

Orodja za delo z delotoki

dr. Marko VIDAK, Medicinska fakulteta Univerze v Ljubljani

dr. Bojan KVERH, Medicinska fakulteta Univerze v Ljubljani

dr. Polonca FERK, Medicinska fakulteta Univerze v Ljubljani

dr. Brane LESKOŠEK, Medicinska fakulteta Univerze v Ljubljani

dr. Maja Zagorščak, Nacionalni inštitut za biologijo

dr. Nadja Nolte, Nacionalni inštitut za biologijo

dr. Carissa Bleker, Nacionalni inštitut za biologijo

dr. Marko Petek, Nacionalni inštitut za biologijo

dr. Anže Županič, Nacionalni inštitut za biologijo

Ljubljana in online, Centralna tehniška knjižnica Univerze v Ljubljani,

Usposabljanje podatkovnih strokovnjakov, 18. – 21. 11. 2024



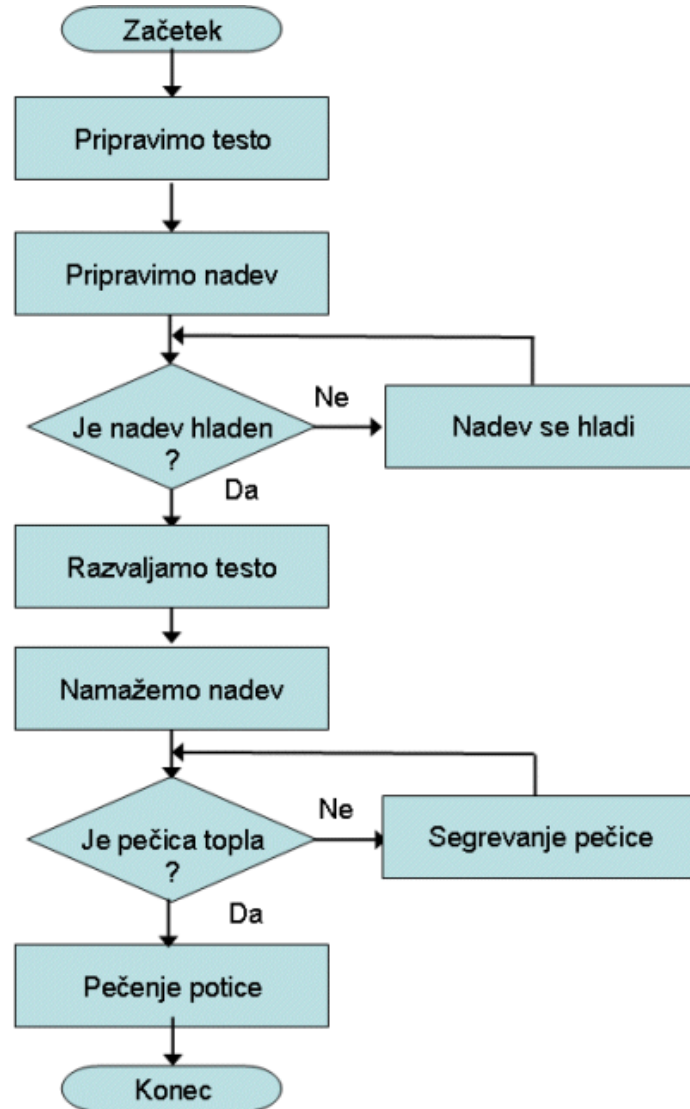
Vsebina

- Uporabljivost (*Reusability*)
- Virtualizacija
- Okolja za delo z delotoki v ukazni vrstici
- Predstavitev okolja Galaxy za delo z delotoki preko grafičnega vmesnika
- Primer uporabe delotoka v Galaxy – računanje podobnosti med aminokislinskimi zaporedji proteinov

Uporabljivost (Reusability)

- Vsaka odprta programska oprema mora biti opremljena z bogato dokumentacijo
- Rezultati zaganjanja morajo biti ponovljivi, ne glede na to, na kakšnem računalniku ali operacijskem sistemu je programska oprema zagnana
- Težave z odvisnostjo od različnih zunanjih knjižnic je mogoče rešiti preko ustvarjanja izoliranih okolij, ki vsebujejo vse odvisne komponente ali stabilne povezave do njih

Diagram delotoka (peka potice)

