" di NCBI;
- **Nucleotide fasta**, per l'upload di files fasta dal database "Nucleotide" di NCBI;
- **Reads fastq SRA**, per l'upload di files fastq dal database "Sequence Read Archive - SRA" di NCBI.

Per poter lanciare un upload da NCBI viene richiesta solo una lista di **Accession Numbers**: potrai infatti scegliere da quale dei 3 databases importare il file o i files, selezionare la specie di appartenenza ed incollare nell'apposito campo gli Accession Numbers.

Le sequenze verranno recuperate direttamente da NCBI, senza necessità di caricare files dal proprio computer.

> **Importante!** Possono essere importate più sequenze solo se appartenenti alla stessa specie. Per sequenze appartenenti a specie diverse sarà necessario effettuare upload separati.

Il video seguente mostra come usare le funzionalità di Upload da NCBI. Le 3 tipologie di upload condividono le stesse modalità e gli stessi requisiti.

upload-from-ncbi-procedure.mp4

8.4 Upload PT Genpat

8.4.1 Istruzioni per l'upload

Per effettuare l'upload delle sequenze del **PT Genpat 2025**, la procedura base è la stessa dell'upload su GenPat, come illustrato dal video in basso.

Le uniche differenze fondamentali rispetto ad un normale upload sono le seguenti:

- È necessario entrare con profilo **LAB Campylobacter**;
- È necessario fornire i metadati specifici per il PT indicati nella tabella dei metadati richiesti;
- È necessario fornire un file zip contenente le reads **grezze** da caricare. Queste dovranno essere rinominate come: "Codice + `_R1.fastq.gz` " (nel caso di illumina paired è necessario fornire entrambi i files R1 e R2; entrambi devono essere rinominati). Esempio: `LABXX-GENPAT25-YY_R1.fastq.gz` e `LABXX-GENPAT25-YY_R2.fastq.gz` . Si vedano le indicazioni nella **tabella dei metadati richiesti**.

Genpat-upload_PT.mp4

Si fa notare che:

1. **le reads dovranno necessariamente essere compresse** in formato `gz` ed avere quindi **estensione** `.fastq.gz` .
2. **la piattaforma non opera alcuna compressione:** come specificato in questa e nelle altre pagine della guida sull'upload, il file da caricare per effettuare l'upload deve essere **un file .zip, già compresso autonomamente dall'utente sul proprio computer**, usando uno strumento di compressione a sua scelta.
3. **l'upload per il PT2025 può essere effettuato solo da utenti con profilo "LAB Campylobacter" che ne abbiano fatto richiesta.**

8.4.2 File dei metadati

Qui sotto sono disponibili i modelli dei files dei metadati da scaricare:

- template.xls
- template.tsv

8.4.3 Descrizione dei metadati richiesti

Nella seguente tabella vengono elencate le colonne presenti da riempire nel file dei metadati, spiegandone il significato ed il dato atteso, per ogni campione sottomesso.

Metadato	Dato atteso	Spiegazione
code	Codice nel formato `LABXX-GENPAT25-YY`	il codice indica l'identificativo del laboratorio (`LABXX` , dove `XX` è un codice numerico progressivo a 2 cifre), il PT (Genpat 2025), il numero del campione (`YY` , dove `YY` è un codice numerico progressivo a 2 cifre, corrispondente al codice dell'aliquota di DNA ricevuta)
species	uno dei due codici `110341G` o `110343G`	il campo è dedicato all'indicazione della specie dei campioni, che va inserita come codice corrispondente, ovvero `110341G` per *Campylobacter jejuni* e `110343G` per *Campylobacter coli*
note	stringa di testo	il campo serve ad accogliere una qualunque nota testuale riguardo il campione e il suo riempimento è opzionale
metadata_type	stringa di testo	il campo deve contenere la stringa `pt2025`
file_format	stringa di testo	il campo deve contenere una delle 2 stringhe di testo `fasta` o `fastq` , per indicare il formato del file caricato
seq_type	stringa di testo	se il campo `file_format` è valorizzato con `fastq` questo campo è obbligatorio e deve contenere uno dei seguenti valori in base al sequenziatore utilizzato: `ion` (Iontorrent), `nanopore` , `illumina_paired` (Illumina short reads, paired-end technology)

8.4.4 Note Aggiuntive

Per approfondimenti sul procedimento di upload manuale si rimanda alla pagina Wiki di GenPat: Upload manuale.

Per approfondimenti sulla costruzione del file dei metadati si rimanda alle pagine Wiki di GenPat sui metadati: preparazione e caricamento dei metadati e Regole metadati e zip.

8.5 Controllo upload

La pagina "**Controllo upload**" (`Uploads > Controllo upload` nel menù di navigazione laterale) contiene l'elenco degli upload effettuati dall'utente, organizzati in una tabella.

Gli elementi del **Controllo upload** possiedono le stesse caratteristiche delle altre tabelle della piattaforma.

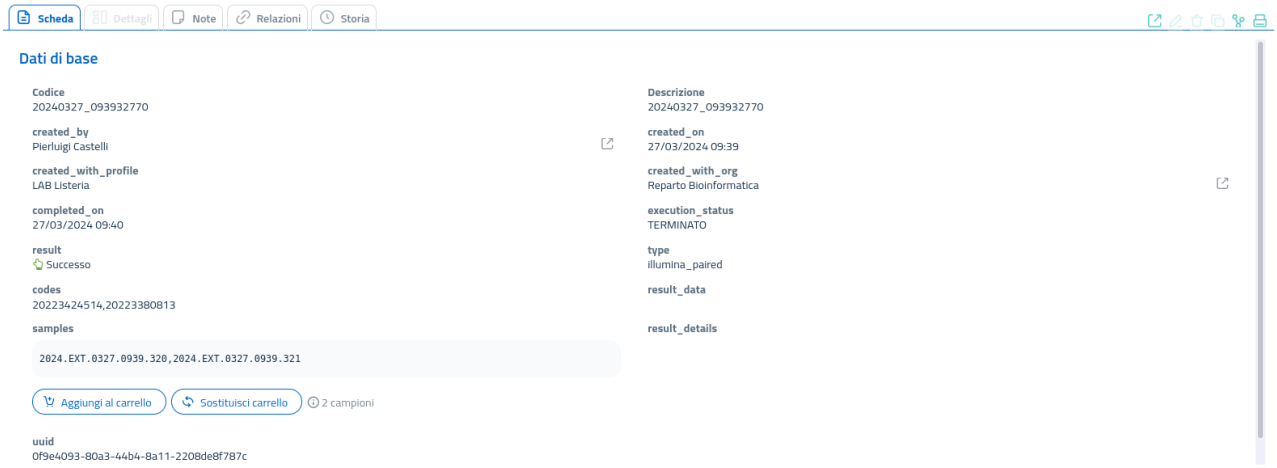
Un upload può avere 3 potenziali esiti:

- **Successo**,
- **Fallimento**,
- **Avvertenze**

8.5.1 Successo

Un upload viene effettuato con **successo** quando tutti i files caricati corrispondono correttamente ai rispettivi metadati. In questo caso, le schede dei metadati vengono rimosse dalla pagina di **Upload** e associate ai files caricati.

Le schede riassuntive degli uploads eseguiti con successo riportano gli identificativi delle sequenze generati dalla piattaforma (campo "samples") e quelli che i files possedevano all'atto dell'upload (campo "codes"). Tali codici possono essere aggiunti al carrello direttamente dalla scheda, tramite l'apposito pulsante "Aggiungi al carrello".



8.5.2 Fallimento

Un upload può risultare in un **fallimento** quando tutti i files che l'utente sta cercando di caricare causano un errore. Nel caso di fallimento, i metadati in attesa di upload rimangono disponibili e visibili nella pagina di **Upload**, in modo da poter essere eventualmente corretti (come dimostrato in questo video).

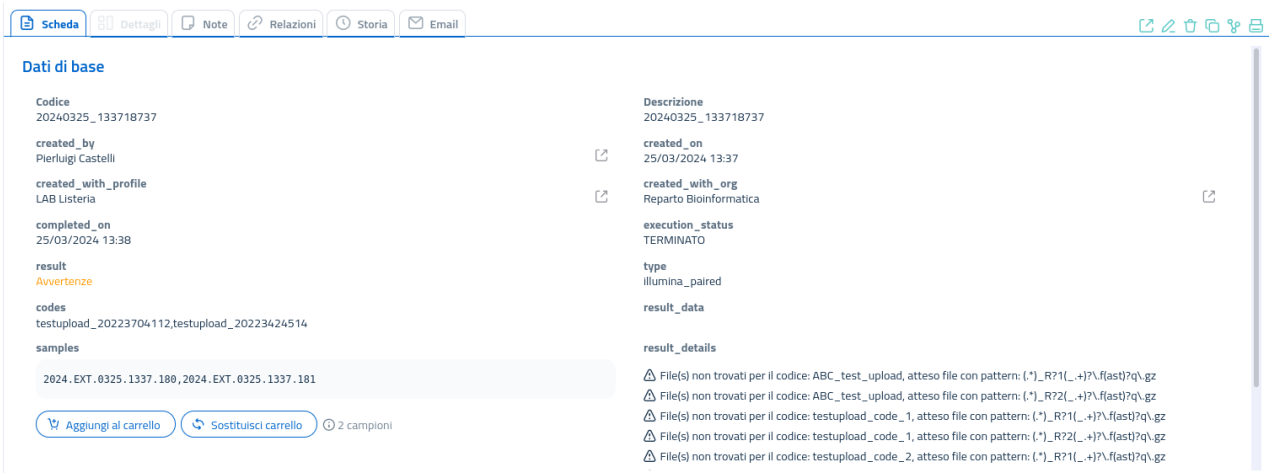
La scheda di un upload fallito fornisce informazioni riguardo il/i motivo/i del fallimento, raggruppate nel campo 'result_details'. Nell'immagine in basso, il processo è fallito perchè nel file zip erano presenti solo files con nomi errati o non corrispondenti a quelli forniti nei metadati:

8.6 Avvertenze

L'upload produce un fallimento solo quando tutti i files in caricamento generano errori: se anche solo uno dei files forniti risulta valido, quel file verrà caricato e la scheda riporterà delle **avvertenze**.

Nelle schede degli upload con avvertenze saranno presenti sia i codici dei campioni caricati con successo, da poter aggiungere al carrello, sia i resoconti degli errori (nel campo "result_details"), per i campioni che invece non sono stati caricati. Nella pagina di **Upload**, solo i metadati dei campioni caricati con successo vengono associati e rimossi dalla lista, gli altri rimangono in coda, in attesa di correzione o rimozione.

Nel caso dell'immagine in basso, alcuni files nello zip fornito possedevano nomi non validi; due files risultavano però corretti. I due files validi sono stati caricati, mentre gli altri hanno dato origine alle avvertenze.



8.7 Approfondimenti

8.7.1 Glossario dei metadati

Nella tabella in basso sono riportate le colonne del modello dei metadati. Per ogni colonna del template scaricabile viene elencato il tipo di metadato richiesto e ne viene data una descrizione e una guida per la compilazione del modello.

Per una spiegazione sui limiti e le richieste riguardanti il file dei metadati e dei file da caricare, si faccia riferimento alla pagina Regole metadati e zip.

Nota: ogni profilo può effettuare upload di tipi specifici tra `official` e `research` (si veda la riga `metadata_type` nella tabella in basso). Il sistema non accetterà upload di tipi diversi da quelli consentiti a quel profilo.

Titolo campo TSV	Metadato	Descrizione	Obbligatorio
`code`	Codice campione	Codice identificativo campione nel sistema del data provider	si
`species`	Specie patogeno	Indicare la specie del patogeno con codifica GENPAT	si
`material`	Materiale	Indicare il materiale del campione con codifica GENPAT	si
`sampling_point`	Luogo Prelievo	Indicare un codice di Luogo Prelievo presente in GENPAT. In caso di campioni ufficiali è possibile inviare il codice istat del comune	si
`sampling_point_address`	Luogo Prelievo	Indirizzo. Da usare in caso sia stat modalita_invio=semplificata	no
`sampling_point_note`	Luogo Prelievo	Utilizzare per indicare il tipo di struttura. In caso modalita_invio=semplificata	no
`sampling_date`	Data prelievo	Utilizzare dettaglio della data al giorno	si
`matrix`	Matrice di isolamento	Utilizzare codice EFSA FOODEX2 della matrice di isolamento. E' necessario specificare almeno il codice FOODEX2 base per alimento, mangimi, animale e ambiente. In particolare: **Food (A0B6Z)**, **Feed (A0BB9)**, **Animals (A04SF)**, **Environment (A0C5Y)**	si
`host`	Specie host	Utilizzare codice EFSA FOODEX2. Specie dell'animale ospite. Da utilizzare in caso di matrice animale	no
`note`	Note aggiuntive	Testo libero opzionale. Utile a descrivere situazioni di anomalia o deviazioni	no
`md5_file_r1`	file rawreads	MD5 checksum del file R1 delle raw reads o del singolo file in caso di sequenziamento single-end	Obbligatorio in caso di `read_format` =fasta e tipo di invio = Ufficiale
`md5_file_r2`	file rawreads	MD5 checksum del file R2 delle raw reads	Obbligatorio in caso di sequenziamento pair-end e di `read_format` =fasta e tipo di invio = Ufficiale
`filename1`	file rawreads	Nome file fasta in caso `read_format` =fasta, oppure nome file R1 delle raw reads o del singolo file in caso di sequenziamento single-end. In caso il campo `code` rappresenti un prefisso del nome file, questo campo può essere evitato. Esempi di nomi dei files corrispondenti ai diversi pattern con `code` ="code": `read_format` =illumina_paired: "code*R1*.fastq.gz" `read_format` =fastq_singleend: "code.*fastq.gz*' `read_format` =*fasta: 'code*.fasta"	no
`filename2`	file rawreads	Nome file R2 in caso di sequenziamento pair-end. In caso il campo `code` rappresenti un prefisso del nome file, questo campo può essere evitato. Esempi di nomi dei files corrispondenti ai diversi pattern con `code` ="code": `read_format` =illumina_paired: "code*R2*.fastq.gz"	no

Titolo campo TSV	Metadato	Descrizione	Obbligatorio
`file_format`	Tipo di file inviato	Può assumere uno dei valori: `fasta`, `fastq`	si
`seq_type`	Tecnologia di sequenziamento	Può assumere uno dei valori: "illumina_paired", "ion", "nanopore". Obbligatorio per tipo di invio = Ufficiale	conditional
`metadata_type`	Tipo di metadato inviato. Vengono eseguite validazioni diverse sui dati in base a questo campo	Può assumere il valore: "official" o "research"	si

8.7.2 Regole metadati e zip

In questa sezione della guida vengono descritte le caratteristiche che devono avere i files e i metadati usati per l'upload, assieme alle regole che essi devono rispettare perchè la loro validazione venga eseguita con successo.

Metadati

> Si invita a consultare il glossario dei metadati per una trattazione specifica sui singoli campi presenti nei modelli dei metadati.

Metadati aggiunti da form

Se i metadati vengono aggiunti manualmente tramite l'opzione Aggiungi da form, o se le schede di uno o più campioni vengono modificate manualmente, l'utente viene guidato nella selezione dei valori dei metadati. Ci sono quindi solo 2 regole da seguire, per garantire una corretta validazione della scheda dei metadati:

- Usare un **Codice** (nome del file) corretto;
- Compilare almeno i **campi obbligatori** per i metadati.

Usare un Codice

Il testo da inserire nel campo "Codice" è il nome del file fasta o fastq che si intende caricare.

- Il codice è costituito dal nome del file, **senza estensione**, che si trova nell'archivio zip che viene caricato;
- Il codice e il nome del file non devono contenere spazi;
- Nel caso di files fastq di *reads paired-end*, il codice **non** deve contenere l'identificatore della *read forward* o *reverse* (ad esempio "_R1" o "_1") e deve essere **unico** (la scheda dei metadati vale infatti per entrambe le *reads*).

> **Nota** Il modello dei metadati offre anche la possibilità di specificare i nomi completi dei files (entrambe le reads per i files paired-end, unico per reads single-end o fasta). Nel caso tali colonne siano riempite, il controllo verrà effettuato sui nomi completi dei files, piuttosto che sul solo codice.

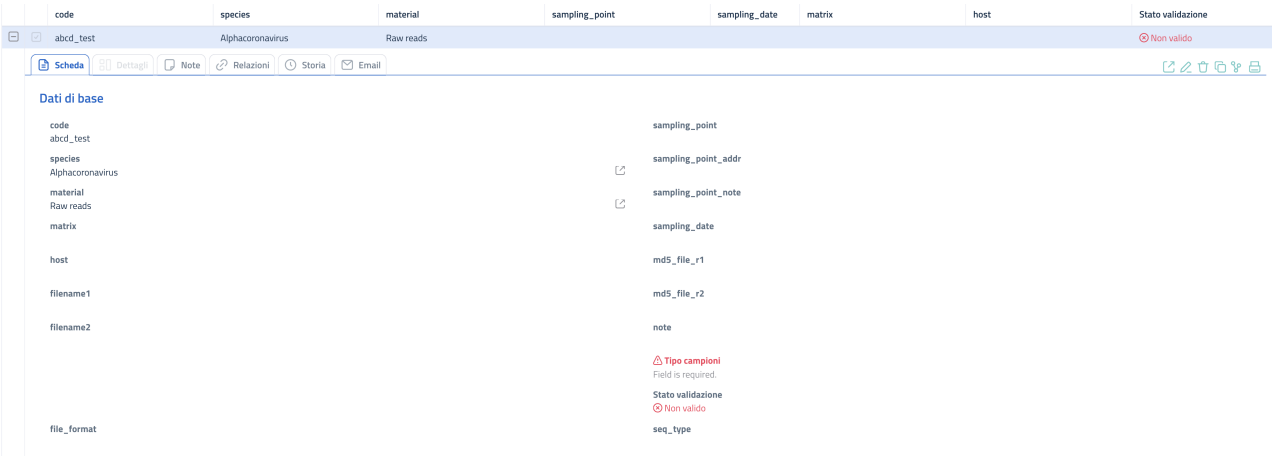
Compilare i campi obbligatori

I campi obbligatori possono variare in base alla tipologia di campione che viene caricato; ad esempio, un campione di ricerca avrà meno campi obbligatori rispetto ad un campione ufficiale, per il quale è invece necessario fornire un maggior numero di metadati.

In generale, i metadati minimi obbligatori possono richiedere di:

- riempire il campo `code`;
- specificare la tipologia di campione (ufficiale, ricerca, *etc.*), il tipo di file (fasta o fastq) e, in caso di fastq, la tecnologia dell'apparato per il sequenziamento (Illumina *paired-end*, Nanopore *single-end*, *etc.*);
- indicare la specie o il materiale.

In fase di validazione vengono evidenziati con dei messaggi di errore i campi obligatori con valori assenti o errati.



TIP Inserendo i metadati da form o utilizzando le funzioni di edit della piattaforma, campi come specie o materiale, possono essere valorizzati tramite un menu a tendina o una specifica tabella. In questi casi la scelta è guidata e garantisce l'inserimento del codice corretto.

Metadati aggiunti tramite file

I templati dei files tabulari per la sottomissione dei metadati sono scaricabili direttamente dalla sezione **Modelli di files dei metadati** della pagina di questo wiki dedicata all'Upload. I modelli sono disponibili sia in formato `.tsv` che in formato `.xls`.

TIP I template sono scaricabili anche dalla piattaforma sia tramite la modale di caricamento dei metadati (bottone `Aggiungi da file`) che da menu contestuale della pagina di upload.

upload-metadata-template.mp4

Metadati richiesti

Ogni colonna del file dei metadati corrisponde ad un campo della scheda metadati.

Nel template del file dei metadati, la seconda riga funge da esempio per quanto riguarda la formattazione e il tipo di testo da inserire in ogni cella.

Valgono le seguenti indicazioni per i campi dei metadati richiesti dal template:

- Il campo `metadata_type` può assumere i seguenti valori:
  `official` - (upload di campioni ufficiali);
  `research` - (campioni caricati a scopo di ricerca).

Ogni profilo utente può effettuare upload di tipi specifici tra quelli sopra elencati e **il sistema non accetterà upload di tipi diversi da quelli consentiti**. Alcuni metadati, inoltre, sono obbligatori solo se l'upload è di campioni ufficiali (`metadata_type` = `official`).