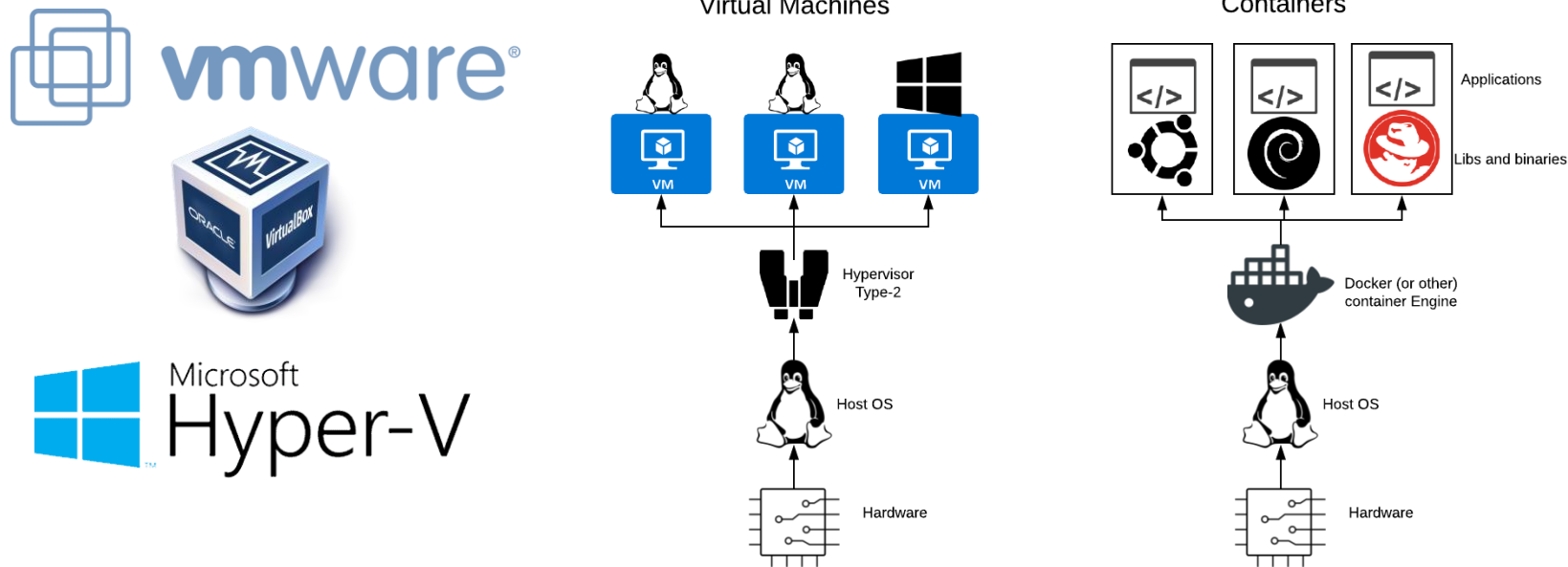


Delotoki v analizi podatkov

- Delotok (angl. *pipeline/workflow*) je zaporedje korakov v procesu analize.
- Osnovni gradnik delotoka je vozlišče - posamezen korak.
- Delotok mora imeti zagotovljen dostop do:
 - orodij, ki so potrebna za izvedbo posameznega koraka,
 - potrebnih vhodnih podatkovnih virov,
 - potrebnih izhodnih podatkovnih virov.

Virtualizacija

- nadomeščanje strojne opreme z navidezno
- virtualizacija računalnikov (**Virtual Machine - VM**) nam omogoča da na enem računalniku ali strežniku istočasno deluje več različnih sistemov, ki so med seboj neodvisni in imajo pripadajočo navidezno strojno opremo



* hipervizor - nadzornik virtualnega stroja (VMM) - virtualizator

Vsebnik (*container*) in okolje (*enviroment*)

- skupina procesov z dovolj veliko mero izolacije
- 'lahka virtualizacija'
- izolacija programov v ločenih imenskih prostorih
- omogoča konsistenco izvajanja poslov: programska oprema + pripadajoče knjižnice + sistemske komponente -> možnost ponovitve eksperimentov in izvajanja poslov tudi v spremenjenem in posodobljenem okolju



INSTALLING APOLLO
Setup guide
Pre-requisites
Install node and yarn
Install jdk
Database configuration
Database schema
Deploy the application
Detailed build instructions
Using Docker to Run Apollo
Apollo Configuration
Chado Export Configuration
Data generation pipeline
Troubleshooting guide
Example Build Script on Unix with MySQL
Adding OpenID Connect Authentication to Apollo
Migration guide
USING APOLLO
Demo
User's Guide

INSTALLING APOLLO
Setup guide
Using Docker to Run Apollo
Running the Container
Production
Apollo Configuration
Chado Export Configuration
Data generation pipeline
Troubleshooting guide
Example Build Script on Unix with MySQL
Adding OpenID Connect Authentication to Apollo
Migration guide
USING APOLLO
Demo

Ubuntu / Debian

```
sudo apt-get install zlib1g zlib1g-dev libexpat1-dev libpng-dev libgd2-noxpm-dev build-essential
```

RedHat / CentOS

```
sudo apt-get install zlib zlib-dev expat-dev libpng-dev libgd2-noxpm-dev build-essential git python
```

It is recommended to use the [default version of JBrowse](#) or [better](#) (though it does not work with JBrowse 2 yet).

There are [additional requirements](#) if doing development with Apollo.

Install node and yarn

Node versions 6-12 have been tested and work. `npm` and `npm install 8` is recommended.`

```
npm install -g yarn
```

Install jdk

Build settings for Apollo specifically. Recent versions of tomcat7 will work, though tomcat 8 and 9 are preferred. If it does not install automatically there are a number of ways to [build tomcat on linux](#):

Using Docker to Run Apollo

You can [install Docker](#) for your system if not previously done.

Running the Container

You can see Apollo straight away:

```
docker run -it -p 8888:8080 gmod/apollo:stable -v /jbrowse/root/directory:/data
```

Open <http://localhost:8888> in a web browser and login with `admin@local.host` / password to get started.

Note: data is not guaranteed to be saved in this manner, but data is `/jbrowse/root/directory` will not be written to either.

Production

To run in production against persistent JBrowse data and a persistent database you should:

Apollo

urejevalnik anotacij genoma, sodelovanje v realnem času

Home Downloads Contribute

What is UniTato?

UniTato is a web-based service that provides access to comprehensive and accurate potato gene models. It unifies the recent potato double monoploid (DM) v4 and v6 gene models of the [Buell Lab's Spud DB](#), and allows further community-based manual curation in a user-friendly web interface powered by Apollo, a genome annotation and editing platform.

How to get started?

UniTato is free and easy to use. You can access the Apollo interface and start exploring the potato genome immediately:
→ [Access the public UniTato Apollo interface](#).

To start contributing manual annotations, please see:
→ [Contribute](#).

If you already have an account:
→ [Login to UniTato Apollo interface](#).

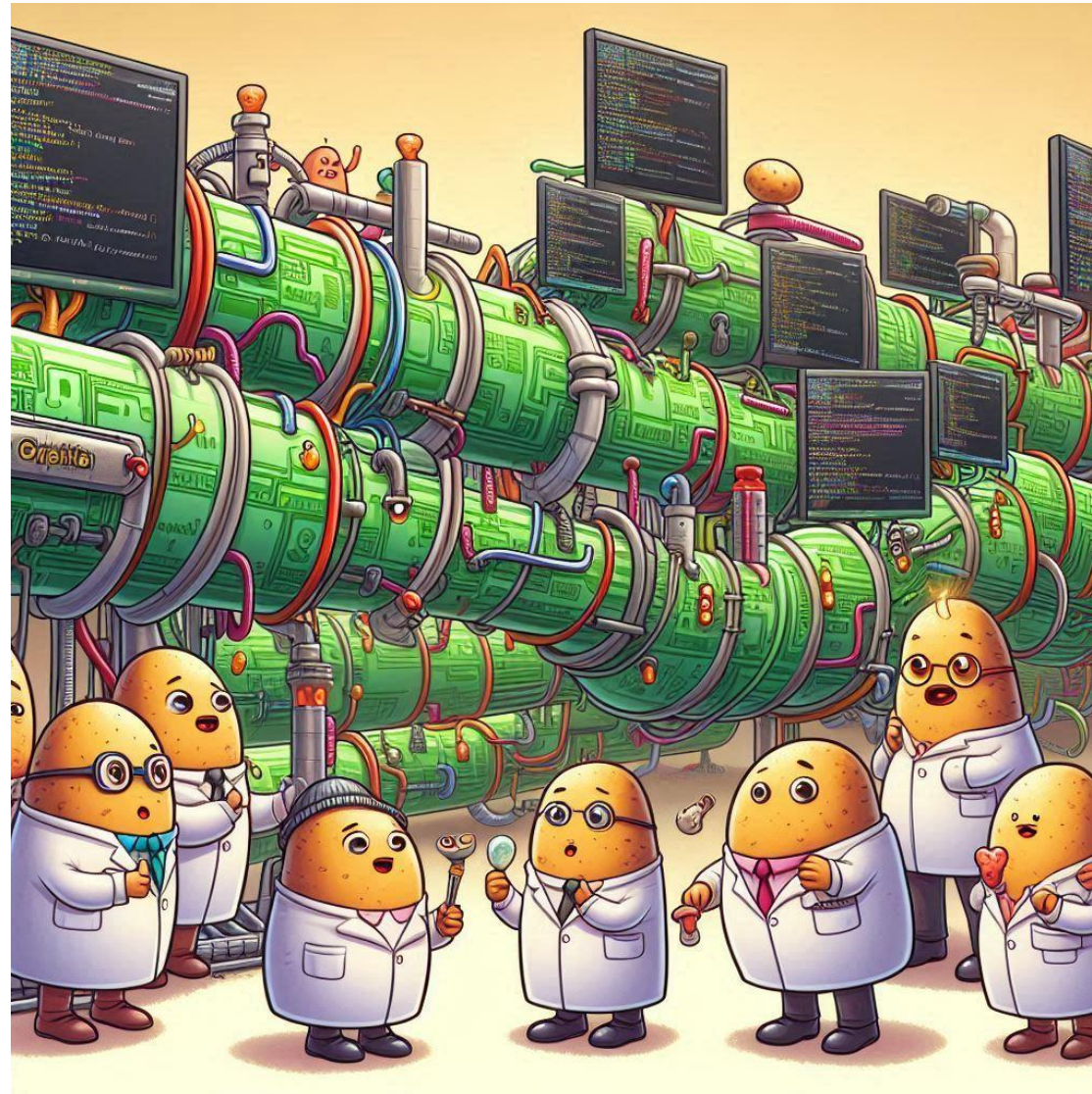
Why use UniTato?