- I campi di `species` , `material` , `host` , `matrix` e `sampling_point` **non** vanno riempiti con nomi tassonomici o descrizioni testuali, ma con **il codice corrispondente nelle tabelle della piattaforma**.

TIP Le tabelle dei codici dei metadati sono consultabili in `Elementi principali > Metadati` del menù laterale della piattaforma. Ulteriori informazioni sono disponibili nella sezione **Metadati** di questa Wiki.

- Il campo `read_format` può assumere i valori:
- `fasta` ;
- `illumina_paired` (files fastq paired-end da apparati Illumina);
- `ion` (file fastq single-end da apparato Iontorrent);
- `nanopore` (file fastq single-end da apparato Nanopore).

Per tutte le altre indicazioni e restrizioni, valgono le informazioni fornite nella sezione Metadati aggiunti da form - Campi obbligatori e nel glossario dei metadati.

Files delle sequenze

I files in formato **fasta** vengono riconosciuti se possiedono estensione `.fasta` o `.fa` . Analogamente, i files in formato **fastq** vengono riconosciuti se possiedono estensione `.fastq.gz` o `.fq.gz` .

Nota: La compressione dei files fastq in formato `.gz` è obbligatoria. I files fasta vengono invece accettati con e senza estensione `.gz` .

Archivio zip

L'archivio deve essere in formato `.zip` . In fase di caricamento, il sistema opererà un controllo delle schede dei metadati, che vengono associati ai campioni corrispondenti. Affinchè la procedura di upload vada a buon fine, è necessario che il nome di ogni file fasta o fastq (esclusa l'estensione del file) corrisponda al valore del metadato `code` inseriso durante il caricamento dei metadati.

Se almeno uno o solo alcuni dei campioni nell'archivio zip presentano corrispondenze con i codici nelle schede validate, il processo di upload verrà portato a termine solo per quei file e con delle **avvertenze**.

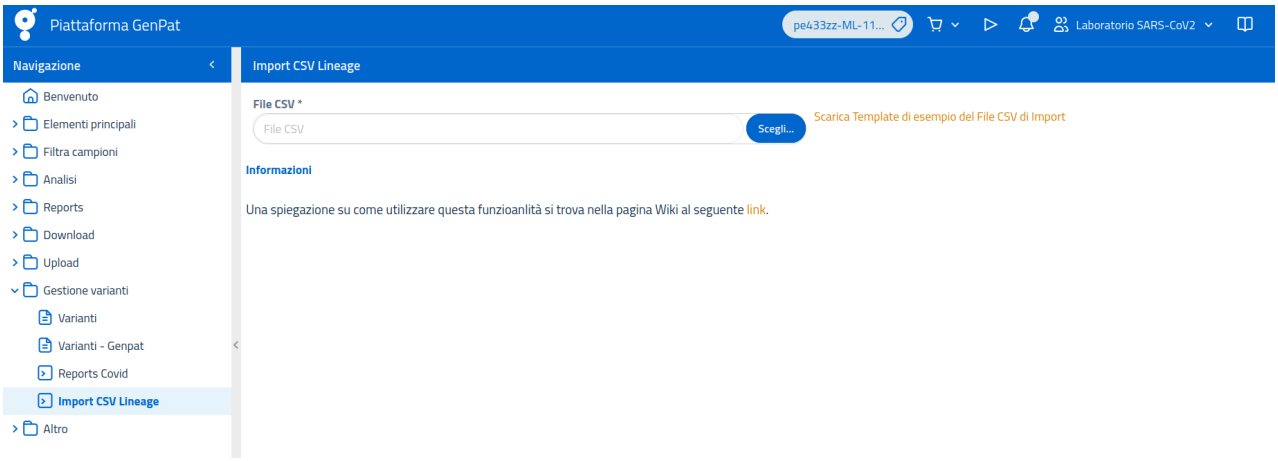
Nel caso in cui l'archivio zip contenga dei files i cui nomi non corrispondono a nessuno dei codici insriti nelle schede dei metadati validate, la richiesta di upload produrrà un **fallimento**.

TIP I files nell'archivio possono trovarsi anche all'interno di un sistema di cartelle e sotto-cartelle, purché la struttura non superi i 10 livelli di profondità.

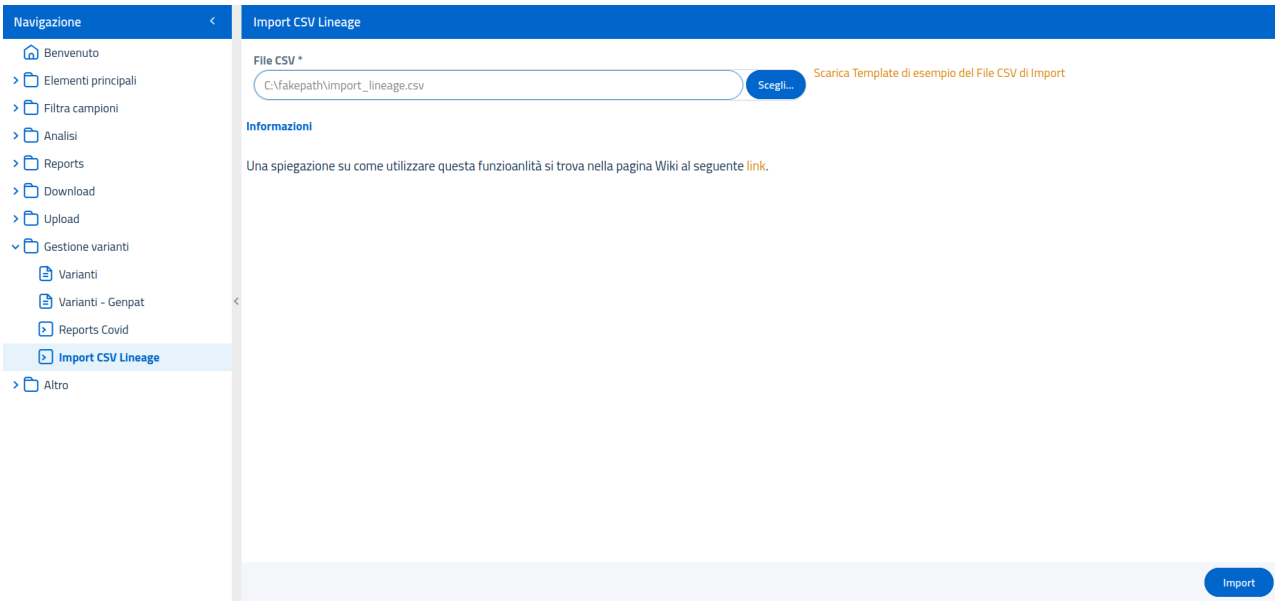
9. Gestione Varianti

9.1 Import CSV Lineage

In piattaforma è presente un'interfaccia che permette di effettuare l'upload dei file di lineage. La pagina, denominata "**Import CSV Lineage**", è raggiungibile tramite la voce di menù **Gestione Varianti**.



Per poter effettuare l'upload del file di lineage è necessario selezionare il file in formato CSV nell'apposito campo e successivamente cliccare sul tasto blu "Import", nell'angolo in basso a destra.



I dati del lineage verranno associati ai campioni automaticamente e verranno anche inseriti i nuovi campioni GisAID.

File CSV di esempio per l'import: import_lineage.csv

10. Dashboards

10.1 SPREAD, Spatiotemporal Pathogen Relationships and Epidemiological Analysis Dashboard (Relazioni spazio-temporali dei patogeni e dashboard delle analisi epidemiologiche)

Prima conosciuto come GrapeTree Extended.

10.1.1 Descrizione

La dashboard si basa su GrapeTree, un programma di visualizzazione di grafici ad albero da file Newick (nwk), totalmente interattivo e appartenente al sistema EnteroBase. GrapeTree supporta la manipolazione sia del layout e sia dei metadati dell'albero. Le figure generate sono prodotte usando gli algoritmi Neighbor-Joining (NJ) e Minimal Spanning Tree (MSTree) o un algoritmo Minimal Spanning Tree migliorato, chiamato MSTree V2. La versione originale è parte integrante di EnteroBase e, per maggiori dettagli, puoi consultare la relativa documentazione.

Per una descrizione formale, puoi consultare la pubblicazione validata in Genome Research.

La versione standalone di GrapeTree è stata integrata, in piattaforma, come strumento principale per la manipolazione dei grafici ad albero e dei metadati; essa, inoltre, al fine di consentire analisi spazio-temporali, è stata ampliata attraverso un sistema integrato di informazioni geografiche (GIS) ed un sistema di visualizzazione basato sui dati temporali. L'applicazione web consente agli utenti di caricare le coordinate geografiche e i dati temporali relativi a ciascun campione e di visualizzarli in maniera tale che la selezione dell'albero si rifletta sulla mappa e viceversa. L'applicazione, inoltre, permette all'utente di riprodurre una visualizzazione temporale sia sulla mappa sia sull'albero.

10.1.2 Come si usa

La dashboard è integrata con la piattaforma; ciò rende possibile lanciare un'analisi e visualizzarne i dati. Accedendo, tuttavia, direttamemte all'indirizzo https://cohesive.izs.it/spread/, puoi caricare e visualizzare un tuo dataset.

> **Nota bene** Il dataset caricato viene visualizzato e gestito esclusivamente lato client nel browser; ciò significa che nessun dato viene trasmesso e non vengono utilizzati cookie di tracciamento. Gli unici dati scaricati da Internet sono il codice di visualizzazione (JavaScript), i font e le eventuali parti necessarie alla creazione della mappa; in ogni caso, puoi anche caricare dati sensibili.

Puoi caricare i file trascinandoli direttamente sopra l'area di rilascio iniziale o utilizzando il bottone di caricamento. Puoi trascinare o caricare un file `.nwk`, seguito da un file `.tsv` contenente i metadati e, facoltativamente, un file `.geojson` contenente informazioni geospaziali relative ai campioni.

Importante!
Il file `.nwk` deve essere caricato sempre prima dei metadati o dei file geoJson.

La dashboard ti consente di scaricare un file JSON completo che include metadati e configurazioni. Il file JSON generato può essere caricato usando il trascina e rilascia oppure le funzionalità di caricamento menzionate in precedenza. Questa funzione risulta utile nel caso tu voglia salvare il tuo lavoro e/o condividerlo.

Upload dell'albero
Puoi trovare un dataset di esempio sulla repository del progetto in https://github.com/genpat-it/spread/tree/main/datasets/test.

10.1.3 Presentazione delle funzioni generali

Sia che tu scelga di caricare dati dal tuo computer o che tu stia visualizzando un dataset pre-elaborato dalla piattaforma, ti troverai di fronte sempre alla stessa area di lavoro configurabile. I principali componenti della dashboard, oltre all'**Albero**, sono:

- **Impostazioni**
- **Mappa**
- **Metadati**
- **Legenda**
- **Video**

Questi componenti possono essere mostrati o nascosti tramite i relativi pulsanti posti nella barra sopra l'area di lavoro principale, in base alle tue esigenze. Prima di entrare nel dettaglio di questi componenti, è opportuno menzionare rapidamente due funzioni presenti sulla barra: la prima è relativa al *cambio lingua* che puoi impostare clccando sull'icona del **mappamondo**; la seconda mostra *informazioni e link utili* e può essere impostata cliccando sull'icona **info**.

Lingue disponibili e traduzioni
Al momento, le lingue disponibili sono l'inglese, il francese e l'italiano; l'applicativo è, tuttavia, facilmente estensibile con nuove traduzioni. Se vuoi contribuire, puoi prendere a modello il file `en.js` o `it.js` presente nella cartella `i18n` sulla repository pubblica del progetto e tradurre i valori presenti. Per qualsiasi ulteriore informazione o supporto in merito puoi scrivere a `bionformatica@izs.it`.

10.1.4 Componenti

Vediamo nel dettaglio le varie parti che compongono la dashboard.

Strumenti albero

Dal menù contestuale, posto in alto a destra dell'area di lavoro dell'albero, si accede a questa sezione che contiene sia strumenti utili per modificare l'aspetto e il layout dell'albero, sia un'opzione per poter riportare l'albero allo stato originario, annullando eventuali cambiamenti e modifiche. Tra le funzioni disponibili, in questa sezione troviamo:

- **Centra albero** che regola le impostazioni di visualizzazione per portare il centro del grafico in posizione centrale all'interno dell'area di lavoro;
- **Ridisegno statico** che ricarica il grafico e lo ridisegna nel layout predefinito;
- **Albero iniziale** che ricarica la pagina e riporta l'albero allo stato e con il layout in cui si trovava quando è stato caricato la prima volta. Ti verrà chiesto di confermare prima di abbandonare la pagina. **Attenzione**: l'uso di questa funzione provocherà la perdita di tutte le modifiche effettuate.

Impostazioni

Attivando le impostazioni, entrerai in una `sezione modificabile` con una serie di `schede`, suddivise in aree tematiche, che ti permettono di configurare l'albero. Ogni scheda può essere espansa o ridotta in base alle tue esigenze. Troverai descritte, qui di seguito, le varie configurazioni disponibili, visibili anche all'interno dell'applicazione cliccando sull'icona `?` della relativa scheda.

Stile nodi

In questa categoria, è possibile trovare alcune opzioni per modificare l'aspetto dei nodi o delle relative informazioni visualizzate.

- **Colora per**: nel menù a cascata di questa sezione, si può selezionare il colore di riempimento dei nodi, in base ai metadati;
- **Mostra/Nascondi etichette**: l'opzione permette di scegliere se visualizzare o meno le etichette dei nodi. È anche possibile scegliere che tipo di etichetta mostrare dal menù a cascata previsto nel campo *Etichette*;
- **Dimensione carattere**: le dimensioni del font delle etichette sono modificabili trascinando l'apposito cursore o specificando le dimensioni desiderate nel campo sottostante;
- **Evidenzia etichette**: usa il campo di ricerca per evidenziare le etichette dei nodi. Sono supportate le Espressioni Regolari;
- **Segmenti individuali**: l'interruttore di questa opzione permette di trasformare i nodi in grafici a torta, ognuno dei quali presenterà una scomposizione degli elementi al suo interno. E' importante sottolineare evidenziare che l'attivazione della funzionalità di scomposizione dei nodi dipenderà dalla categoria selezionata nel menu *Colora per*.

Dimensione nodi

- **Dimensione raggio (%):** la dimensione dei nodi può essere aumentata o diminuita trascinando l'apposito cursore o specificando le dimensioni nel campo.
- **Curtosi (%):** l'area dei nodi è correlata di default al numero di elementi che ne fanno parte. Usando il cursore o il campo in questa sezione, le dimensioni dei nodi possono essere diminuite o aumentate sulla base della distribuzione della curtosi dei nodi. L'aumento del valore di curtosi comporta l'incremento della percentuale di crescita della dimensione del nodo per ogni membro in esso contenuto; quindi, i nodi con un maggior numero di membri avranno un'area più grande e risalteranno maggiormente.

Stile rami

In questa sezione sono presenti le seguenti opzioni utili a personalizzare i rami dell'albero:

- **Mostra/Nascondi etichette:** l'opzione permette di scegliere se mostrare o nascondere le etichette dei rami;
- **Dimensione carattere:** permette di scegliere le dimensioni del testo delle etichette o tramite l'uso del cursore o inserendo un valore nell'apposito campo;
- **Adattamento in scala (%):** questa opzione permette di aumentare/diminuire la lunghezza dei rami attraverso l'utilizzo del cursore oppure specificando un valore nel campo sottostante;
- **Comprimi i rami:** usando il cursore o scrivendo un valore nel campo di questa opzione, tutti i rami più corti del valore selezionato verranno compressi. I nodi che ricadono in tale intervallo verranno fusi. I valori delle lunghezze dei rami verranno adeguati in scala a quelli definiti nei dati dell'albero originale;
- **Scala logaritmica:** selezionando questa opzione, le lunghezze di tutti i rami verranno mostrate in scala logaritmica.

Taglio rami

Le impostazioni di questa sezione permettono di rappresentare i rami in modi diversi in relazione alla loro lunghezza (i valori di lunghezza dei rami vengono riprodotti in scala rispetto alle lunghezze definite nei dati del grafico originale). Per aumentare o diminuire la scala puoi inserire un valore numerico nel campo *Taglia rami più lunghi di*. In seguito puoi scegliere se mostrarli, nasconderli o ridurli:

- **Mostra:** mostra i rami più lunghi del valore inserito (predefinito);
- **Nascondi:** rende trasparenti i rami più lunghi del valore scelto. Nonostante i rami non siano mostrati, rimane possibile interagirvi;
- **Riduci:** i rami più lunghi del valore di cutoff specificato verranno ridotti al valore di cutoff. I rami ridotti saranno rappresentati con linee tratteggiate per indicare che sono interessati da una riduzione.

Rendering

Questa sezione contiene alcune opzioni che permettono di controllare il posizionamento dei nodi nel grafico. Un pulsante permette di alternare due modalità: *Statico* e *Dinamico*.

- **Statico:** in questa modalità, il layout dell'albero viene calcolato alla generazione del grafico e viene mantenuto statico (ma interattivo). La lunghezza relativa dei rami e la loro scala (sulla base dei dati del grafico originale) saranno sempre mantenute fintanto che l'opzione **Lunghezza del ramo reale** resta selezionata.
- **Dinamico:** i nodi vengono posizionati dinamicamente in maniera simile ad una modalità Force-directed ma tenderanno a sparpagliarsi per distanziarsi dai nodi vicini. In questa modalità, l'albero può essere spostato liberamente, trascinandolo con il mouse; i nodi si ridisporranno sempre automaticamente per mantenere le distanze dai nodi circostanti, a seguito dello spostamento. La modalità *Dinamico* può essere usata per migliorare l'estetica del grafico ma la scala dei rami verrà modificata rispetto all'originale: i rami NON saranno in scala in questa modalità. L'opzione **Solo selezionati** può essere attivata per applicare la scelta solo ai nodi selezionati.

Impostazioni dell'albero

Mappa

Questa sezione permette di visualizzare una mappa interattiva se, per i campioni, sono state definite delle coordinate geografiche (tramite file metadata o tramite file geoJSON). Sulla mappa vengono disegnati dei `marcatori`, con un raggio più o meno lungo, in base al numero di campioni presenti in una determinata area. Cliccando su ogni marcatore, puoi aprire un popup informativo contenente la lista dei campioni presenti all'interno.

I marcatori riflettono i colori del tematismo scelto per la visualizzazione e la selezione dell'albero. Puoi interagire con la mappa anche tramite il suo **menu contestuale**, accessibile con un click del tasto destro del mouse su di essa oppure tramite il pulsante in alto a destra raffigurante tre punti verticali.

Strumenti mappa

Tramite le opzioni presenti in questa sezione potrai controllare l'aspetto della mappa.

- **Aggrega per coordinate/per metadati:** il pulsante ti consente di aggregare i punti sulla mappa in base alla vicinanza geografica delle coordinate fornite o in base ai valori dei metadati caricati. Nel primo caso è possibile **impostare un valore** di *delta* che determina il grado di aggregazione.
- **Raggio minimo/massimo del marcatore:** tramite questi due campi, è possibile intervenire sulla dimensione dei punti sulla mappa definendo un raggio minimo ed un raggio massimo. La dimensione iniziale è relativa alla quantità di nodi presenti sulle stesse coordinate geografiche.
- **Modalità diagramma/Modalità heatmap:** questo pulsante permette di cambiare la modalità di visualizzazione dei punti sulla mappa. Con la prima opzione, (selezionata di default), i punti saranno visualizzati con grafici a torta in base alle categorie dei nodi dell'albero laddove, con la seconda, ai nodi viene applicata una sfocatura.

Mappa dell'albero

Metadati

Attivando il pulsante **Metadati**, potrai visualizzare i metadati all'interno di una tabella interattiva. La tabella permette di spostare le colonne, estenderle e/o ridurle, ordinarle o filtrarle in base a delle chiavi di ricerca. Cliccando su una riga, puoi selezionare/deselezionare i campioni presenti nell'albero.

Come per la mappa, anche la tabella dei metadati ha degli strumenti ai quali si può accedere cliccando sulla tabella con il tasto destro del mouse o, in alternativa, usando il bottone in alto a destra con i tre punti verticali.