- UniTato contains the most up-to-date and unified potato gene models, which can help you avoid missing or wrongly annotated genes.
- UniTato allows you to browse, search, and download manually curated potato gene models.

UniTato tracks

The following tracks are available for exploration in Apollo:

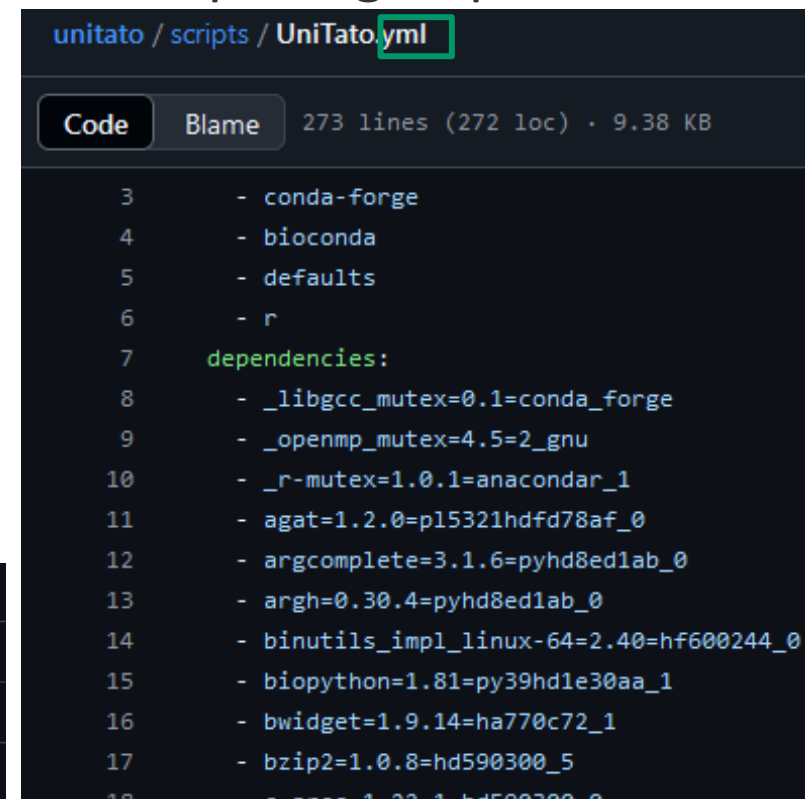
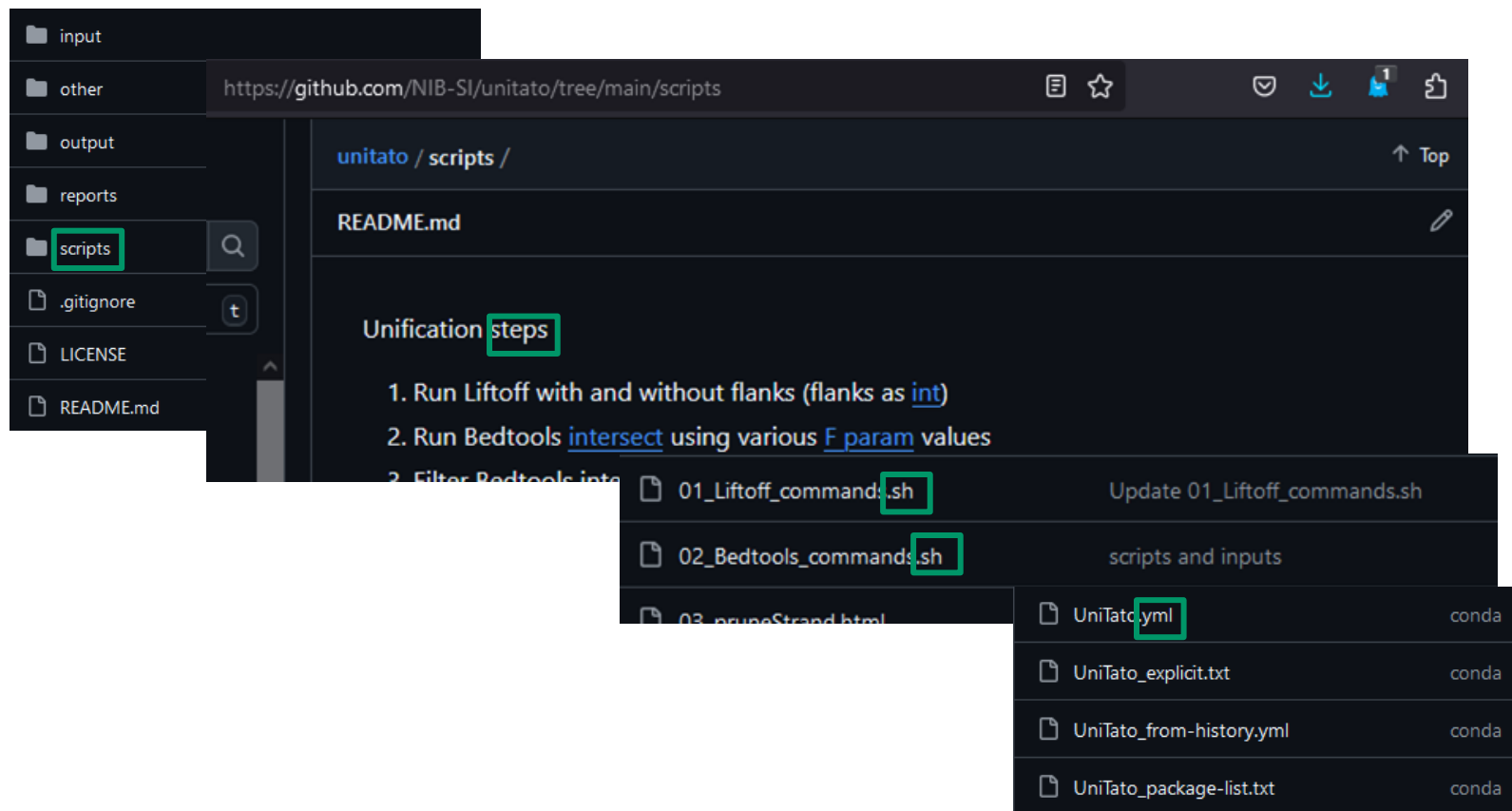
Okolja za delo z delotoki v ukazni vrstici



MINIMUM: uporaba sistemov za upravljanje s paketi in orodji



ustvarjanje okolja Conda
na podlagi zapisa YAML



OPTIMUM - upravitelji (*workflow managers*) pri pisanju ponovljivih delotokov

- Ponovljivi delotoki - **FAIR**ifikacija
- Nadzor različic in vsebnikov
- Odprta koda, strukturiranost, enostavno vzporedno delovanje
- Povečanje obsega za delovanje, npr. HPC
- Rastoča skupnost, standardizirani bioinfo delotoki



Jezik (language)	Python	Groovy
Podatki (data)	datoteke (<i>everything is a file</i>)	datoteke in vrednosti (<i>can use both files and values</i>)
Izvajanje (execution)	delovni direktorij (<i>working directory</i>)	več direktorijev (<i>each job in its own directory</i>)
'Dry runs'	+	-
Spremljanje spremembe kode	-	+

github.com/NIB-SI/ARMOR/blob/master/Snakefile

ARMOR / Snakefile

Code Blame Executable File · 665 lines (615 loc) · 27.3 KB

```
5
6  ## Configuration file
7  if len(config) == 0:
8      if Path("config.yaml").exists():
9          configfile: Path("config.yaml")
10     else:
11         raise FileNotFoundError("Make sure there is a config.yaml file in {cwd}"
12                                " or specify one with the --configfile cmdline parameter.")
13
14  # Make sure that all expected variables from the config file are in the config
15  # dictionary
16  configVars = [
17      'annotation', 'organism', 'build', 'release', 'salmonk', 'readlength',
18      'fldMean', 'fldSD', 'design', 'contrast', 'genesets',
19      'ncores', 'fqext1', 'fqext2', 'fqsuffix', 'output',
20      'useCondaR', 'Rbin', 'run_trimming', 'run_STAR', 'run_DRIMSeq',
21      'run_camera'
22  ]
23  fileVars = [
24      'txome', 'genome', 'gtf', 'salmonindex', 'STARindex', 'metatxt', 'FASTQ'
25  ]
```

<https://github.com/NIB-SI/ARMOR/blob/master/README-NIB-SI.md>

```
64  ## ----- ##
65  ## Target definitions
66  ## ----- ##
67  ## Run all analyses
68  rule all:
69      input:
70          outputdir / "MultiQC" / "multiqc_report.html",
71          outputdir / "outputR" / "shiny_sce.rds"
72
73  rule setup:
74      input:
75          outputdir / "Rout" / "pkginstall_state.txt",
76          outputdir / "Rout" / "softwareversions.done"
77
78  ## FastQC on original (untrimmed) files
79  rule runfastqc:
80      input:
81          expand(outputdir / "FastQC/{sample}_" f"{config['fqext1']}_fastqc.zip", sample = samples.names[samples.type
82          expand(outputdir / "FastQC/{sample}_" f"{config['fqext2']}_fastqc.zip", sample = samples.names[samples.type
83          expand(outputdir / "FastQC/{sample}_fastqc.zip", sample = samples.names[samples.type == 'SE'].values.tolist(
```

my-first-nextflow-pipeline / main-STAR-params.nf

nadjano correct order 034eb90 · 10 months ago History

Code Blame 89 lines (63 loc) · 1.67 KB

```
1  #!/usr/bin/env nextflow
2  params.genomeFasta = "$baseDir/test_RNA_seq/ggal_1_48850000_49020000.Ggal171.500bpflank.fa"
3  params.publish_dir="$baseDir/results"
4
5
6  genomeFasta = Channel.fromPath(params.genomeFasta)
7
8  // from File Pairs
9  fastqFiles = Channel
10     .fromFilePairs('./test_RNA_seq/ggal_*_{1,2}.fq')
11     .view()
```

```
script:
"""
STAR --genomeDir STAR_index \
--readFilesIn $read1 $read2 \
--outFileNamePrefix ${sample_name}_intron_len_${intronMax}/ \
```