Metadati dell'albero

Legenda

Questo è uno strumento chiave sia per l'interazione con l'albero sia per la visualizzazione/selezione temporale dei dati. Tramite gli **Strumenti legenda**, è possibile cambiare categoria e colore dell'albero, personalizzare i colori o selezionare campioni da una determinata categoria.

Il video seguente mostra come accedere al menu contestuale della legenda e alle sue opzioni, come interagire con la legenda e come modificare le preferenze e la palette di colori.

Legenda dell'albero

Video

In modalità **Video**, la legenda si trasforma in una vera e propria timeline. Il video player, posizionato in basso a destra, permette di iterare sugli elementi della legenda ad una velocità configurabile. Le funzioni video sono completamente metadata driven quindi, se all'interno dei metadati sono presenti informazioni temporali, puoi colorare la legenda per tale informazione, ordinarla in ordine ascendente e, di conseguenza, avere una timeline per misurare la diffusione e lo sviluppo nel tempo di un agente patogeno.

Nel video seguente troverai un possibile caso d'uso.

Modalità video della legenda

Macro

Il componente Macro è accessibile tramite l'icona `fast-forward` situata nell'angolo in alto a destra dell'applicazione. Questa sezione offre una raccolta in continua espansione di sequenze di azioni (macro), che possono essere ricercate e categorizzate per tag.

Il componente consente di cercare le macro e, utilizzando chiavi specifiche (`title:` , `tag:` , `desc:`), è possibile restringere la ricerca per concentrarsi sugli elementi di interesse.

Se hai suggerimenti per nuove macro, saremo lieti di riceverli!

Zoom dei cluster

L'applicazione può essere usata per visualizzare risultati generati con ReporTree. Molto brevemente, ReporTree permette la generazione di una serie di zoom sui cluster specificando i campioni di interesse e alcuni parametri quali soglie e distanze; in questo contesto, è possibile indicare alla dashboard la loro presenza e, di conseguenza, essere visualizzati. Sull'interfaccia dell'albero, cliccando sul pulsante **Zoom disponibili** si ha accesso agli **Strumenti zoom** per poter avere un focus sugli eventuali cluster disponibili. È possibile selezionare il *tipo di zoom*, la *soglia* e il *cluster* al quale il campione può essere associato, così come è possibile selezionare i **nodi coinvolti**. Le funzioni permettono, inoltre, di poter **aprire in una nuova scheda** un cluster modificato, potendo decidere se mantenere o meno le **impostazioni correnti**.

Nel video seguente troverai un possibile caso d'uso.

Zoom dei cluster

Albero

Questo è ovviamente il componente principale e, pertanto, è sempre visibile. Esso mostra la visualizzazione ad albero dei dati ed è completamente interattivo. Con questo componente è possibile interagire direttamante tramite il menu contestuale o tramite tutti gli altri componenti.

Puoi interagire direttamente con l'albero, spostando i singoli nodi o posizionandoti sopra di essi per vedere dei tooltip; puoi anche selezionarli premendo `Shift` e cliccando su di essi oppure disegnando un'area con il mouse. La selezione si rifletterà anche sui componenti **Mappa** e **Metadati**.

CONSIGLIO

Per abilitare la selezione nell'area di lavoro del grafico, si tenga premuto il tasto `Shift` della tastiera: fintanto che il cursore viene mostrato con la forma di un mirino, sarà possibile selezionare o deselezionare i singoli nodi cliccando su di essi; sarà altresì possibile selezionare più nodi tracciando un'area di selezione (è necessario trascinare il cursore mentre si tiene premuto il tasto sinistro del mouse).

Attività più interessanti sicuramente possono essere fatte tramite il **menu contestuale** o, come detto, tramite l'interazione con gli altri componenti (trattati nelle relative sezioni). Puoi accedere al **menu contestuale** facendo click con il tasto destro del mouse sull'albero oppure tramite bottone in alto a destra con tre punti verticali. Le voci del menu sono abbastanza auto esplicative e ti invitiamo a provarle per capire meglio, tramite l'interazione, a cosa servono. Di base, vanno a modificare la visualizzazione dell'albero.

Ci concentriamo, qui, sulla funzione di *salvataggio*.

Salva SPREAD

Tramite questo pulsante, hai accesso ad una modale che permette di salvare e/o esportare SPREAD in diversi formati:

- **Salva come json completo (.json)**: permette di salvare il lavoro corrente in un file `.json` completo di tutte le informazioni sull'albero, sui relativi metadati e sulle configurazioni correnti dei diversi componenti. Questa funzionalità è fondamentale poichè rappresenta l'unico modo per salvare i progressi del tuo lavoro di modifica di SPREAD, riprenderlo in un secondo momento o condividerlo con altri. Nel `.json` vengono salvate informazioni quali: il posizionamento e il colore dei nodi, la categoria corrente, la modalità di visualizzazione dei nodi sulla mappa, la configurazione del layout dell'area di lavoro e persino le schede delle impostazioni che sono espanse o compresse.
- **Esporta newick (.nwk)**: esporta i dati nel formato `.nwk`
- **Esporta metadata (.tsv)**: esporta i metadati nel formato `.tsv`
- **Esporta geoJson (.geoJson)**: esporta i metadati spaziali nel formato `.geojson`

Tutti i file prodotti sono compatibili per essere ricaricati successivamente nella dashboard come accennato in Come si usa.

Strumenti albero e funzione Salva SPREAD

10.2 Auspice

La dashboard Auspice, integrata alla piattaforma, permette di visualizzare, tramite il software omonimo, (consulta la guida ufficiale) gli alberi filogenetici, i metadati associati e i dati geografici prodotti dalla pipeline Augur.

L'accesso all'integrazione è possibile dalla sezione "Visualizza analisi" che apre la tabella di tutte le analisi e dei relativi *workflow*. L'accesso alla dashboard Auspice è disponibile nella sezione *Dati risultato* della scheda di ogni analisi **Augur** terminata con successo.

10.2.1 Guida rapida

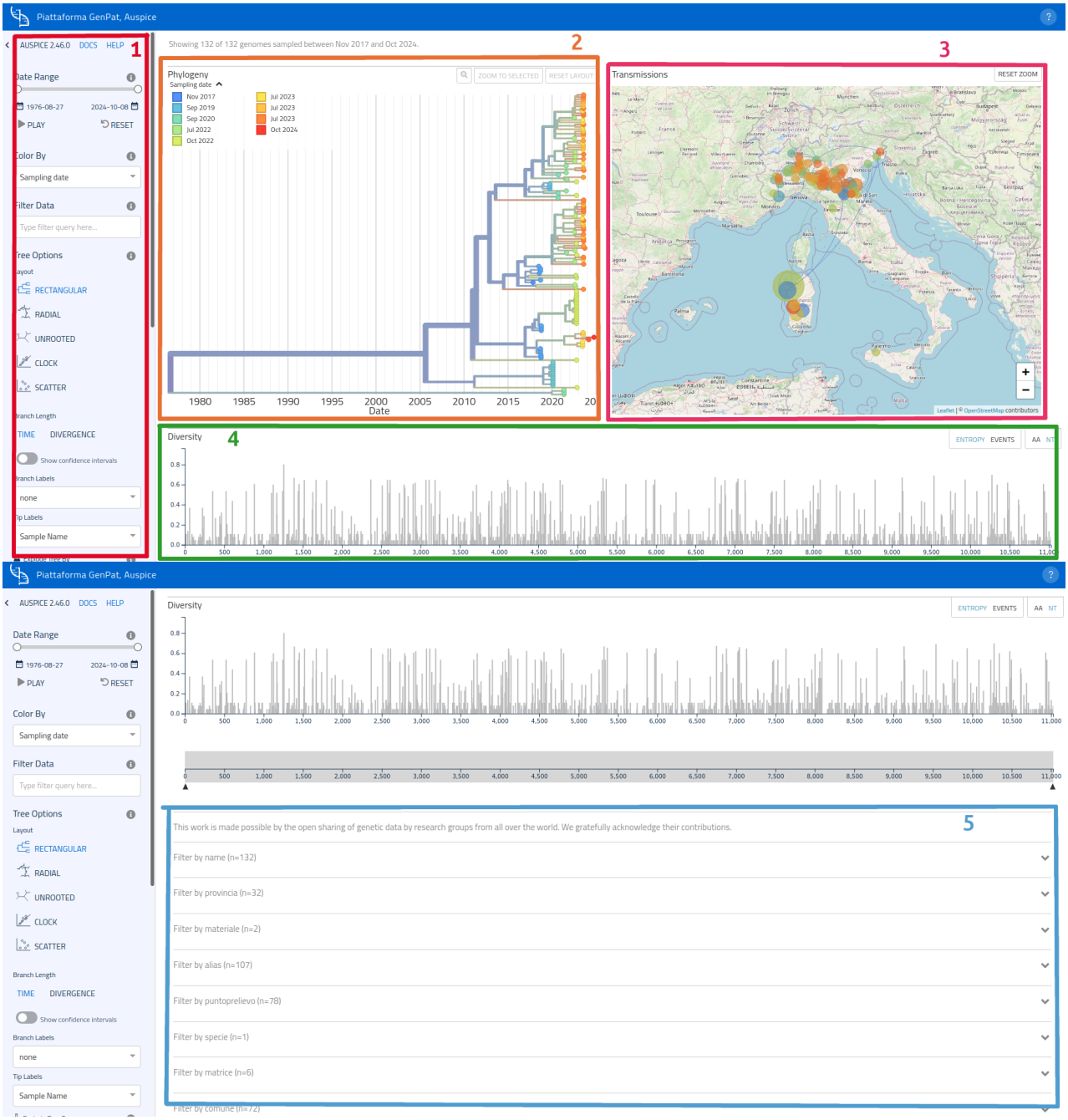
Il seguente video introduttivo costituisce una guida rapida all'interfaccia di Auspice e contiene dimostrazioni pratiche all'uso del menù, della mappa e dei grafici.

Demo Auspice

10.2.2 Interfaccia principale

L'interfaccia di Auspice è divisa in 5 zone principali:

- una barra laterale delle opzioni a sinistra (**1**);
- l'albero filogenetico a sinistra (**2**);
- una mappa a destra, dove vengono visualizzati i *clusters* sulla base dei dati geografici (**3**);
- grafico della diversità (del quale si può scegliere di visualizzare entropia o eventi), al di sotto di mappa e albero (**4**);
- una zona filtri, sotto il grafico (**5**).



Barra delle opzioni

La barra delle opzioni offre accesso alle pagine del manuale ufficiale di Auspice (*Docs*) ed è divisa in sottocategorie, in modo da gestire diverse caratteristiche dell'aspetto dei dati:

- Visualizzazione, colori e filtro dei dati (*Filter data*);
- Opzioni dell'albero;
- Opzioni della mappa;
- Opzioni di animazione;
- Layout e visualizzazione dei pannelli.

La descrizione del menù è disponibile nella guida ufficiale di Auspice; inoltre, ad ogni elemento del menù laterale è associato un bottone "**info**" ( ) che, al passaggio del puntatore, mostrerà una descrizione dell'opzione.

Una dimostrazione pratica sull'uso del menù è presente nel video introduttivo mentre, in questa pagina del manuale, le sezioni dedicate illustreranno alcune caratteristiche di ogni elemento dell'interfaccia. Per eventuali approfondimenti, si rimanda alla guida ufficiale di Auspice.

Albero filogenetico

Layout

L'albero filogenetico può essere visualizzato come pannello sulla sinistra (predefinito) o sull'intera larghezza della pagina (puoi attivare la modalità "**PIENO**", per il layout di pagina, tra le opzioni del **Display** del menù laterale).

Tramite il menù laterale, è inoltre possibile scegliere tra 5 diversi tipi di visualizzazione dell'albero (Rettangolare, Radiale, Non radicati, Orologio, Scatter).

Nel menù laterale sono anche presenti opzioni per la visualizzazione dei rami (**Branche**):

- i rami possono rappresentare le distanze temporali tra i campioni o la divergenza tra gli stessi (*Lunghezza di branca*);
- possono essere nascosti o mostrati gli intervalli di confidenza, tramite l'apposito pulsante, e le etichette dei rami (*Etichette di branche*);
- è possibile scegliere cosa mostrare nelle etichette delle foglie (*Tip Labels*);
- si può selezionare l'opzione per esplodere l'albero in base ai diversi metadati associati (*Explode Tree By*).

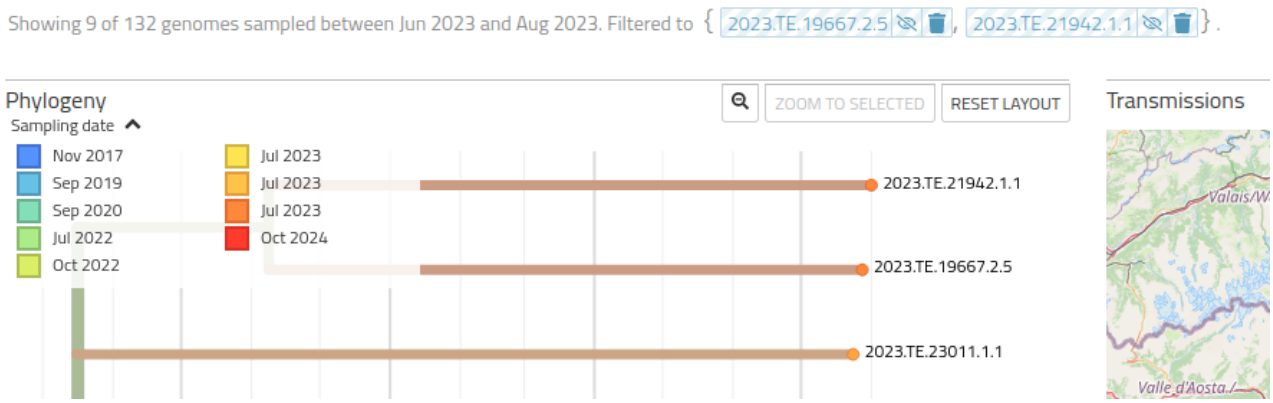
Legenda

La **legenda**, nell'angolo in alto a sinistra del riquadro *Filogenesi*, può essere visualizzata o nascosta cliccando sull'apposito pulsante a freccia (^); inoltre, al passaggio del puntatore, gli elementi della legenda - che corrispondono alle date di campionamento intese come "foglie" dell'albero filogenetico -, verranno evidenziate nel grafico.

Ingrandimento e campioni in evidenza

Posizionando il puntatore del mouse su un campione nell'albero, verrà visualizzata un'etichetta riassuntiva contenente informazioni sul quel campione. Cliccando invece sul campione o sui rami dell'albero, il grafico ingrandirà la regione di interesse e visualizzerà le informazioni complete sul campione. Le opzioni di ingrandimento potranno essere gestite dall'utente tramite i tre appositi pulsanti in alto: .

Cliccando sui campioni nel grafico, questi verranno aggiunti ad un filtro, usato per la visualizzazione dei campioni; la lista completa dei campioni nel filtro è visibile sopra l'area del grafico:



Sarà quindi possibile nascondere o visualizzare singolarmente i campioni selezionati (nascondendo tutti gli altri nel grafico); analogamente, è possibile eliminare singolarmente i campioni dal filtro. Nella sezione **Filter Data** del menù laterale è invece presente un'opzione che consente di eliminare in blocco i campioni dal filtro o di attivare/disattivare la visualizzazione di tutti i campioni selezionati.

Mappa

Per navigare nella mappa, sono disponibili alcuni semplici comandi:

- Usare la rotella del mouse tenendo premuto il tasto `Ctrl` per regolare lo zoom;
- Nell'angolo in basso a destra sono presenti i tasti **+/-** per aumentare o diminuire lo zoom. Fare doppio clic su una zona della mappa equivale a premere il tasto **+**;
- Tenendo premuto `Shift` (`Maiusc` su alcune tastiere) si può creare un'area di selezione con il mouse, che permette di ingrandire l'area selezionata;
- Lo zoom può essere resettato con l'apposito pulsante in alto a destra.

La mappa ed il grafico sono reciprocamente legati: la selezione di un nodo sul grafico evidenzierà anche la sua posizione sulla mappa.

Le opzioni nel menù laterale per la mappa permettono di scegliere il livello di risoluzione geografica (dipende dal livello di accuratezza dei metadati) e se mostrare o meno le linee di trasmissione.

Grafico della diversità

Sotto l'albero e la mappa è presente un terzo grafico (*Nucleotide diversity of genome*), che riporta la diversità tra i campioni, tenendo conto dei nucleotidi o degli amminoacidi (questi ultimi non disponibili per l'integrazione in piattaforma). Il grafico, inoltre, può mostrare la pagina "*Entropia*" o la pagina "*Eventi*" (il pulsante per passare da una pagina all'altra è nell'angolo in alto a destra) ed è possibile aumentare lo zoom sul grafico selezionando una finestra di valori nella linea graduata sottostante il grafico stesso.

Filtri

Al di sotto del grafico della diversità, sono elencati dei filtri applicabili, uno per ogni tipo di metadato associato ai campioni. Cliccando sulla riga di un filtro, la relativa scheda si espanderà, permettendo quindi di selezionare la tipologia di filtro:

Esempio: per filtrare i campioni isolati nella provincia di Ravenna, è necessario aprire la scheda di metadati "*Filtrare per Provincia*" e cliccare su "Ravenna".

Animazione

I campioni possono essere mostrati sul grafico e sulla mappa tramite un'animazione, partendo dai più vicini alla radice e andando verso le foglie. Dal menù laterale, si possono scegliere la velocità dell'animazione, l'attivazione della modalità loop e di quella cumulativa (mantiene la visualizzazione di tutti i nodi nella mappa dopo la loro prima apparizione e non li visualizza solo temporaneamente).

11. SOP

11.1 NGSManagerSOP

Questa pagina contiene l'elenco delle versioni della pipeline NGSManagerSOP e la lista di analisi (steps) che essa comprende.

> Il software NGSManagerSOP è sotto controllo di versione alla corrispondente pagina del servizio Gitea (interna a IZSAM).

11.1.1 Versioni

Versione NGSManagerSOP-2.3

(dal 23/06/2025)

Comprende i seguenti software di bioinformatica:

- fastqc: 0.11.5
- fastp: 0.23.4
- shovill: 1.1.0
- mlst: 2.23.0 - DB Pasteur aggiornato al 2025-06-23
- chewbbaca: 2.8.5

Corrispondenti agli step di Nextflow:

- step_1PP_trimming__fastp.nf
- step_2AS_denovo__shovill.nf
- step_4TY_MLST__mlst.nf
- step_4TY_flaA__flaA.nf
- step_4TY_cgMLST__chewbbaca.nf
- step_4TY_wgMLST__chewbbaca.nf

Versione NGSManagerSOP-2.2

(dal 12/11/2024)

Comprende i seguenti software di bioinformatica:

- fastqc: 0.11.5
- fastp: 0.23.4
- shovill: 1.1.0
- mlst: 2.23.0 - DB Pasteur aggiornato al 2024-11-06
- chewbbaca: 2.8.5

Corrispondenti agli step di Nextflow:

- step_1PP_trimming__fastp.nf
- step_2AS_denovo__shovill.nf
- step_4TY_MLST__mlst.nf
- step_4TY_flaA__flaA.nf
- step_4TY_cgMLST__chewbbaca.nf
- step_4TY_wgMLST__chewbbaca.nf

Versione NGSManagerSOP-2.1

(dal 28/08/2024)

Comprende i seguenti software di bioinformatica:

- fastqc: 0.11.5
- fastp: 0.23.1
- shovill: 1.1.0
- mlst: 2.23.0 - DB Pasteur aggiornato al 2024-08-19
- chewbbaca: 2.8.5

Corrispondenti agli step di Nextflow:

- step_1PP_trimming__fastp.nf
- step_2AS_denovo__shovill.nf
- step_4TY_MLST__mlst.nf
- step_4TY_flaA__flaA.nf
- step_4TY_cgMLST__chewbbaca.nf
- step_4TY_wgMLST__chewbbaca.nf

Versione NGSManagerSOP-2.0

(dal 27/09/2023)

Comprende i seguenti software di bioinformatica:

- fastqc: 0.11.5
- fastp: 0.23.1
- shovill: 1.1.0
- mlst: 2.23.0 - DB Pasteur aggiornato al 2023-09-27
- chewbbaca: 2.8.5

Corrispondenti agli step di Nextflow:

- step_1PP_trimming__fastp.nf
- step_2AS_denovo__shovill.nf
- step_4TY_MLST__mlst.nf
- step_4TY_flaA__flaA.nf
- step_4TY_cgMLST__chewbbaca.nf
- step_4TY_wgMLST__chewbbaca.nf

Versione NGSManagerSOP-1.1

(dal 10/05/2023 al 26/09/2023)

E' stato modificato il software mlst passando alla versione 2.23.0

Versione NGSManagerSOP-1.0

(dal 06/02/2021 fino al 09/05/2023)

Comprende i seguenti software di bioinformatica:

- fastqc: 0.11.5
- trimmomatic: 0.36
- spades: 3.11.1
- mlst: 2.16.0
- chewbbaca: 2.8.4

Corrispondenti agli step di Nextflow

- step_1PP_trimming__trimmomatic.nf
- step_2AS_denovo__spades.nf
- step_4TY_MLST__mlst.nf
- step_4TY_flaA__flaA.nf
- step_4TY_cgMLST__chewbbaca.nf
- step_4TY_wgMLST__chewbbaca.nf

11.1.2 Validazione e riproducibilità

Datasets

I tests di validazione e riproducibilità sono eseguiti sui datasets i cui indirizzi sono riportati di seguito:

- dataset *Listeria monocytogenes* - izs.te.b912.1.sop001
- dataset *Campylobacter* - izs.te.b3.1.3.sop071

Validazione

- Validazione pipeline NGSManagerSOP - izs.te.b912.1.sop001 (*Listeria monocytogenes*)
- Validazione pipeline NGSManagerSOP - izs.te.b3.1.3.sop071 (*Campylobacter*)

Riproducibilità

- Riproducibilità risultati della pipeline NGSManagerSOP - izs.te.b912.1.sop001 (*Listeria monocytogenes*)
- Riproducibilità risultati della pipeline NGSManagerSOP - izs.te.b3.1.3.sop071 (*Campylobacter*)

Controllo esiti tests

I links sopra riportati portano alle pagine riassuntive delle builds per la SOP *Listeria* e *Campylobacter*, per i processi di validazione e riproducibilità, rispettivamente. Per tutte le pagine sarà possibile visualizzare un grafico dell'andamento dei processi recenti lanciati per la validazione o per la riproducibilità.