<https://github.com/nadjano/my-first-nextflow-pipeline>

```
/*
 * Define the workflow
 */
workflow {

    star_index = build_STAR_index(genomeFasta)

    index_with_reads = star_index.combine(fastqFiles_flat)

    index_with_reads.view()

    // Lets add some mapping parameters

    intron_max_length = [100, 1000]

    run_STAR_alignment(index_with_reads, intron_max_length)

}
```



Okolje Galaxy

- Spletno okolje za **uporabo bioinformacijskih orodij in izvajanje delotokov** preko **grafičnega vmesnika** (*Graphical User Interface* - GUI).
- Omogoča uporabo številnih bioinformatičnih orodij **brez poznavanja ukazne vrstice**.
- Galaxy je sicer nastal na področju ved o življenju vendar se sedaj uporablja tudi na področju drugih ved.

Orodje brez grafičnega vmesnika ...

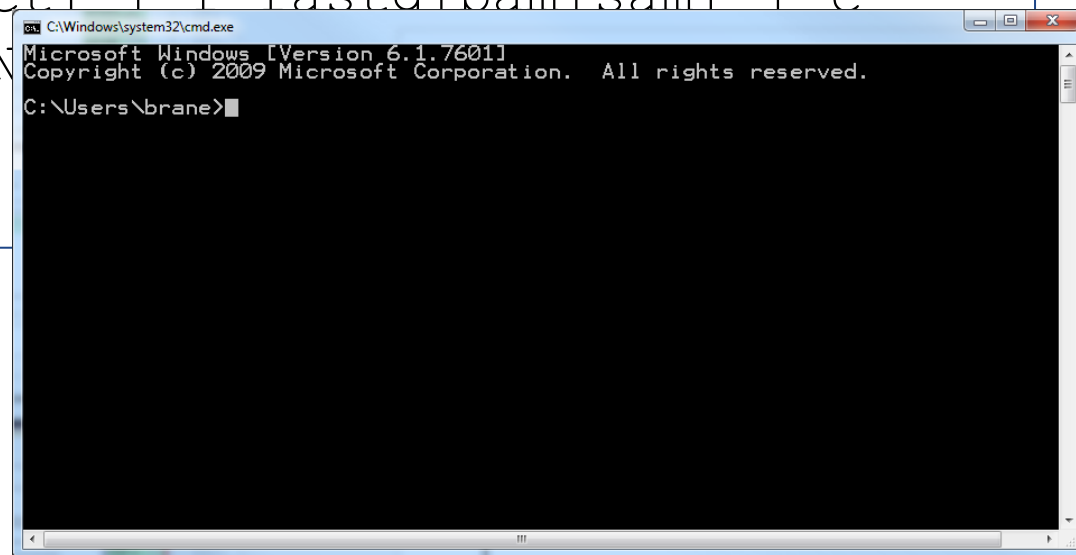
```
$ fastqc --help
```

FastQC - A high throughput sequence QC analysis tool

SYNOPSIS

```
fastqc seqfile1 seqfile2 .. seqfileN
```

```
fastqc [-o output dir] [--(no)extract] [-f fastq|bam|sam] [-c  
contaminant file] seqfile1 .. seqfileN
```



... in z grafičnim vmesnikom Galaxy:

The screenshot displays the Galaxy web interface with three main panels:

- Tools Panel (Left):** A sidebar with a search bar and a list of tool categories including Get Data, Send Data, Lift-Over, Text Manipulation, Convert Formats, Filter and Sort, Join, Subtract and Group, and NGS: QC and manipulation. Under NGS: QC and manipulation, the 'FastQC Read Quality reports' tool is highlighted.
- FastQC Read Quality reports (Galaxy Tool Version 0.63) Panel (Center):** The main configuration area for the FastQC tool. It includes fields for 'Short read data from your current history' (set to '15: child-mother.bam'), 'Contaminant list' (set to 'No selection'), and 'Submodule and Limit specifying file' (set to 'No selection'). An 'Execute' button is at the bottom. Below the configuration, there is a 'Purpose' section explaining that FastQC aims to provide a simple way to do quality control checks on raw sequence data. It also lists the main functions of FastQC: Import of data from BAM, SAM or FastQ files; Providing a quick overview to tell you in which areas there may be problems; Summary graphs and tables to quickly assess your data; Export of results to an HTML based permanent report; and Offline operation to allow automated generation of reports without running the interactive application.
- History Panel (Right):** A list of previous jobs. The top job is '0: converted BAM'. Below it, several jobs are listed, including '1: SAM-to-BAM on data', '0: converted BAM', '0: Filter SAM on data 8', '0: Filter SAM on data 7', '0: Map with BWA for Illumina on data 4 and data 3: mapped reads', '0: Map with BWA for Illumina on data 2 and data 1: mapped reads', '0: FastQC imported: raw_child-ds-2.fq.html', '0: FastQC imported: raw_child-ds-1.fq.html', '0: FastQC imported: raw_mother-ds-2.fq', '0: FastQC imported: raw_mother-ds-1.fq', '0: FastQC imported: raw_child-ds-2.fq', and '0: FastQC imported: raw_child-ds-1.fq'. Each job entry has icons for viewing, editing, and deleting.