Per controllare lo stato di validazione o di riproducibilità della pipeline è sufficiente **accedere ai link** riportati precedentemente e **verificare che tutte le caselle colorate del grafico siano verdi**.

> Gli stati dei singoli campioni del dataset di validazione sono presenti nel file .csv elencato in alto nella pagina. Esso riporta gli esiti come "PASS" (se i risultati ricadono nei limiti accettati per la validazione) o "FAIL" (nel caso i valori ottenuti non dovessero essere accettabili per la validazione).

Il video seguente mostra brevemente come verificare lo stato di validazione della pipeline NGSmanagerSOP. La procedura è identica per la verifica della riproducibilità, eccetto che per l'assenza dei files dei dataset di validazione.

12. Integrazioni SILAB

12.1 **Parte di Accettazione e automatismo iniziale**

12.1.1 **Logica funzionamento automatismo Listeria (LATO SILAB)**

1. Si impostano con **Blocco elaborazione RdP** tutte le accettazioni con tutti i seguenti requisiti:

   STATO <> ('ELIMINATA','REFERTATA')
   Hanno almeno un campione distribuito per **Estrazione acidi nucleici: DNA e metodo 'cf' oppure 'EA'**
   la specie del profilo di distribuzione rientra nelle specie della tabella di parametri ***BIOINFO_PARAMETRI**** **con tipologia** LISTERIA**
   non hanno nessun campione distribuito per il gruppo **LYN**:
   - *Determinazione del Multi-Locus Sequence Typing*
   - *Determinazione del core genome MLST (cgMLST)*
   - *Determinazione del Clonal Complex*

   Non hanno ancora il **Blocco elaborazione RdP** In questo modo non si rischia che venga elaborato RdP prima della fine degli esami.
2. Si processano tutte le accettazioni del punto 1:

   Se l'esito dell'esame Estrazione acidi nucleici: DNA è FAVOREVOLE:
   - si crea campione figlio con materiale DNA, se non presente
   - si distribuisce il campione creato per Sequenziamento WGS, se non presente

   Se l'esito dell'esame Estrazione acidi nucleici: DNA è SFAVOREVOLE:
   - Se è già presente, si provvede ad eliminare la distribuzione per Sequenziamento WGS su DNA
   - si elimina il Blocco RdP impostato precedentemente.
3. Nei casi in cui (es. NRG 2023/TE/11493) viene accettato direttamente il DNA per la listeria monocytogenes, e si distribuisce il WGS, non viene messo il blocco Rdp (perchè è solo una distribuzione) ma l'automatismo per risultati ed allegati è lo stesso.

12.1.2 **Logica funzionamento automatismo Campylobacter coli e jejuni (LATO SILAB)**

Al momento dell'accettazione campioni di Campylobacter coli e jejuni, l'operatore addetto deve distribuire i campioni per il gruppo **CAMPCEP** che comprende:

- *Campylobacter spp: Identificazione* con *MALDI – TOF (Spettrometria di massa)*
- *Estrazione acidi nucleici: DNA* con *chimico-fisico*
- *Antibiogramma Gram negativi- Campylobacter* con *MICRODILUIZIONE*

A volte, invece del metodo *MALDI – TOF (Spettrometria di massa)*, c'è necessità di distribuire metodo **PCR** per l'identificazione. In questi casi, si è deciso che comunque deve essere eseguita la distribuzione del gruppo sopra indicato, e modificare ilmetodo successivamente dalla maschera di acquisizione dati.

E' IMPORTANTE CHE SI FACCIA PRIMA DEGLI SCHEMI SISD ALTRIMENTI POI SARA' PIU' COMPLICATA LA MODIFICA DEL METODO IN QUANTO CONSIDERATA SUGLI SCHEMI

L'automatismo per Campylobacter ha inizio con:

1. Si impostano con **Blocco elaborazione RdP** tutte le accettazioni con tutti i seguenti requisiti:

   STATO <> ('ELIMINATA','REFERTATA')
   Hanno almeno un campione distribuito per **Estrazione acidi nucleici: DNA e metodo 'cf' oppure 'EA'**
   la specie del profilo di distribuzione rientra nelle specie della tabella di parametri ***BIOINFO_PARAMETRI*** **con tipologia** CAMPYLO**
   non hanno nessun campione distribuito per il gruppo **CAMRAW**:
   - *Determinazione del Multi-Locus Sequence Typing*
   - *Determinazione del core genome MLST (cgMLST)*
   - *Determinazione del Clonal Complex*
   - *Determinazione wgMLST*
   - *Campylobacter: determinazione del flaA SVR*

   Non hanno ancora il **Blocco elaborazione RdP**
   Non hanno già il WGS distribuito In questo modo non si rischia che venga elaborato RdP prima della fine degli esami.
2. Si processano tutte le accettazioni del punto 1, e:

   Se l'esito dell'esame Estrazione acidi nucleici: DNA è FAVOREVOLE e l'identificazione ha dato come esito un campylo jejuni o coli

   si crea campione figlio con materiale DNA, se non presente
   si distribuisce il campione creato per Sequenziamento WGS, se non presente

   Se l'esito dell'esame Estrazione acidi nucleici: DNA oppure l'identificazione NON ha dato come esito un campylo jejuni o coli:

   Se è già presente, si provvede ad eliminare la distribuzione per Sequenziamento WGS su DNA
   si elimina il Blocco RdP impostato precedentemente.

12.1.3 **Logica funzionamento automatismo Brucella (LATO SILAB)**

Al momento dell'accettazione campioni Brucella, l'operatore addetto deve distribuire i campioni per il gruppo **N7 (BRUCELLA IDENTIFICAZIONE)** che comprende:

- **Brucella spp: Identificazione** con **PCR-RFLP**
- **Brucella spp: Identificazione** con **PCR Bruce-Ladder**
- **Brucella spp: Identificazione** con **PCR Ladder Suis**
- **Estrazione acidi nucleici: DNA** con **Estrazione automatica**

Il Laboratorio ha deciso di far sempre distribuire Brucella spp: Identificazione con tutti e trew i metodi perchè la momento dell'accettazione non si sa quale metodo distribuire (dipende dal tipo della brucella). Poi loro eliminano le distribuzioni che non devono essere acquisite.

L'automatismo per Brucella ha inizio con:

1. Si impostano con **Blocco elaborazione RdP** tutte le accettazioni con tutti i seguenti requisiti:

   STATO <> ('ELIMINATA','REFERTATA')
   Hanno almeno un campione distribuito per **Estrazione acidi nucleici: DNA e metodo 'cf' oppure 'EA'**
   la specie del profilo di distribuzione rientra nelle specie della tabella di parametri ***BIOINFO_PARAMETRI*** **con tipologia** BRUCELLA**
   non hanno nessun campione distribuito per il gruppo **BRRAW**:
   Determinazione del Multi-Locus Sequence Typing
   Determinazione del core genome MLST (cgMLST)
   Analisi degli SNPs
   Non hanno ancora il **Blocco elaborazione RdP**
   Non hanno già il WGS distribuito In questo modo non si rischia che venga elaborato RdP prima della fine degli esami.
2. Si processano tutte le accettazioni del punto 1, e:

   Se l'esito dell'esame Estrazione acidi nucleici: DNA è FAVOREVOLE e l'identificazione ha dato come esito una brucella
   si crea campione figlio con materiale DNA, se non presente
   si distribuisce il campione creato per Sequenziamento WGS, se non presente
   Se l'esito dell'esame Estrazione acidi nucleici: DNA oppure l'identificazione NON ha dato come esito una brucella:
   Se è già presente, si provvede ad eliminare la distribuzione per Sequenziamento WGS su DNA
   si elimina il Blocco RdP impostato precedentemente.

12.1.4 **Logica funzionamento automatismo Salmonella UOMO (LATO SILAB)**

Al momento dell'accettazione campioni Salmonella, l'operatore addetto deve distribuire i campioni per il gruppo **SM1 (SALMONELLA IDENTIFICAZIONE - WGS/DNA)** che comprende: - **Salmonella spp: Identificazione** con **Identificazione sierologica** - **Estrazione acidi nucleici: DNA** con **Estrazione automatica**

L'automatismo per Salmonella ha inizio con:

1. Si impostano con **Blocco elaborazione RdP** tutte le accettazioni con tutti i seguenti requisiti:

   STATO <> ('ELIMINATA','REFERTATA')
   Specie Origine = UOMO
   Hanno almeno un campione distribuito per **Estrazione acidi nucleici: DNA e metodo 'cf' oppure 'EA'**
   la specie del profilo di distribuzione rientra nelle specie della tabella di parametri **BIOINFO_PARAMETRI** con tipologia **SALMONELLA**
   non hanno nessun campione distribuito per il gruppo **SMRAW**:
   - *Determinazione del Multi-Locus Sequence Typing*
   - *Determinazione del core genome MLST (cgMLST)*

   Non hanno ancora il **Blocco elaborazione RdP**
   Non hanno già il WGS distribuito
   In questo modo non si rischia che venga elaborato RdP prima della fine degli esami.
2. Si processano tutte le accettazioni del punto 1, e:

   Se l'esito dell'esame Estrazione acidi nucleici: DNA è FAVOREVOLE e l'identificazione ha dato come esito una Salmonella
   - si crea campione figlio con materiale DNA, se non presente
   - si distribuisce il campione creato per Sequenziamento WGS, se non presente

   Se l'esito dell'esame Estrazione acidi nucleici: DNA oppure l'identificazione NON ha dato come esito una salmonella:
   - Se è già presente, si provvede ad eliminare la distribuzione per Sequenziamento WGS su DNA
   - si elimina il Blocco RdP impostato precedentemente.

12.2 **Esecuzione ed Inserimento dati**

12.2.1 **DA BIOINFOTOOLS**

A. ESITI ESAMI

Per ciascun campione delle RUN da elaborare, si prendono i risultati dei vari accertamenti:

1. **WGS**:

   Per i campioni del tender si splitta il file separando le colonne con il carattere tab "**\t**" e si considerano le colonne **totReads** e **avgQual** per prendere i valori dei Total rawreads e Qscore
   Per tutti gli altri campioni si splitta il file separando le colonne con il carattere "," e si considerano le colonne **Total_rawReads** e **Mean_rawAvgQual** per prendere i valori dei Total rawreads e Qscore
   Se ci sono errori (file non trovato..) si restituisce "NON DISPONIBILI N. reads totali, Q-score".
   E' importante, affinchè vada bene inserimento dei risultati sul SILAB, che i due valori (readsa totali e QScore) siano divisi dal carattere '#' (es. 12345#30)
2. **MLST**:

   Considera il file della tipologia *...._mlst.tsv* e lo splitta considerando come separatore il carattere "**;**".
   Se il file non viene trovato si restituisce ANALISI BIOINFORMATICA NON COMPLETA
   Si considera il valore della colonna **ST**. Se si trova il carattere **-** (trattino), si controlla il numero di volte in cui è presente il carattere "?" tra i loci della colonna "loci(varianti)"; nel caso in cui si presenta più di una volta, si restituisce "**ANALISI BIOINFORMATICA NON COMPLETA**", altrimenti "**Trovato nuovo ST. Raw reads inviate all'Istituto Pasteur per l'assegnazione dei codici**".
3. **CC**:

   Considera il file della tipologia *...._mlst_cc.csv* e lo splitta considerando come separatore il carattere "**;**".
   Se il file non viene trovato si restituisce ANALISI BIOINFORMATICA NON COMPLETA
   Si considera il valore della colonna **CC**. Se si trova il carattere **-** (trattino), si controlla il numero di volte in cui è presente il carattere "?" tra i loci della colonna "loci(varianti)"; nel caso in cui si presenta più di una volta, si restituisce "**ANALISI BIOINFORMATICA NON COMPLETA**", altrimenti "**Trovato nuovo CC. Raw reads inviate all'Istituto Pasteur per l'assegnazione dei codici**".
4. **cgMLST**:

   Considera il file della tipologia **4TY_cgMLST/DS_*chewbbaca/qc/result/*_import_chewbbaca_check.csv** e lo splitta considerando come separatore il carattere "**,**".
   Se il file non viene trovato si restituisce ANALISI BIOINFORMATICA NON COMPLETA
   Si considera il valore della colonna **calledNum**.
5. **wgMLST**:

   Considera il file della tipologia **4TY_wgMLST/DS_*chewbbaca/qc/result/*_chewbbaca_check.csv** e lo splitta considerando come separatore il carattere "**,**".
   Se il file non viene trovato si restituisce ANALISI BIOINFORMATICA NON COMPLETA
   Si considera il valore della colonna **calledNum**.
6. **flaA**:

   Considera il file della tipologia **4TY_flaA/DS_*flaA/result/*_mbn_flaa_cc.csv** e lo splitta considerando come separatore il carattere "**;**".
   Se il file non viene trovato si restituisce ANALISI BIOINFORMATICA NON COMPLETA
   Si considera il valore della colonna **loci(varianti)** :
   - se contiene un numero, lo si considera com risultato
   - se contiene due numeri separati da , li si sonsidera entrambi come risultato
   - se contiene ? o ~ come risultato bisogna considerare "Trovato nuovo flaA. fasta inviate al database di riferimento pubMLST C.coli per assegnazione dei codici."

7. **SNPs**:

   Considera il file della tipologia **2AS_mapping/DS_*snippy/result/_snippy/meta/*_import_coverage.csv** e lo splitta considerando come separatore il carattere "**,**".
   Se il file non viene trovato si restituisce ANALISI BIOINFORMATICA NON COMPLETA
   Si considerano i valore delle colonne **COV** e **H_COV** considerando un solo decimale; **H_COV** viene riportato in valore percentuale.
   se *COV** (Coverage verticale) ha un valore inferiore a 30 si avrà esito: Analisi Bioinformatica non completa (Coverage veriicale <30X)
   se *H_COV** (Coverage orizzonatale) ha un valore inferiore a 97 si avrà esito: Analisi Bioinformatica non completa (Coverage orizzontale <97%)

Per ciascuno dei risultati dovrà essere indicato, oltre al DS ed esito, anche il tipo esame (wgs,cc,st,cgmlst,fla,wgmlst). Di seguito un esempio di risultati: `dsId;descrizioneEsito;tipoesito` `13335207;1954198#32.58;wgs` `13335207;ST883;mlst` `13335207;CCST-21 complex;cc` `13335207;-;flaa` `13335207;677;cgmlst` `13335207;1077;wgmlst` `13335207;32#99;SNPs` `13493744;2599284#32.5;wgs` `13493744;ST3335;mlst` `13493744;CCST-206 complex;cc` `13493744;14;flaa` `13493744;678;cgmlst` `13493744;1062;wgmlst` `13493744;30#97;SNPs`

Terminato il ciclo per la generazione del file risultati, si esegue la chiamata ad un metodo SILAB (tramite CURL) per inserire gli esiti **schemigrafici/schemadettaglio/insert/risultato/biomassivonoschema/** passando elenco dei risultati.

A1. Inserimento esiti lato SILAB

I valori limite (in base a quanto indicato nelle SOP) con cui effettuare i controlli sui dati ricevuti, e anche esiti di default per eventuali errori, sono inseriti in una apposita tabella BIOINFO_SOP_CHECKLIMIT distinti per SOP, Specie ed Esame.
L'inserimento dei risultati in SILAB, viene gestito dal PKG **BIOINFO_ACQUISIZIONE** del DB di SILAB eseguito dal nuovo metodo dell'applicativo SILAB biomassivonoschema nel seguente modo:

- si leggono i risultati passati ordinandoli per tipoesito in modo da analizzare prima il wgs
- Il DS indicato per ciascuna tipologia è sempre quello relativo all'esame WGS; per distinguere i vari campioni si agisce sulla tipologia indicata
- **Per gli esiti del WGS**
  SI CONTROLLA INNANZI TUTTO SE IL CAMPIONE RIENTRA NELLA PROCEDURA SOP e lo si fa attraverso la nuova funzione oracle di SILAB BIOINFO_ACQUISIZIONE.CTRL_SOP_LYN_CAMP.
  Se non rientra nella SOP, si procede all'inserimento del risultato proseguendo poi con il campione tipologia wgs successivo, altrimenti si procederà come descritto di seguito:
    In caso di esito *NON DISPONIBILI N. reads totali, Q-score*, si imposta il giudizio dell'esame in SFAVOREVOLE e si prosegue con il campione tipologia wgs successivo.
    In caso contrario, si effettua il controllo sui valori dell'esito. Se NON rientrano nei limiti indicati si setta il giudizio dell'esame in SFAVOREVOLE, si imposta la descrizione esito come indicato dalla SOP in base a quale valore non è rientrato nei limiti, e si prosegue con il campione tipologia wgs successivo.
    Se il risultato dei tot.rawreads e QSCORE rientrano entrambi nei limiti:
      si setta il giudizio in FAVOREVOLE
      si inserisce l'esito WGS
      si inserisce un campione figlio con materiale **Raw reads**
      si distribuisce il campione appena creato utilizzando il gruppo di distribuzione:
        LYN per listeria
        CAMRAW per Campylobacter
        BRRAW per Brucella
        SMRAW per Salmonella
      si prosegue con gli esiti dello stesso DS ma tipologia diversa da wgs:
        In caso di CC, MLST, flaA, se il risultato è un trattino "-", si legge la descrizione dell'esito dalla tabella di configurazione in base alla specie, sop ed esame.
        Per tutte le tipologie di esami, si effettua il controllo del risultato con gli eventuali limiti indicati dalla SOP e inseriti nella tabella di configurazione. Se l'esito non rientra nei limiti, si provvederà a settare il giudizio SFAVOREVOLE ed inserire come esito, l'eventuale messaggio impostato da tabella di configurazione.
- nel caso di **Campylobacter**, si controlla l'esito dell'identificazione (in MALDI o PCR) e, **se FAVOREVOLE (quindi un Campylo), mettiamo la specie isolata del campione come quella identificata**; **se SFAVOREVOLE invece, si provvederà ad eliminare tutte le eventuali distribuzioni eseguite successivamente su DNA e RAWREADS, compreso eventuali risultati** (?????DA VERIFICARE???)

B. File da allegare per cgMLST e wgMLST

Per ciascun campione delle RUN da elaborare, si leggono i files generati da cgMLST e wgMLST:

1. **cgMLST**:

   Considera il file della tipologia **4TY_cgMLST/DS_*chewbbaca/result/chewbbacacrc32*sv**
   legge il file e lo trasforma base64
   Crea una riga con indicazione DS,FILE base64, tipo esame (cgmlst)
   Abbiamo un file per ciascun campione
2. **wgMLST**:

   Considera il file della tipologia **4TY_wgMLST/DS_*chewbbaca/result/chewbbacacrc32*sv**
   legge il file e lo trasforma base64
   Crea una riga con indicazione DS,FILE base64, tipo esame (wgmlst)
   Abbiamo un file per ciascun campione
3. **SNPs**:

   Considera il file della tipologia *2AS_mapping/DS_snippy/result/*_snippy/result/*snps_only.vcf
   legge il file e lo trasforma base64
   Crea una riga con indicazione DS,FILE base64, tipo esame (SNPs)

Terminato il ciclo per lettura dei files, si esegue la chiamata ad un metodo SILAB (tramite CURL) per inserire gli allegati a ciascu NRG **tabelle/allegato/insert/bioallegatomassivo/** passando elenco allegati nel formato `dsId;fileStr;tipoEsame;estensione` dove in fileStr c'è il file formato base64.

B1. Inserimento allegati lato SILAB

L'inserimento degli allegati in SILAB, viene gestito dal metodo SILAB bioallegatomassivo che provvede a:

- Leggere quanto passato da bioinfotools
- raggruppare per esame (cgmlst o wgmlst o SNPs) e per accettazione: se un NRG ha più campioni, i risultati saranno uniti in un unico allegato, identificando il risultato di ciascun campione dal campo file (descritto di seguito).
- aggiungere, per ciascun risultato di campione, alla colonna "file", il codice univoco campione seguito dall'identificativo (es. 2023.TE.34.1.1_1-1-1 dove 2023.TE.34.1 è il codice UK e 1-1-1 identificativo del campione). **E' importante notare che l'identificato considerato è quello relativo al campione raw reads e non DNA**
- inserire l'allegato di ciascun NRG distinto per esame (es. 2023TE34_cgmlst.csv; in caso di campylobacter si avranno due file uno per cgmlst e l'altro per wgmlst 2023TE34_wgmlst.csv; in caso di brucella si avranno due file uno per cgmlst e uno per SNps). Gli allegati inseriti non avranno la validazione effettuata, ma si valideranno in automatico in fase di validazione esiti da silab.

12.3 **Elenco Profili selezionabili per esami di sequenziamento**

Nella tabellea **A** sono elencati i profili che devono essere validi e selezionabili per gli esami di sequenziamento, come indicato dall'incontro del 2.05.2024 con il reparto di Biotecnologie. Tutti gli altri profili attualmente esistenti, non inclusi in tabella, dovrebbero essere disattivati

12.3.1 **Tabella A** - Profili Validi

ACCERTAMENTO	METODO	METODICA	MATERIALE	SPECIE
Profili relativi alle SOP:				
Sequenziamento WGS	**Illumina**	IZS TE B2.1.9 SOP032 2023 Rev. 0	ACIDO DEOSSIRIBONUCLEICO (DNA)	Brucella
Sequenziamento WGS	**Illumina**	IZS TE B2.1.9 SOP032 2023 Rev. 0	ACIDO DEOSSIRIBONUCLEICO (DNA)	Campylobacter coli
Sequenziamento WGS	**Illumina**	IZS TE B2.1.9 SOP032 2023 Rev. 0	ACIDO DEOSSIRIBONUCLEICO (DNA)	Campylobacter jejuni
Sequenziamento WGS	**Illumina**	IZS TE B2.1.9 SOP032 2023 Rev. 0	ACIDO DEOSSIRIBONUCLEICO (DNA)	Listeria monocytogenes
Sequenziamento WGS	**Illumina**	Metodo interno	ACIDO DEOSSIRIBONUCLEICO (DNA)	Salmonella
Profili per Metagenomica SARS-CoV-2:				
Metagenomica SARS-CoV-2	**Illumina**		ACIDO RIBONUCLEICO (RNA)	SARS-CoV-2
Metagenomica SARS-CoV-2: Controllo Positivo	**Illumina**		ACIDO RIBONUCLEICO (RNA)	SARS-CoV-2
Metagenomica SARS-CoV-2: Controllo Negativo	**Illumina**		ACQUA DISTILLATA	.
Profili fuori SOP da utilizzare in tutti I casi in cui non deve bisogna far riferimento alla SOP accreditata; esitono stessi profili per ciscun metodo Illumina, iontorrent e Minioin nanopore:				
Sequenziamento WGS	**Illumina**		ACQUA DISTILLATA	.
Sequenziamento WGS	**Illumina**		ACIDO DEOSSIRIBONUCLEICO (DNA)	BATTERI
Sequenziamento WGS	**Illumina**		ACIDO DEOSSIRIBONUCLEICO (DNA)	VIRUS
Sequenziamento WGS	**Illumina**		ACIDO RIBONUCLEICO (RNA)	VIRUS
Sequenziamento WGS	**Iontorrent**		ACQUA DISTILLATA	.
Sequenziamento WGS	**Iontorrent**		ACIDO DEOSSIRIBONUCLEICO (DNA)	VIRUS
Sequenziamento WGS	**Iontorrent**		ACIDO DEOSSIRIBONUCLEICO (DNA)	BATTERI
Sequenziamento WGS	**Iontorrent**		ACIDO RIBONUCLEICO (RNA)	VIRUS
Sequenziamento WGS	**MinIon Nanopore**		ACQUA DISTILLATA	.

Sequenziamento WGS	**MinIon Nanopore**		ACIDO DEOSSIRIBONUCLEICO (DNA)	BATTERI
Sequenziamento WGS	**MinIon Nanopore**		ACIDO DEOSSIRIBONUCLEICO (DNA)	VIRUS
Sequenziamento WGS	**MinIon Nanopore**		ACIDO RIBONUCLEICO (RNA)	VIRUS
Profili per Ampliseq da usare per DNA su qualsiasi specie:				
Ampliseq	**Illumina**		ACIDO DEOSSIRIBONUCLEICO (DNA)	TUTTE
Profili per Metagenomica SG; per I metodi illumina e Minion sono ripetuti stessi profili; Per il metodo Analisi bioinformatica sui dati ngs c'è per materiale FILE:				
Metagenomica SG	**Illumina**		ACIDO DEOSSIRIBONUCLEICO (DNA)	TUTTE
Metagenomica SG	**Illumina**		ACQUA DISTILLATA	.
Metagenomica SG	**Analisi bioinformatica sui dati ngs**		FILE	TUTTE
Metagenomica SG	**MinIon Nanopore**		ACIDO DEOSSIRIBONUCLEICO (DNA)	TUTTE
Metagenomica SG	**MinIon Nanopore**		ACQUA DISTILLATA	.
Profili per Metagenomica16S:				
Metagenomica 16S	**Illumina**		ACIDO DEOSSIRIBONUCLEICO (DNA)	TUTTE
Profili per Metatrascrittomica; per I due metodi associati abbiamo stessi profili:				
Metatrascrittomica	**Illumina**		ACIDO RIBONUCLEICO (RNA)	TUTTE
Metatrascrittomica	**Illumina**		FILE	TUTTE
Metatrascrittomica	**MinIon Nanopore**		FILE	TUTTE
Metatrascrittomica	**MinIon Nanopore**		ACIDO RIBONUCLEICO (RNA)	TUTTE
Profili per Metagenomica SGTarget, con solo il metodo Illumina:				
Metagenomica SGTarget	**Illumina**		ACIDO DEOSSIRIBONUCLEICO (DNA)	BATTERI
Metagenomica SGTarget	**Illumina**		ACIDO RIBONUCLEICO (RNA)	BATTERI
Metagenomica SGTarget	**Illumina**		ACIDO RIBONUCLEICO (RNA)	VIRUS

Metagenomica SGTarget	**Illumina**		ACIDO DEOSSIRIBONUCLEICO (DNA)	VIRUS

12.4 **Tabella di configurazione per esclusione in automatismo del giro completo SOP**

Nell'incontro del 02.05.2024 con il reparto di biotecnologie, si è deciso di utilizzare i motivi per l'indicazione di effettuare o meno il "passaggio" completo come da procedura SOP (sequeziamento, cgmlst....), come da configurazione tabella **B**

12.4.1 **Tabella B** - Configurazione esecuzione automatismo SOP

ACCERTAMENTO	METODO	MOTIVO	Automatismo	METODICA	MATERIALE	SPECIE
Sequenziamento WGS	**Illumina**	**Verifica materiale di riferimento (2024-54)**	**NON COMPLETO - Solo WGS**		ACIDO DEOSSIRIBONUCLEICO (DNA)	BATTERI
Sequenziamento WGS	**Iontorrent**	**Verifica materiale di riferimento (2024-54)**	**NON COMPLETO - Solo WGS**		ACIDO DEOSSIRIBONUCLEICO (DNA)	BATTERI

13. Novità

13.1 2026

13.1.1 Gennaio 2026

Versione 26.01.1

La versione 26.01.1 introduce aggiornamenti all'esperienza d'uso e al layout di *Lancia Analisi*, *Download*, *Upload da NCBI* e *Report*: la scelta degli strumenti è più descrittiva, la ricerca è stata migliorata ed è possibile salvare le proprie preferenze. In questa versione sono presenti anche ottimizzazioni mirate al miglioramento delle performance.

Gli aggiornamenti di questo rilascio riguardano le seguenti aree:

- aggiornamenti UI/UX
- selezione e salvataggio preferenze
- performance e ottimizzazioni
- miglioramenti della localizzazione e risoluzione bug minori.

Aggiornamenti UI/UX

In questa versione sono stati aggiornati i layout e i comportamenti di *Lancia Analisi*, *Download*, *Upload da NCBI* e *Report*, con un cambiamento generale dell'interfaccia e l'introduzione di strumenti più descrittivi.

Le principali novita' includono:

- menu strumenti con testo descrittivo e link al wiki visibile (dove disponibile) prima della selezione;
- indicizzazione delle descrizioni per una ricerca più efficace e precisa;
- correzione del comportamento della ricerca nella vista ad albero, per renderlo coerente con la vista a *card*;
- migliorie UI, tra cui l'utilizzo di scrollbar personalizzate e la risoluzione di problematiche legate al ridimensionamento della finestra;
- miglioramento delle traduzioni di alcuni elementi della UI.

Aggiornamenti UI/UX

Selezione e salvataggio preferenze

Per le pagine *Lancia Analisi*, *Download*, *Upload da NCBI* e *Report* è ora possibile gestire le proprie preferenze tramite un sistema dedicato, che include:

- supporto alle preferenze utente, per i tool preferiti, in base al ruolo;
- aggiunta della categoria *Preferiti* per i tool contrassegnati con la stella.

Selezione e salvataggio preferenze

Performance e ottimizzazioni

Sono state migliorate le prestazioni di rendering delle classi e delle viste con bufferizzazione insieme ad animazioni piu' fluide:

- carico iniziale ottimizzato per classi, viste e join view;
- render progressivo con animazioni piu' morbide quando gli elementi entrano in viewport.

13.2 2025

13.2.1 Dicembre 2025

Versione 25.12.1

La versione 25.12.1 propone un aggiornamento in merito alle modalità di fruizione di alcune funzionalità, nell'ottica di garantire all'utente un'esperienza ulteriormente fluida nell'utilizzo della piattaforma.

Gli aggiornamenti di questo rilascio riguardano le seguenti aree:

- help desk
- lancia analisi e download
- funzioni di stampa
- miglioramenti della performance e risoluzione bug minori.

Help desk

Questo rilascio consente all'utente di contattare l'Help Desk direttamente dalla piattaforma, mantenendo la possibilità di descrivere l'issue, anche tramite immagini/video, dell'eventuale criticità da risolvere. Il video che segue mostra come usufruire della funzionalità *Help Desk*:

Help Desk update
Per maggiori informazioni, si prega consultare la pagina dedicata

Lancia Analisi e download

A partire da questo rilascio, la visualizzazione dei tool viene filtrata tramite la preselezione della prima categoria presente, anziché *All*; inoltre, i filtri applicati rimarranno memorizzati durante la sessione di lancio. Il video che segue mostra tale aggiornamento:

Preselezione categoria

Funzioni di stampa

Con questo rilascio, si ripristina l'opzione di stampa per:

- **singola scheda analisi** tramite il pulsante , in alto a destra della scheda;
- **una selezione multipla di schede** accessibile dal menu contestuale tramite il pulsante .

13.2.2 Settembre 2025

Versione 25.09.1

La versione 25.09.1 di GenPat introduce importanti miglioramenti che rendono l'esperienza utente più flessibile. Questo rilascio si concentra su due aree principali:

- **Gestione vista e miglioramenti esperienza d'uso delle tabelle** - Maggiore controllo sulla visualizzazione dei dati
- **Miglioramenti funzionalità di download** - Miglioramenti dell'esperienza utente

Gestione vista e miglioramenti esperienza d'uso delle tabelle

Le colonne delle tabelle presenti in GenPat adattano automaticamente la loro larghezza in base allo spazio disponibile. Da questo rilascio è possibile svincolare questo comportamento automatico e adattare le colonne in base al contenuto o alle vostre preferenze personali. Quando la larghezza complessiva della tabella supera lo spazio orizzontale disponibile, si attiva automaticamente lo scroll orizzontale.

Suggerimento Ricorda di salvare le tue preferenze di visualizzazione per ritrovarle configurate al prossimo accesso alla piattaforma.

Video su come cambiare vista delle tabelle.
Sempre sulle tabelle, sono stati introdotti i seguenti cambiamenti:

- nuovo pulsante nel menu contestuale delle colonne che permette di adattare istantaneamente la loro larghezza al contenuto;
- l'apertura della scheda di un record è stata circoscritta al click sul toggle **[+]** (quando presente), resta invece disponibile l'uso del doppio click sull'intera riga;
- pulsante di aggiunta filtri ridisegnato per una maggiore evidenza;

Video dimostrativo miglioramenti esperienza d'uso dell tabelle.

Nota Alcune delle modifiche descritte riguardano solo determinate tipologie di tabella, quelle definite come 'classi'.

Miglioramenti funzionalità di download

Questo rilascio include diverse ottimizzazioni che migliorano stabilità e usabilità per le funzionalità di download introdotte nella versione 25.04.1:

- possibilità di richiedere una notifica per evitare attese prolungate durante i processi di preparazione file;
- le nuove funzionalità sono ora disponibili anche per l'area **Download** della piattaforma;
- risoluzione di bug minori per un'esperienza più affidabile.

Video dimostrativo richiesta notifiche.

13.2.3 Aprile 2025

Versione 25.04.1

In questo rilascio:

- Nuova visualizzazione per la cartella dei risultati delle analisi
- Sezione dedicata ai campioni pubblici

Risultati analisi

I risultati delle analisi (Controllo Analisi) sono ora visibili tramite un sistema che permette di navigare e cercare i file generati senza lasciare la piattaforma. Sono stati inoltre introdotti strumenti di download che consentono di scaricare sia il singolo file generato che un insieme di più file a livello di cartella, selezione multipla e ricerca. Il video seguente mostra in sintesi le diverse operazioni rese disponibili dalle nuove funzioni.

analysis-results-view.mp4

Campioni pubblici

Per i profili abilitati è stata resa disponibile una nuova sezione che raccoglie tutti i campioni pubblici caricati sulla piattaforma. La sezione Campioni Genpat conterrà quindi solo i campioni relativi al proprio profilo (sia pubblici che privati).

13.2.4 Marzo 2025

Versione 25.03.1

In questo rilascio:

- Funzioni per l'invio di campioni ufficiali ad EFSA direttamente tramite la piattaforma.
- Novità sul wiki e gestione multilingua.
- Personalizzazioni della piattaforma configurabili in base al dominio.

EFSA Submissions

A partire da Campioni Genpat, è possibile avviare la procedura di invio di campioni ufficiali ad EFSA. Questa funzionalità è disponibile per i profili autorizzati. Maggiori informazioni saranno presto pubblicate su questo wiki, insieme a brevi video tutorial.

Novità sul wiki e gestione multilingua

Con questa versione, il wiki della piattaforma utilizza un sistema di documentazione basato su Docker che combina **MkDocs** con macro e plugin personalizzati. Questo sistema è stato rilasciato e reso liberamente disponibile su GitHub dal reparto di Bioinformatica.

Inoltre, il wiki supporta ora un sistema di localizzazione per la gestione multilingua. Diversi contenuti sono già stati tradotti in inglese, mentre altri sono in fase di scrittura. All'interno della piattaforma, i link vengono generati dinamicamente in base alle preferenze linguistiche dell'utente.

Personalizzazioni della piattaforma configurabili in base al dominio

Da questo rilascio, la piattaforma consente di personalizzare alcuni aspetti della **UI**, nonché i contenuti delle pagine di **Login** e della **Welcome page**, tramite parametri di configurazione collegati alla URL.

13.3 2024

13.3.1 Ottobre 2024

Versione 24.10.1

In questa versione sono stati rilasciati aggiornamenti relativi all'esperienza d'uso delle modalità di upload.

> **Nota Bene**
> Le procedure di preparazione e caricamento di metadati e sequenze rimangono invariate. Sono state riorganizzate esclusivamente a livello di interfaccia utente.

Di seguito una lista delle novità introdotte:

- le procedure di upload da NCBI e upload di file locali sono state separate in due aree distinte;
- il flusso di preparazione dei metadati e relativo caricamento dei file è stato reso più coerente ed entrambe le azioni avvengono all'interno della stessa pagina di upload;
- è stata aggiunta la possibilità di validare e procedere all'upload su una selezione specifica di metadati;
- i pulsanti di aggiunta dei metadati sono stati accorpati in un menu specifico ed affiancati con voci per l'accesso rapido al wiki;
- è ora possibile validare i metadati senza procedere all'upload;
- per chi ha i permessi necessari, è ora possibile aggiungere punti prelievo da sorgenti esterne, se non già presenti in GenPat.

Upload proprie sequenze

upload-files.mp4

Upload da NCBI

upload-from-ncbi-procedure.mp4

Aggiunta punto prelievo

upload-add-sampling-point.mp4

> Tutti i dettagli sono disponibili nella sezione Uploads di questo wiki.

13.3.2 Settembre 2024

Versione 24.09.1

In questo rilascio, il componente dati della piattaforma CMDBuild® è stato aggiornato dalla versione 3.4 alla versione 3.4.3, con l'integrazione di miglioramenti significativi nell'utilizzo delle classi e delle viste.

La nuova versione introduce anche alcuni miglioramenti dell'esperienza d'uso della piattaforma insieme alla risoluzione di bug minori.

Preferiti nel menu di navigazione

Aggiunta la possibilità di inserire nel menu di navigazione principale elementi preferiti per un più rapido accesso.

favourites-items.mp4

Ordinamento colonne

Da questa versione è possibile ordinare alfabeticamente la lista delle colonne da rendere visibili o non visibili sulle tabelle.

ordinamento-colonne.mp4

Operatore OR nei filtri avanzati

Aggiunta la possibilità di utilizzare l'operatore `OR` nei filtri avanzati sulle tabelle.

operatore-or-nei-filtri-avanzati.mp4

Changelog completo

Al seguente link è possibile consultare il changelog completo del componente https://www.cmdbuild.org/en/download/changelog

13.3.3 Luglio 2024

Versione 24.07.1

In questo rilascio sono stati introdotti un nuovo strumento per le analisi, alcuni cambiamenti che riguardano l'esperienza utente insieme alla risoluzione di bug minori.

Pharokka

È stato aggiunto in Genpat il software Pharokka - annotation tool for bacteriophage genomes. Questo tool consente di ottenere l'annotazione del genoma specificamente per i batteriofagi. Ora è disponibile per l'analisi 4AN_genes (4AN_genes__pharokka).

Esperienza utente

Nella versione corrente sono state introdotte nuove funzioni che coinvolgono l'esperienza d'uso dell'applicativo. La prima riguarda una gestione più avanzata delle notifiche di errore, coprendo un range più ampio di casistiche per fornire messaggi più significativi sugli errori incontrati. Questa funzione include anche un controllo sui messaggi esistenti per evitare duplicazioni e un bottone per la chiusura massiva dei messaggi a schermo. La seconda ha a che fare con le notifiche e riguarda l'introduzione di un bottone che consente di ignorarle tutte.

notifications-ignore-all.mp4

13.3.4 Marzo 2024

Versione 24.03.1

In questo rilascio introduzione di un nuovo sistema di **Upload**, insieme alla risoluzione di bug minori.