Orodja

Osrednje okno: nastavitve, izvajanje, pregled rezultatov

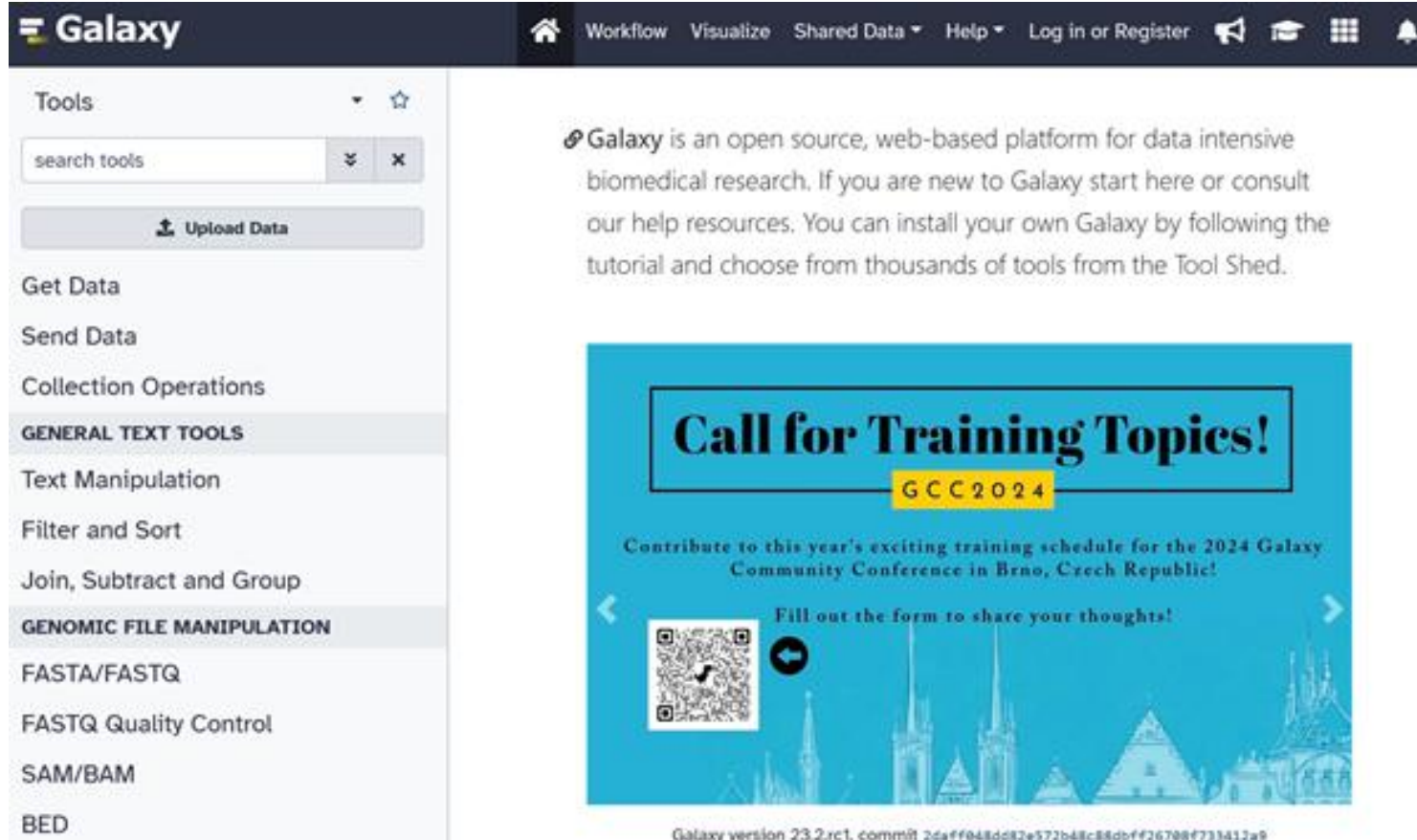
Zgodovina

Različice okolja Galaxy

- Obstaja več odprtih sistemov Galaxy.
- Najpomembnejše so:
 - **UseGalaxy.org** - ameriški strežnik Galaxy
 - **UseGalaxy.org.au** - avstralski strežnik UseGalaxy
 - **UseGalaxy.eu** - evropski strežnik Galaxy (Freiburg, DE)
- Lokalizirane različice: Vsak si Galaxy lahko namesti sam. Izziv je vzdrževanje ter povezava z lokalno gručo in arhivom, npr. Galaxy na ELIXIR-SI HPC.