Upload

Nella versione corrente è possibile trovare una nuova modalità di caricamento delle sequenze in GenPat. Il nuovo sistema raccoglie tutte le tipologie di upload all'interno della stessa stessa pagina e per l'**upload manuale** è stata prevista una fase di caricamento e validazione dei metadati al fine di ridurre i margini d'errore.

La nuova pagina è accessibile tramite menu delle **Azioni** in header, tramite bottone **Upload** nella sezione `Azioni principali` della pagina di benvenuto e attraverso menù di navigazione laterale:

```
Upload/
    ...
    Upload page
    ...
```

Le vecchie modalità sono ancora disponibili ed accessibili tramite menu di navigazione laterale e verranno tenute per permettere una fase di passaggio graduale al nuovo sistema.

```
Upload/
    ...
    ...
    Upload legacy/
```

Maggiori informazioni possono essere trovate nella sezione Upload di questo wiki.

13.3.5 Febbraio 2024

Versione 24.02.1

In questo rilascio novità per le funzioni del bottone **wiki**, introduzione del servizio **help desk** insieme ad aggiustamenti e risoluzione bug minori.

Cerca nel wiki

Facendo click sull'icona **wiki** in alto a destra dell'applicativo potrai trovare un menu contestuale con diverse azioni. Una di queste è la funzione di ricerca rapida sul wiki, direttamente a partire dalla piattaforma, come mostrato nel video.

search_on_wiki.mp4

Help desk

Altra voce presente nel nuovo menu contestuale del wiki è quella dedicata all' **help desk**. Tramite questa azione si aprirà una modale che permette di selezionare un'area di riferimento ed aprire un ticket direttamente tramite email o tramite link diretto a **Redmine** per gli utenti registrati al servizio, senza dover lasciare l'applicativo.

help-desk.mp4

13.4 2023

13.4.1 Novembre 2023

In questa sezione verranno messi a disposizione descrizioni e video utili sulle novità introdotte dai rilasci dell'applicativo.

Versione 23.11.1

Nella versione corrente cambiamenti sulle funzioni di **segnalazione errori** per i risultati degli esami insieme ad aggiustamenti e risoluzione bug minori.

- Segnala errore dei risultati esami
- Risoluzione bug minori

Segnala errore dei risultati esami

Le funzioni gestite dalla pagina **Setta Flag Default dei Risultati Esami** sono state inserite direttamente nella pagina Risultati esami.

All'interno del relativo menu contestuale potrai trovare infatti la voce **Segnala errore**. Questa voce permette di marcare come errati gli elementi selezionati attraverso la tabella.

> Puoi trovare maggiori informazioni nella pagina dedicata di questo wiki.

mark-error-exam-results.mp4

Risoluzione bug minori

In questo rilascio troverai anche:

- aggiustamenti di layout sulla selezione multipla all'interno delle tabelle;
- risoluzione bug di visualizzazione delle icone sul sistema di avvisi;
- correzione di errori sulle label di selezione reference nel lancia analisi.

13.4.2 Ottobre 2023

Versione 23.10.4

Nella versione corrente potrai trovare il bottone **Aggiungi al carrello** in nuove aree dell'applicativo: Campioni Genpat, Campioni Alias, Relazioni Alias-Genpat, Risultati esami, Motivi-Genpat e nelle pagine di Check sotto Reports/Dettaglio/Checks.

Troverai anche alcuni cambiamenti nella UI delle schede di Controllo analisi e di Controllo downloads.

Versione 23.10.3

In questo rilascio cambiamenti sulle funzioni di **Export** e **Stampa** per le tabelle e due nuove funzioni per la tabella **Campioni Genpat**.

- Esporta CSV
- Esporta e copia codici campioni

Esporta CSV

Le funzioni di stampa presenti nella toolbar delle tabelle, sono stati integrati all'interno dei menu contestuali ed ora consentono di **esportare come file csv l'intera tabella** o di **esportare solo gli elementi selezionati**.

export-csv.mp4

Esporta e copia codici campioni

Per la tabella **Campioni Genpat**, sono state introdotte due nuove funzioni:

- **Esporta codici da selezione** che permette di esportare un file CSV con i soli codici campione
- **Copia codici da selezione** che permette di copiare negli appunti i codici campione separati da virgola

I comandi possono essere trovati all'interno del menu contestuale della tabella. Di seguito un breve video con degli esempi d'uso di queste nuove funzionalità.

export-copy-codes.mp4

Versione 23.10.2

In questa versione sono stati introdotti nuovi meccanismi per consentire la conservazione di un'analisi, insieme ad altri miglioramenti e risoluzione bug.

- Conserva analisi
- Miglioramenti e risoluzione bug

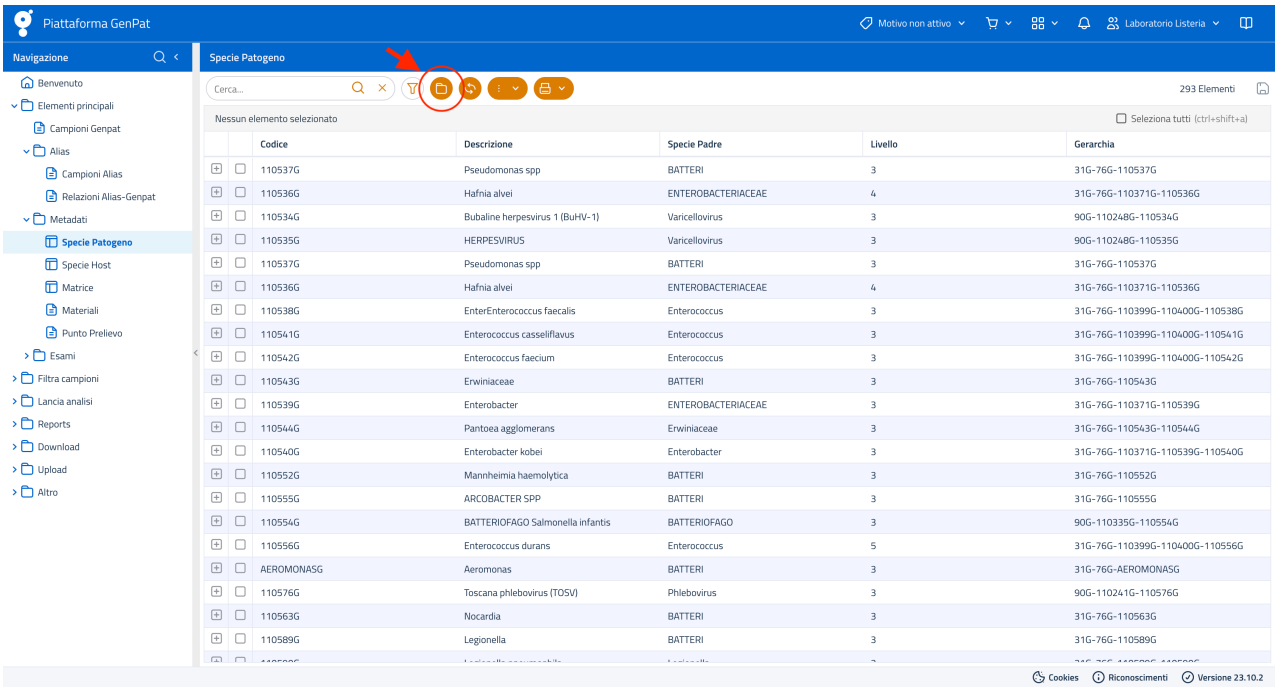
Conserva analisi

All'interno della scheda analisi è stata inserita la voce modificabile "**Conserva analisi**", attraverso cui potrai decidere se conservare indefinitamente un accertamento eseguito e non importabile, impedendo che i relativi dati vengano rimossi dal sistema una volta scaduto il tempo limite di un mese.

keep-analyses.mp4

Miglioramenti e risoluzione bug

Per le tabelle che contengono informazioni sulla struttura gerarchica degli elementi, è stato operato un miglioramento suull'accessibilità dello strumento **Vedi gerarchia**: il bottone è stato spostato direttamente nella toolbar e non sarà più presente come voce all'interno del menu contestuale.



Troverai anche alcuni aggiustamenti di layout sulle schede delle tabelle. In questo rilascio sono stati corretti anche alcuni bug minori.

Versione 23.10.1

In questo rilascio una nuova versione del sistema di **download** insieme alla risoluzione di problematiche minori.

- Aggiornamento sistema Download
- Risoluzione problematiche minori

Aggiornamento sistema Download

Con la nuova versione della pagina di download è possibile scaricare i risultati di qualunque analisi *Single Sample* disponibile in GenPat, in una struttura uniformata al sistema di lancio analisi.

Il nuovo download introduce la possibilità di scaricare i risultati di un'analisi filtrando in base a:

- tool bioinformatico con cui sono stati prodotti;
- sottocategoria di files tra quelli prodotti dall'analisi selezionata.

Per un dato set di campioni, tali novità permettono di accedere più facilmente ai risultati:

- di analisi della stessa tipologia ma eseguite con software diverso;
- di un'analisi meno recente.

Il video seguente dimostra l'uso del nuovo sistema di download, evidenziandone i vantaggi per l'utente. La guida completa è invece disponibile alla pagina Wiki "Download".

Genpat-download.mp4

Risoluzione problematiche minori

Nell'area di gestione del carrello, la voce di default per l'aggiunta di campioni è ora **Da codici** invece di **Da csv**. Risolto un bug che non permetteva di aprire tabelle direttamente all'interno delle schede principali. Migliorato strumento di ricerca per il menu di navigaziione principale. Risolto un bug che non permetteva di visulaizzare filtri nella vista tabulare nelle pagine di Lancia analisi, Download e Reports.

13.4.3 Settembre 2023

Versione 23.09.3

In questo rilascio potrai trovare miglioramenti di esperienza utente insieme a nuove funzionalità relative ai **motivi**.

- Modifica visibilità di un motivo
- Lista motivi semplificata

Modifica visibilità di un motivo

Quando si crea un motivo è possibile indicarne il tipo di visibilità, questa può assumere due valori **Solo io** o **Il mio gruppo**: nel primo caso il motivo può essere visto solo da chi ha effettivamente creato il motivo, nel secondo può essere visto da tutti gli utenti che appartengono allo stesso gruppo.

Da questo rilascio sarà possibile modificare la visibilità di un motivo precedentemente creato direttamente dalla sua scheda, nell'area **Motivi**, come brevemente mostrato nel video che segue.

> **IMPORTANTE**
> Potrai cambiare la visibilità del motivo solo se sei stato tu a crearlo.

change-tag-visibility.mp4

Lista motivi semplificata

La tabella contenente la lista dei motivi disponibili per l'attivazione è stata semplificata, al suo interno troverai ora solo i tuoi motivi o del tuo gruppo.

Potrai visualizzare i motivi di sistema (ad esempio quelli derivati dalle **RUN**) tramite una nuova vista presente nel menu di sinistra chiamata **Motivi pubblici**. Puoi trovarla all'interno di **Filtra campioni / Motivi**.

Potrai ora scegliere un motivo da attivare tra due sorgenti, **Miei e del mio gruppo** oppure **Pubblici**. Il video che segue mostra brevemente come selezionare l'uno o l'altro attraverso la pagina di gestione dei motivi.

tag_source_selection.mp4

Versione 23.09.2

A decorrere del 27/09/2023, per uniformarsi ai requisiti delle pipeline "WGS - OneHealth analytical pipeline" di EFSA (https://dev.azure.com/efsa-devops/EFSA/_git/efsa.wgs.onehealth) e garantire la validità delle analisi effettuate, il software chewBBACA per il calcolo dei profili allelici è stato aggiornato dalla versione 2.8.4 alla versione 2.8.5.

L'aggiornamento interessa la pipeline relativa ai batteri di **NGSManager**, il quale passa quindi dalla versione 1.1 alla **versione 2.0** (cronologia degli aggiornamenti della pipeline NGSManager: NGSManagerSOP).

Ora NGSManager usa (in conformità con la pipeline EFSA):

- **fastp 0.23.1** per il trimming;
- **shovill 1.1.0** per l'assembly;
- **chewBBACA 2.8.5** per calcolare la chiamata allelica.

Per *Listeria monocytogenes* è stato anche aggiornato lo schema cgMLST. Il sistema di sorveglianza per *Listeria monocytogenes* utilizzerà quindi la nuova modalità di calcolo dei profili allelici, che prevede:

- schema `l_mono_chewie_1748_220623`;
- encoding `crc32`.

Sono stati ricalcolati con la nuova pipeline NGSManager: - trimming, assembly e profilo allelico per *Listeria monocytogenes*; - trimming, assembly, profilo allelico ed esami di typing per *Campylobacter*, *Salmonella* ed *Escherichia coli*.

I nuovi profili allelici utilizzano quindi l'encoding crc32.

Versione 23.09.1

In questo rilascio potrai trovare l'informativa sui cookie accessibile direttamente nel footer.

see-cookies-policy.mp4

13.4.4 Luglio 2023

Versione 23.07.1

In questo rilascio potrai trovare miglioramenti di esperienza utente insieme a nuove funzionalità relative al **Lancia analisi**:

- Lancia analisi con RIS_CD
- Filtri avanzati per viste

Lancia analisi con RIS_CD

Da questa versione è possibile utilizzare un file csv risultati (RIS_CD) per le analisi. Il caricameno del csv permette di procedere allo step 2 della procedura senza prendere in considerazione carrello o motivo attivo.

load_ris_cd.mp4

Filtri avanzati per viste

Le funzionalità di filtro avanzato sono ora disponibili per tutte le tabelle di tipo `Vista`.

- Scopri di più sui filtri avanzati

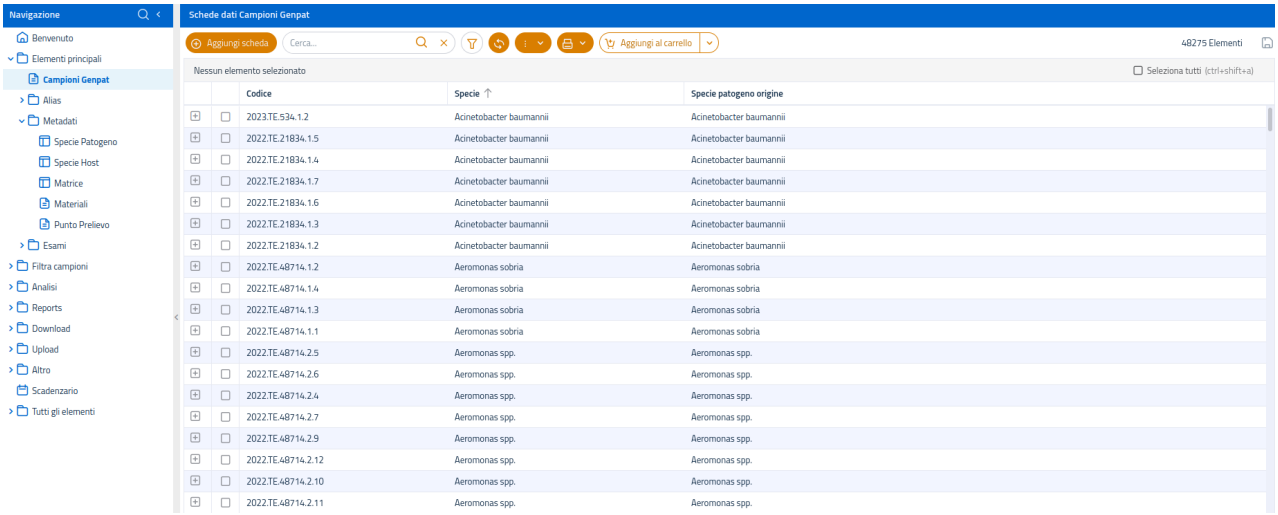
13.4.5 Giugno 2023

Versione 23.06.2

Per garantire uniformità di codifiche agli utenti esterni, nella piattaforma GenPat i metadati dei campioni ora aderiscono alle specifiche EFSA FOODEX2.

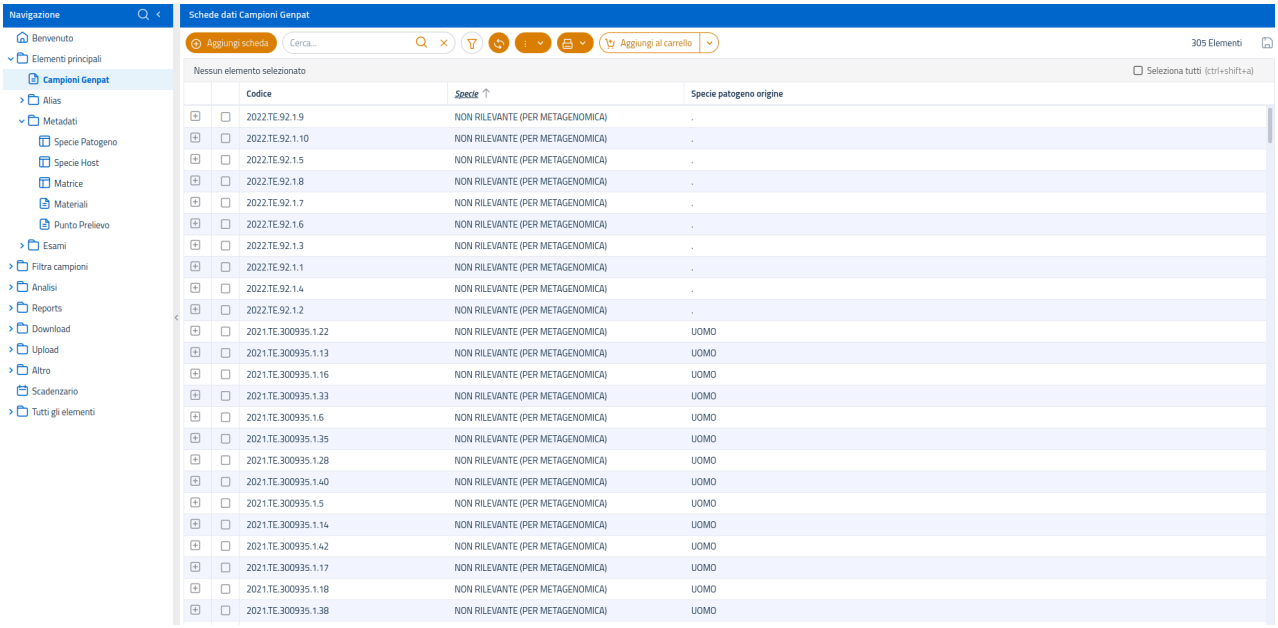
È stata implementata la categoria di "specie GenPat", che riporta i virus e batteri analizzati.

Nella sezione Campioni GenPat, nel campo "Specie" viene quindi ora riportato il risultato del mapping tra la specie di origine del campione e la specie propria dell'applicativo GenPat, ovvero con fonte "Genpat (Map da EFSA)". La specie patogena di origine rimane visibile nell'apposito campo "Specie Patogeno Origine", come mostrato nell'immagine seguente.



I casi di "MAPPING NON TROVATO" sono dovuti a specie contrassegante come "" o all'assegnazione di una Specie Patogeno diversa da una specie batterica o virale (ad esempio "BOVINO").

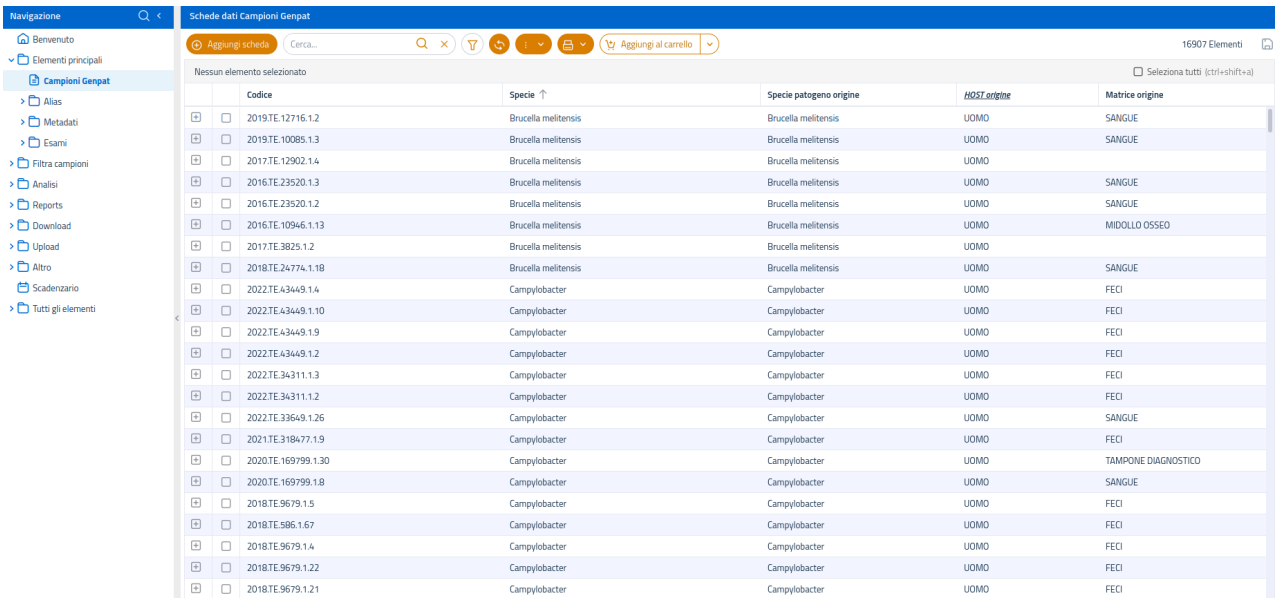
I casi di "NON RILEVANTE (PER METAGENOMICA)" sono dovuti a specie patogeno contrassegante come "TUTTE" e a materiale patogeno identificato come "ACIDO RIBONUCLEICO (RNA)".



È importante utilizzare la Specie GenPat (riconoscibile dal codice che termina con "G") anche negli upload.

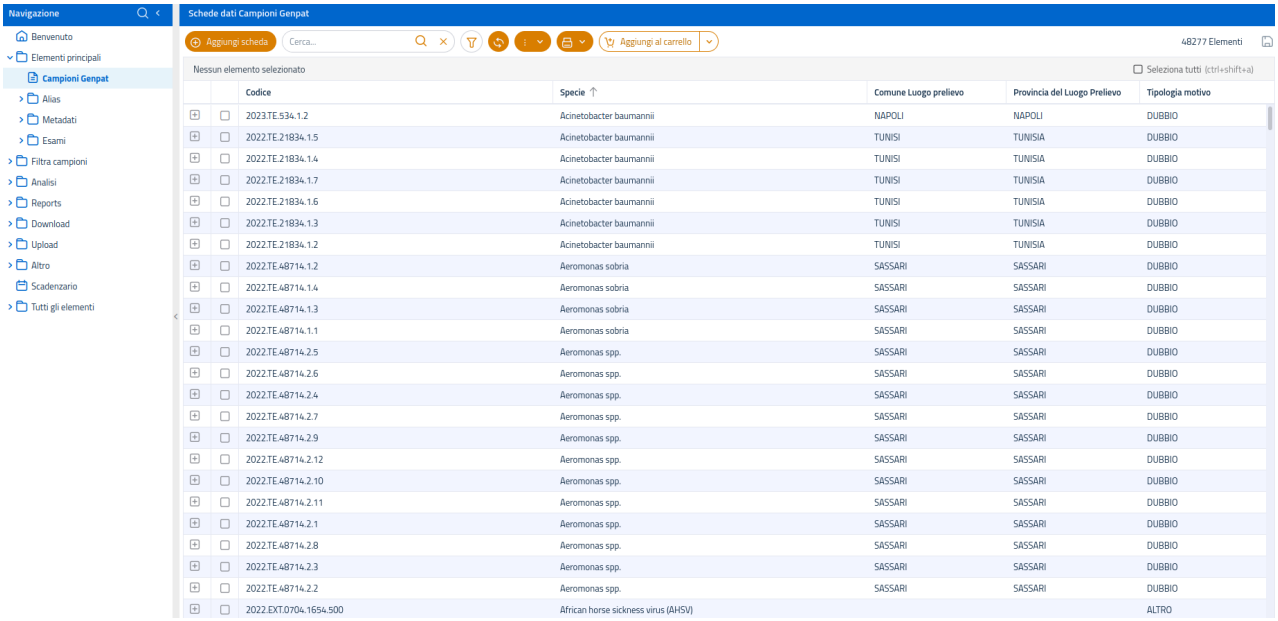
È inoltre fondamentale riportare nei file per upload (CMP_ALIAS.csv) i relativi codici specie della fonte "GenPat (Map da EFSA)". Ad esempio, nel file di upload per Listeria, il codice della specie deve essere "86G" (corrispondente alla fonte GenPat) e non con "86".

Su campioni GenPat è stato inoltre implementato il mapping di specie, host e matrice come da indicazioni EFSA Foodex2. Restano visibili comunque anche le informazioni di host e matrice originali (antecedenti al mapping con EFSA), indicati con le etichette "Matrice origine" e "Host origine".



Sono state aggiunte le seguenti informazioni nella sezione "Campioni GenPat", consultabili attivando la colonna appropriata:

- "Comune Luogo Prelievo" e "Provincia Luogo Prelievo";
- "Tipologia Motivo", da dove è possibile capire se un campione appartiene o meno ad un motivo ufficiale.



Le informazioni disponibili nella sezione "Alias" rimangono invariate (così come indicate in origine).

Versione 23.06.1

In quest'ultimo rilascio è stato introdotto un sistema automatico all'interno del lancia analisi che permette di visualizzare, per i tool che hanno una pagina wiki dedicata, una breve descrizione estratta dalla documentazione insieme ad un link diretto ad essa.

wiki-for-tools.mp4

13.4.6 Maggio 2023

Versione 23.05.4

Introdotto strumento di ricerca per il menu di navigazione. Nel video che segue un rapido esempio di utilizzo.

main-actions.mp4

Versione 23.05.3

Nella versione 23.05.3 sono stati risolti errori testuali segnalati per alcuni elementi di interfaccia utente.

Versione 23.05.2

In questo rilascio sono stati risolti alcuni problemi minori riscontrati sulle localizzazioni degli elementi di UI presenti in header e footer. Inoltre è stato tradotto in italiano il cotenuto presente nella modale dei **Riconoscimenti**.

Versione 23.05.1

In questo rilascio potrai trovare miglioramenti di esperienza utente insieme a nuove funzionalità relative alla visualizzazione della versione di GenPat:

Azioni principali

Nella **Welcome page** e in header vengono mostrati dei bottoni per le azioni principali che possono essere intraprese in GenPat (`Analisi` , `Downloads` , `Uploads` , etc.), da questa versione le voci riflettono automaticamente i permessi del gruppo cui fai parte. Pertanto queste voci potranno essere visibili o non visibili.

Nel seguente video un esempio di come il cambio gruppo può andare ad influire sulla visualizzazione delle azioni possibili.

main-actions.mp4

Versione di GenPat

Il numero di versione di GenPat è ora visibile in basso a destra dell'applicativo. Nella maggior parte dei casi non è richiesta alcuna azione da parte tua ma l'aggiornamento di alcune componenti del sistema potrebbero richiedere un reload dell'applicazione tramite modale durante la tua sessione di lavoro.

> **Importante!** Se non vuoi perdere il tuo lavoro in corso puoi momentaneamente evitare di ricaricare l'applicativo rispondendo `No` alla modale. Avrai modo di lanciare un reload successivamente, cliccando sul numero di versione posizionato in basso a destra. Puoi capire che un reload è necessario perchè vedrai un'icona identificativa del reload ed un asterisco al termine del numero di versione. Di seguito un breve video esplicativo.

genpat-version.mp4

13.4.7 Aprile 2023

Versione 23.04.1

In questo rilascio potrai trovare miglioramenti di esperienza utente insieme a nuove funzionalità relative al lancio analisi:

Miglioramenti sistema notifiche

Aggiunta la possibilità di ignorare le notifiche e rimuoverle dunque dall'elenco. Le notifiche avranno quindi tre tipologie di stato:

- **non lette** vengono segnalate da pallino blu di fianco alla data della notifica
- **lette** segnalate da pallino grigio di fianco alla data della notifica
- **ignorate** non mostrate in elenco

Per ignorare una notificha puoi semplicemente cliccare sul relativo **cestino** come mostrato nel seguente video.

ignore-notification.mp4

Selezione multipla di reference

Accertamento: 2AS_mapping, metodo: ivar

È stata abilitata la selezione multipla di reference per i mapping effettuati con accertamento: 2AS_mapping, metodo: ivar.

Quando più di un reference è selezionato tra i risultati si avrà una cartella aggiuntiva 'aggregate', nella quale è disponibile un riepilogo di tutti i valori di 'coverage' ottenuti. È possibile importare il risultato dell'esame in genpat - verrà creato un 'risultato esame' per ogni reference indicato.

Nuovo modulo: Mapping di reference segmentati

È stato aggiunto il modulo: 'Mapping di reference segmentati' che effettua un mapping con i reference selezionati utilizzando il metodo 'ivar' ed infine costruisce un 'multifasta' utilizzando i mapping ottenuti - seguendo l'ordine con cui i references sono specificati.

È possibile importare il risultato dell'esame in genpat - verrà creato un 'risultato esame' per ogni reference indicato ed in aggiunta un risultato per il 'multifasta'.

> **Importante** L'ordinamento dei reference in questo caso influisce sull'esecuzione dell'analisi.

Di seguito un video esplicativo su come effettuare una selezione multipla di reference.

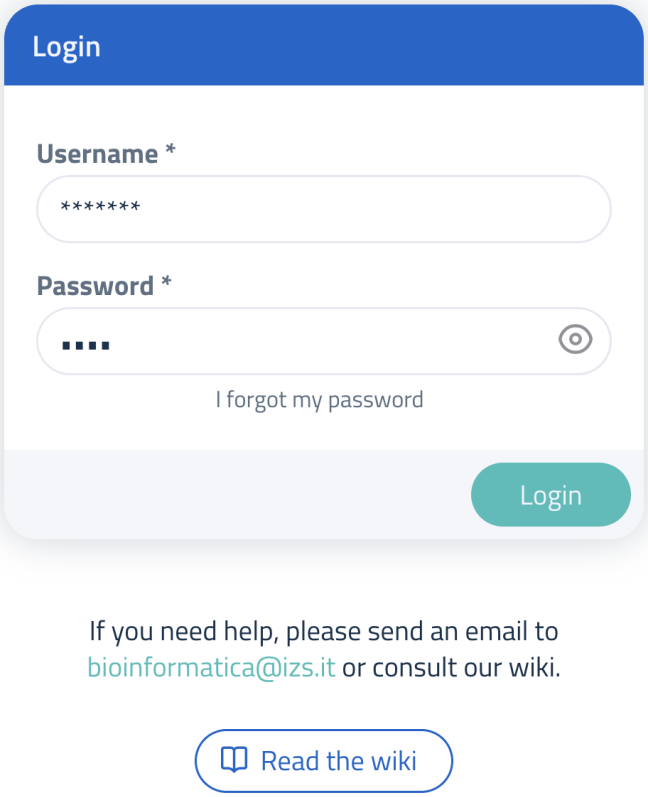
multireference.mp4

14. Supporto

14.1 Gestione password

La piattaforma è stata estesa con un **sistema di autenticazione esterno**, di conseguenza è possibile che per le procedure relative al cambio o al reset della password sarai reindirizzatə ad un servizio al di fuori di essa.

Mentre per il recupero della password puoi utilizzare il tip **Ho dimenticato la password** presente sotto la maschera di login.

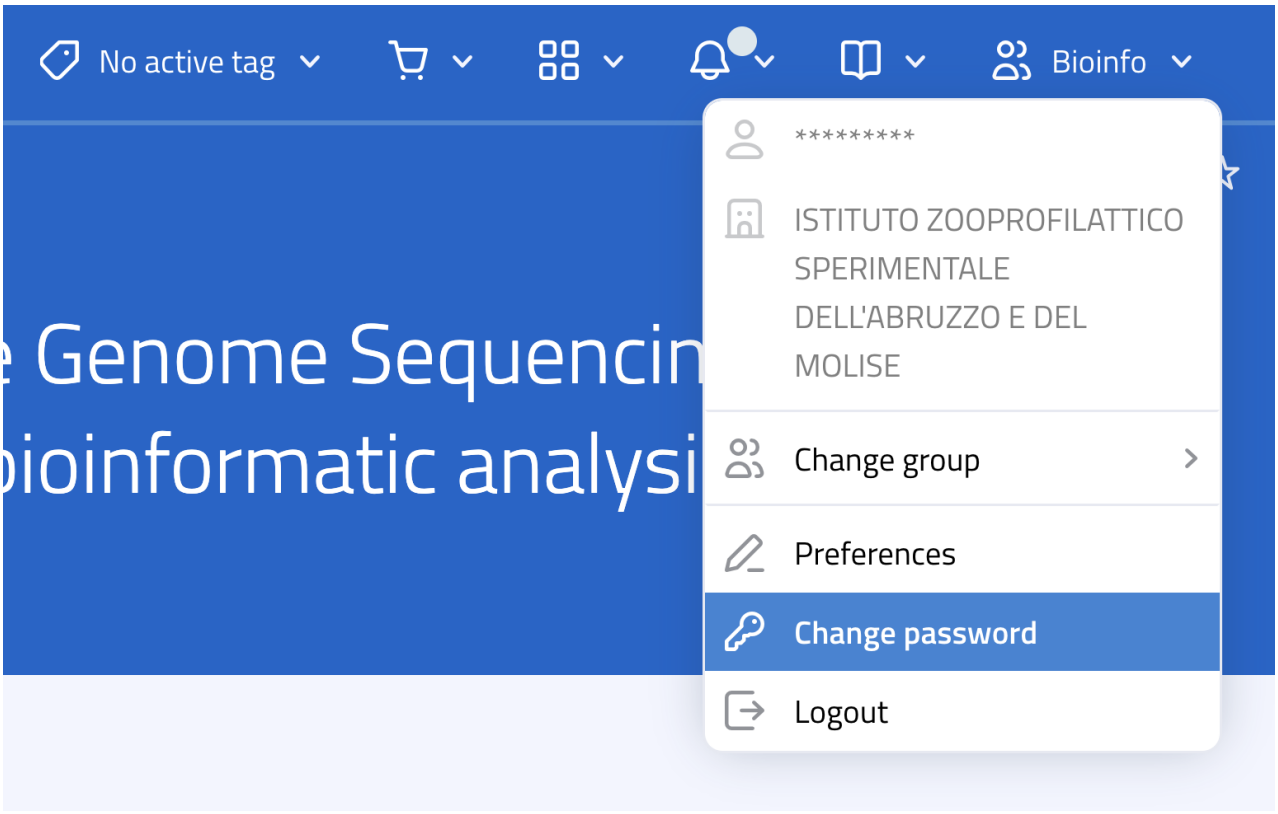


Cambia password

14.1.1 Procedure di cambio password

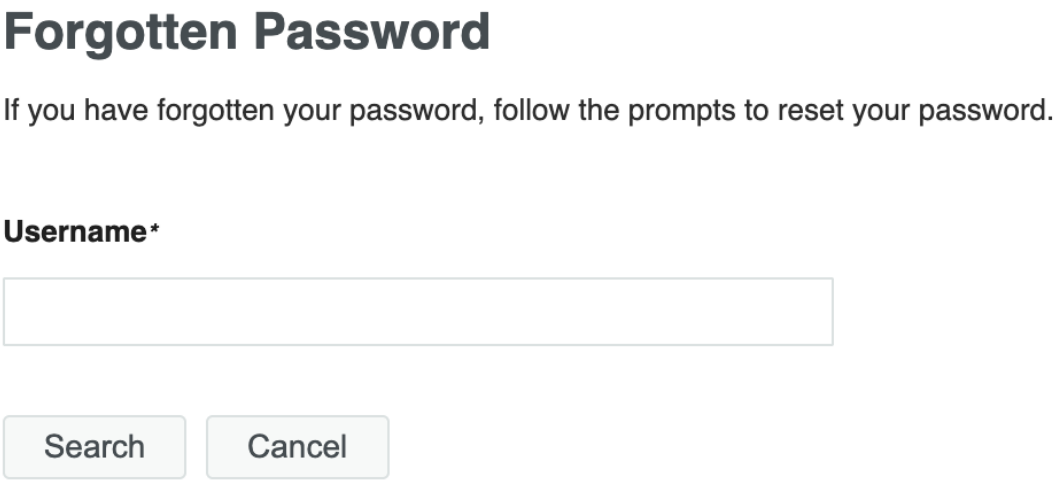
Se la mail con la quale sei registrato in piattaforma è `@izs.it` verrai reindirizzatə al servizio di **Self Service Password Reset (SSPR)** del tuo Istituto, una guida dettagliata è presente in area riservata.

Per modificare la tua password, utilizza la voce **Cambia password** disponibile nel **menu utente** in alto a destra sulla piattaforma.



Cambia password

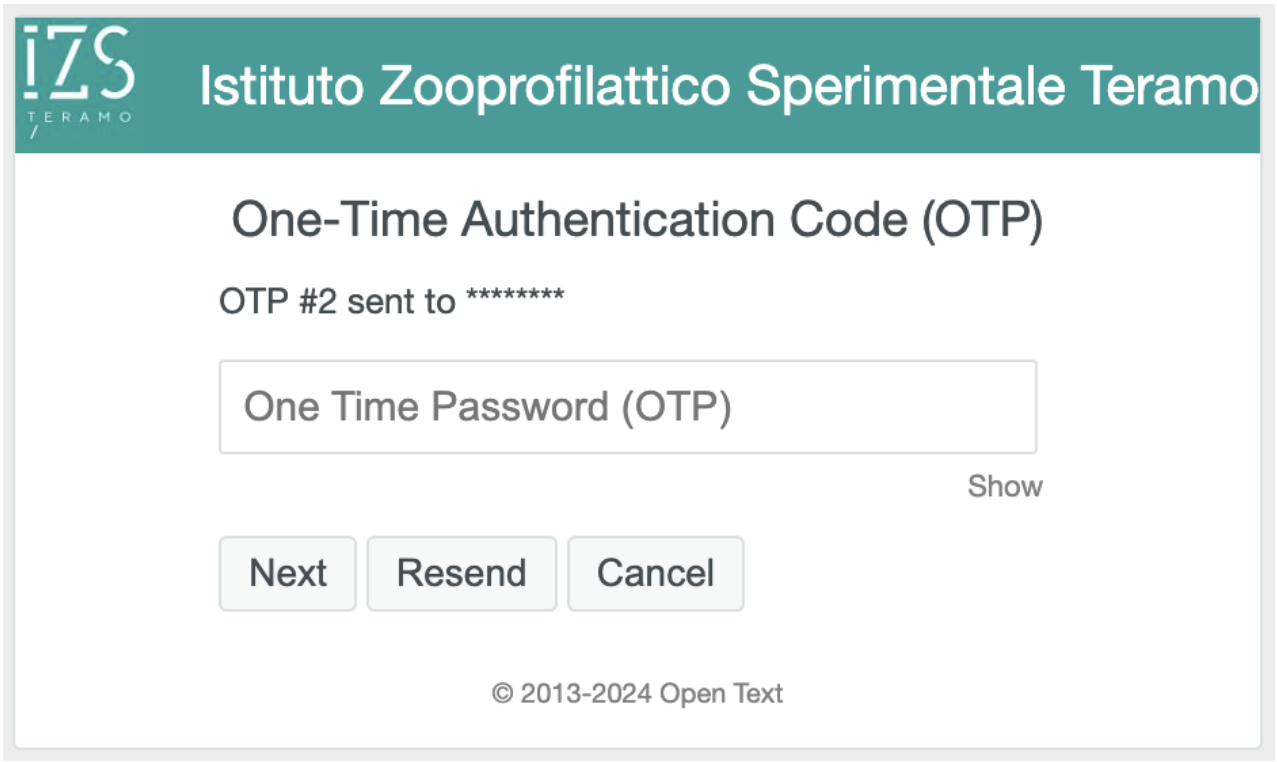
Seguendo la voce sopra menzionata `Cambia password` si aprirà una finestra modale che ti informerà brevemente sui passaggi necessari. Cliccando su `Procedi` verrai reindirizzatə direttamente al servizio esterno https://portale.izs.it/sspr/public/forgottenpassword.



Servizio cambio password

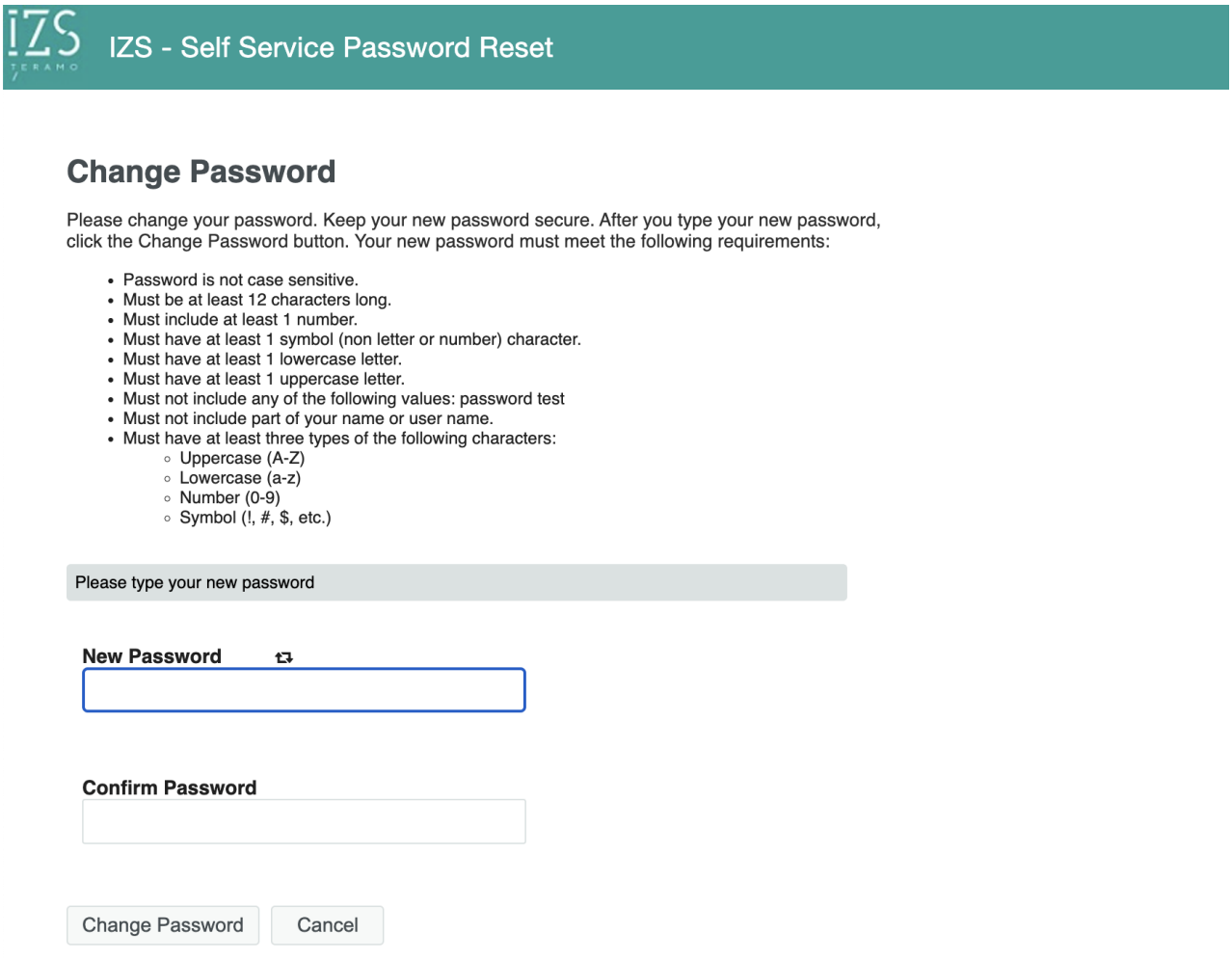
Suggerimenti Puoi cambiare la **lingua** tramite l'incona del globo in alto a destra. Il titolo e la descrizione della pagina potrebbero fare riferimento alle procedure di **password dimenticata**, ma non preoccuparti: si tratta dello stesso flusso utilizzato anche per il cambio password.

Ti verrà richiesto di inserire il tuo **nome utente** (lo stesso usato per accedere alla piattaforma). Dopo una verifica, riceverai una **email** all'indirizzo associato al tuo account, contenente un **codice OTP** (*One Time Password*).



Inserimento codice OTP

Una volta inserito il codice OTP, potrai impostare una **nuova password**.



Inserimento codice OTP

Il reset potrebbe richiedere alcuni secondi. Una **barra di avanzamento** ti mostrerà lo stato dell'operazione; al termine, potrai tornare in piattaforma.

14.1.2 Password dimenticata

Se hai dimenticato la password, clicca sul link **Ho dimenticato la password** sotto la maschera di login per avviare la procedura di recupero.

I passaggi sono identici a quelli descritti nella sezione Procedure di cambio password.

14.1.3 Caratteristiche della password

Per assicurare un ragionevole livello di sicurezza, la password dovrà necessariamente avere le seguenti caratteristiche:

- **Lunghezza minima**: Minimo 12 caratteri (sarebbe meglio utilizzare password complesse di 15 o più caratteri);
- **Unicità**: il sistema non consente di impostare una password uguale a quelle utilizzate in precedenza.
- **Complessità**: la password deve soddisfare **almeno tre** dei seguenti quattro criteri:
- maiuscola (A-Z)
- minuscola (a-z)
- numero (0-9)
- simbolo (! , #, $, ecc.)
- caratteri di altre lingue non elencati sopra
- **NON** deve includere parte del nome proprio o del nome utente.

Consiglio finale: utilizza sempre password lunghe, uniche e non riconducibili a dati personali. Valuta l'uso di un gestore di password sicuro per memorizzarle.

14.2 Help desk

Per richieste di supporto è possibile utilizzare le funzioni di **Help desk**, accessibili tramite la relativa voce presente nel menu contestuale del wiki, posto in alto a destra della piattaforma. Al click si aprirà una modale che consentirà, senza dover abbandonare l'applicativo:

- di selezionare l'area problematica di riferimento;
- di descrivere l'eventuale criticità, anche attraverso un'immagine o un video di max. 25MB (formati accettati: .png, .jpg, .jpeg, .mp4).

Il video che segue mostra la procedura:

Help Desk

14.3 Video Utili

In questa sezione sono messi a disposizione dell'utente video utili per la formazione o l'utilizzo generale della piattaforma. I video presenti potrebbero far riferimento ad eventi o attività di formazione istituite nell'ambito di altri progetti. La videoteca digitale verrà nel tempo accresciuta e categorizzata, con l'aggiunta di ulteriori video.

14.3.1 NGS & Bioinformatics - ERFAN

Video realizzati nell'ambito del progetto ERFAN

Lesson 1 - 22/02/2022

NGS Bioinformatics Lesson - 1

Lesson 2 - 01/03/2022

NGS Bioinformatics Lesson - 2

14.4 Scorciatoie da tastiera

Le funzioni principali possono essere lanciate tramite tastiera attraverso particolari combinazioni predefinite. Alcune combinazioni possono funzionare ad un livello generale dell'applicativo oppure solo all'interno di determinate pagine. Di seguito, proponiamo una lista delle scorciatoie definite di default.

Importante

La lista è in continuo aggiornamento e, pertanto, ti invitiamo a visitare spesso questa pagina.

14.4.1 Lista comandi

Funzione	Combinazione	Pagine
Abilita/disabilita la selezione multipla	`ctrl+shift+m`	Classi e Viste (pagine con griglia selezionabile)
Seleziona tutto	`ctrl+shift+a`	Classi e Viste (pagine con griglia selezionabile)
Svuota carrello	`alt+shift+e`	Tutto
Aggiungi al carrello	`alt+shift+a`	Pagine dove è presente il pulsante *Aggiungi al carrello*
Sostituisci carrello	`alt+shift+r`	Pagine dove è presente il pulsante *Aggiungi al carrello*
Rimuovi da carrello	`alt+shift+d`	Pagine dove è presente il pulsante *Aggiungi al carrello*
Attiva/Disattiva carrello	`alt+shift+c`	Tutto
Disattiva tag	`alt+shift+t`	Tutto

14.5 FAQ

In questa pagina sono elencate le domande frequenti (FAQ) riguardo l'uso della piattaforma e le corrispondenti risposte. Prima di contattare l'assistenza, per favore controllate se il vostro problema può essere risolto leggendo una delle risposte in questa pagina.

14.5.1 Gestione dati

Come cambio il carattere usato come separatore delle colonne in Excel?

Risposta: Il separatore delle colonne può essere specificato a diversi livelli. Qui sotto riportiamo delle guide per i casi principali:

- Per poter modificare il separatore predefinto delle colonne in Excel è necessario cambiare il carattere da usare come separatore tra le opzioni di sistema Windows. La seguente guida esterna illustra come fare: https://stolasinformatica.eu/guida/cambiare-separatore-elenco-csv/;
- Per importare in Excel un file con un separatore differente, si può fare riferimento a questa guida: https://www.excelpertutti.com/da-csv-a-excel-convertire-csv-in-excel/;
- Per dividere le righe di un file di testo usando un altro carattere, sfruttando le funzionalità interne di Excel, è possibile seguire questa guida: https://it.extendoffice.com/documents/excel/1786-excel-split-text-by-space.html;
- Se invece si vuole convertire un file si può usare uno strumento online come quello disponibile a questo link: https://onlinecsvtools.com/change-csv-delimiter.

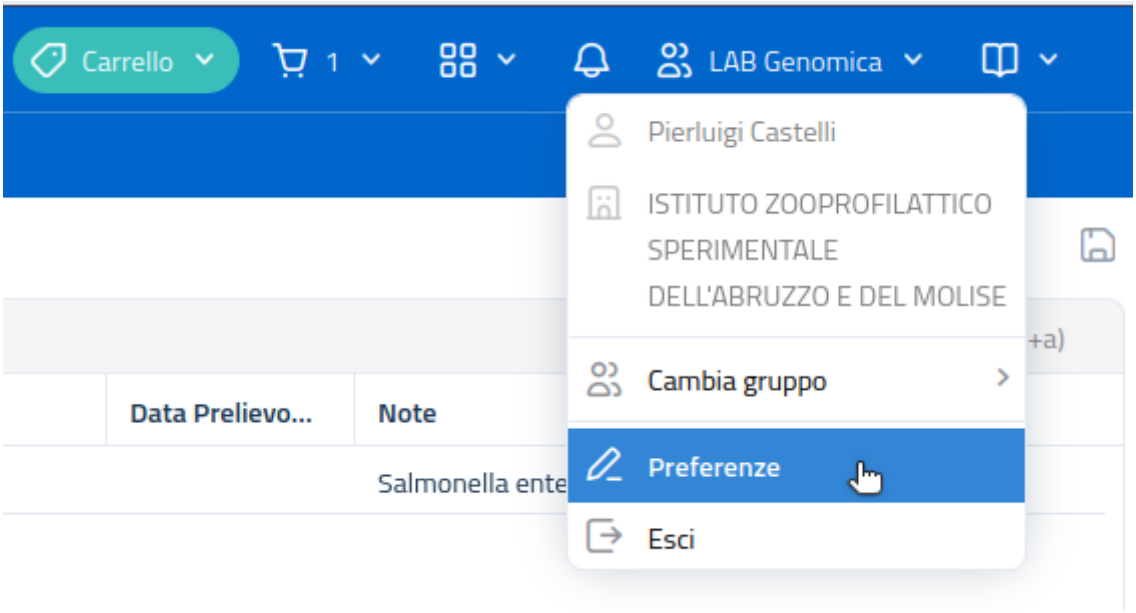
Nota: I files `.csv` che vengono creati per la piattaforma devono avere come separatore di elenco il punto e virgola (`;`), mentre il file che viene creato per il samplesheet deve usare la virgola (`,`) come separatore.

Non riesco ad esportare una tabella come CSV.

Risposta: L'esportazione in CSV di una tabella della piattaforma viene gestita mediante una finestra popup. Controllare che il browser abbia aperto altre finestre e, in caso contrario, verificare che il browser non abbia bloccato la visualizzazione di popup del sito. Normalmente quasi tutti i browsers visualizzano un piccolo banner in alto nella pagina, nel quale si chiede all'utente come gestire la richiesta di visualizzazione di un popup.

Si può cambiare il carattere utilizzato come separatore di colonna per l'esportazione dei files CSV?

Risposta: Sì, in piattaforma è possibile modificare il separatore di colonna predefinito, usato per l'esportazione in CSV, nella sezione **Preferenze** del menù utente: una volta entrati nelle preferenze, si può modificare il carattere usato come separatore nella voce "**Separatore CSV**". Una volta salvate le modifiche, da quel momento in avanti, quando si esporteranno liste di campioni come CSV, verrà usato il carattere selezionato come separatore.



14.5.2 Carrello e tag

Ricevo un messaggio di errore quando tento di aggiungere al carrello molti campioni.

Risposta: Il carrello può contenere un massimo di 2000 campioni. In caso si debba lavorare su un numero superiore, suggeriamo di creare un tag/motivo.

Come faccio a duplicare/clonare un tag?

Risposta: Attiva il tag da duplicare, raggiungi la sezione `Esplora campioni` seleziona e aggiungi i campioni al carrello e crea un nuovo tag a partire dal carrello.

14.5.3 Interfaccia

Come posso visualizzare colonne con ulteriori informazioni nelle tabelle dei campioni?

Risposta: Ulteriori colonne, comprese quelle contenenti informazioni sui risultati (come specie calcolata, quality check, N50, PFGE, percentuale di alleli chiamati, *etc.*) possono essere visualizzate come descritto nella pagina Wiki "Funzionalità tabelle".

Ulteriori informazioni, non disponibili direttamente nella tabella dei campioni, si possono trovare nella pagina dei "Risultati Esami" (dal menù di Navigazione laterale, `Elementi principali > Esami > Risultati Esami`).

14.5.4 Controllo e visualizzazione dati

Dove trovo i check di un campione?

Risposta: Nel menù di Navigazione laterale sono presenti le tabelle di Check: `Reports > Dettaglio > Quality Check Principale` e `Reports > Dettaglio > Checks`. Usando un filtro (carrello attivo o tag/motivo) si possono filtrare le righe delle tabelle per risalire al/ai campione/i di interesse. Le pagine di Check sono descritte nella relativa pagina Wiki "Pagine di Check".

Dove posso trovare le informazioni sulle raw reads di un campione?

Risposta: Le informazioni sulle raw reads sono visualizzabili nella tabella "Check Reads" (`Reports > Dettaglio > Checks > Check Reads`). Inoltre, come per i risultati delle analisi, le raw reads dei campioni sono scaricabili tramite le funzionalità di Download. In questo caso il tasto da selezionare è "0sq_rawreads".

14.5.5 Analisi

Come mai GrapeTree non riesce a calcolare il Minimum Spanning Tree?

Risposta: Di solito la causa è che non ci sono alleli di differenza tra i campioni analizzati: GrapeTree non riesce a creare il MST su un dataset con più campioni aventi 0 differenze. Un'altra possibile causa è il numero di campioni: GrapeTree non è in grado di calcolare un albero se si usano solo 2 campioni o meno.

Sui campioni esterni posso lanciare GrapeTree o devo lanciare prima ChewBBACA?

Risposta: GrapeTree richiede come input i profili allelici da ChewBBACA (o altro tool disponibile per **4TY_cgMLST**); perchè i profili siano utilizzabili, i risultati di ChewBBACA devono essere stati importati. Sui campioni esterni non vengono lanciate analisi automaticamente (al contrario dei campioni interni), quando essi vengono aggiunti alla piattaforma. Prima di lanciare GrapeTree su campioni esterni bisogna assicurarsi che ChewBBACA sia stato lanciato ed i relativi risultati importati. La tabella `Elementi principali > Esami > Risultati Esami` permette di verificare che **4TY_cgMLST** sia stato eseguito.

Cosa succede se non conosco l'ospite del mio campione o non voglio farne la deplezione?

Risposta: Se non si desidera (o non è possibile) eseguire il processo di deplezione dell'ospite, si può lanciare direttamente il mapping. I softwares per la ricostruzione del genoma funzioneranno ugualmente.

> **ATTENZIONE:** Questo approccio è accettabile unicamente se il campione è di ottima qualità e si è certi di ottenere buoni risultati, altrimenti i dati saranno inutilizzabili e la piattaforma accumulerà dati di cattiva qualità, danneggiando le analisi future degli utenti.

Come faccio se non conosco il reference migliore per il mio genoma virale?

Risposta: L'analisi "3TX_species - identificazione di specie" può essere lanciata usando il software "vdabricate", che identifica il reference migliore nel DB_Virus.

Come posso ottenere informazioni su programmi/opzioni di analisi che sono state svolte automaticamente o da altri utenti del mio gruppo?

Risposta: Fintanto che un'analisi è stata eseguita ed importata, i risultati (report e QC compresi) sono scaricabili dalla pagina Download (Wiki Download), inoltre i files "Logs / cfg" sono disponibili per ogni analisi. Tuttavia si tenga a mente che il sistema di download prepara sempre i risultati dell'analisi più recente con le caratteristiche specificate. Nel caso in cui si desideri investigare l'escuzione di analisi meno recenti, si può usare la funzione di download tramite RIS_CD.

Non vedo le analisi di altri utenti/analisi lanciate automaticamente in “Controllo analisi”. Non mi sono accessibili?

Risposta: Le schede delle analisi lanciate da altri utenti non possono essere visualizzate in "Controllo analisi": solo gli amministratori possono vedere tutte le analisi lanciate sul sistema. I risultati di quelle analisi sono comunque disponibili per il download, quindi si ha in ogni caso accesso ai risultati, ai report ed ai QC, compresi i dati visualizzati nelle tabelle di Check.

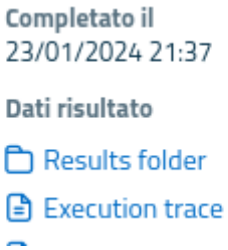
Nel caso dei campioni siano stati analizzati più di una volta con gli stessi softwares, come faccio a visualizzare solo le informazioni relative ad una determinata esecuzione?

Risposta: Ogni analisi lanciata possiede un codice univoco. In "Risultati Esami" è possibile filtrare analisi eseguite su campioni con lo stesso NRG ma DS diversi, sfruttando il campo "Cerca" o i filtri nella colonna "Codice". Se si conosce il codice sarà sufficiente eseguire una ricerca per quella stringa di testo, altrimenti è possibile risalirvi in base alla data di esecuzione (ad esempio cercando o filtrando per "210923" si ottengono le analisi lanciate 23 settembre 2021; analogamente si può specificare un'ora: "240208152038" sta per l'8 febbraio 2024 alle 15.20 e 38 secondi).

È possibile vedere con che programmi/opzioni sono state svolte le analisi?

Risposta: Certamente. Nella cartella di ogni analisi sono conservati i log ed un file ".cfg", il quale corrisponde allo script Bash eseguito per lanciare il tool, con tutti i parametri indicati. Di conseguenza i programmi, le loro versioni, le versioni Docker ed i comandi ed opzioni usati per il lancio dell'analisi vengono sempre conservati e sono consultabili in diversi modi:

- Per le analisi eseguite ed accessibili nella pagina "Controllo analisi", è possibile ottenere informazioni su quanto lanciato leggendo il file "Execution trace", disponibile nella scheda dell'analisi completata e che contiene la versione dei vari tool eseguiti:



- Per le analisi importate nel sistema, ovvero presenti in "Risultati Esami", è possibile scaricarne i risultati ed indicare tra le azioni di download anche "Logs / cfg". Aprendo la cartella per il download dei risultati dell'analisi, nella cartella "logs" sarà presente il file di configurazione, contenente lo script Bash eseguito ed i parametri.

- I tools (con relative versioni) che concorrono al calcolo dei profili allelici sono anche descritti nella pagina "NGSManagerSOP".

14.5.6 Gestione profilo

Come faccio a cambiare profilo?

Risposta: Il profilo si cambia cliccando sul nome del profilo corrente (in alto a destra) e poi su "Cambia profilo". La procedura è anche illustrata nella videoguida "Profili utente".

Come faccio a cambiare la mia password?

Risposta: Trovi una descrizione delle procedure di cambio password nella pagina Gestione password.

Piattaforma GenPat - Manuale d'uso