Galaxy ZDA

- <https://usegalaxy.org>



The screenshot shows the Galaxy web interface. The top navigation bar includes the Galaxy logo, a home icon, and links for Workflow, Visualize, Shared Data, Help, Log in or Register, and notification icons. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button, followed by categories like 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', 'FASTQ Quality Control', 'SAM/BAM', and 'BED'. The main content area features a welcome message: 'Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.' Below this is a blue banner for the 'Call for Training Topics!' for GCC 2024, with a QR code and a link to fill out a form. The banner also includes the text 'Contribute to this year's exciting training schedule for the 2024 Galaxy Community Conference in Brno, Czech Republic!' and a small image of a city skyline. At the bottom, the version information is displayed: 'Galaxy version 23.2.rc1, commit 2da9ff048dd82e572b48c88dbff26798f733412a9'.



Galaxy Avstralija

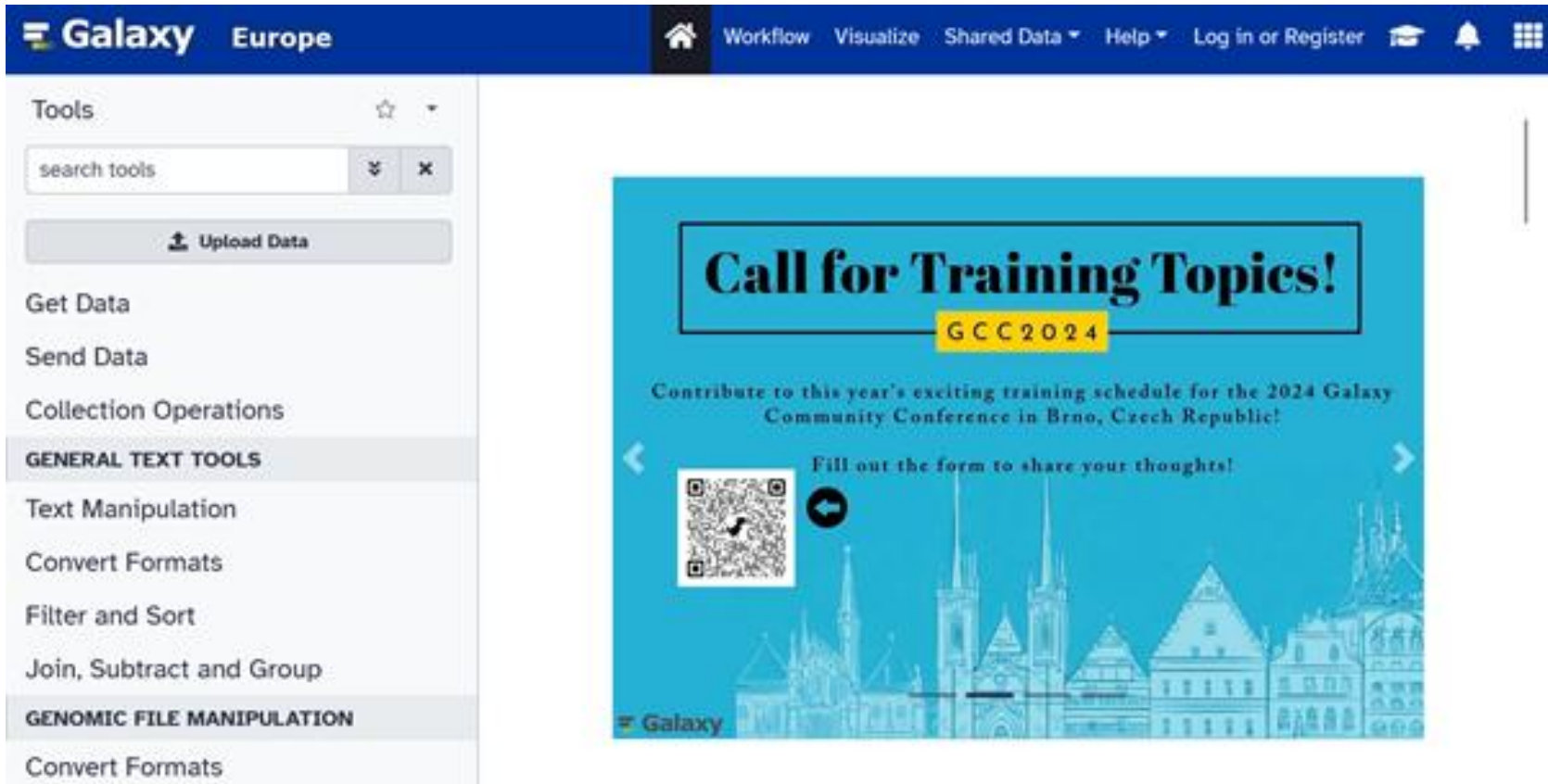
- <https://usegalaxy.org.au>

The screenshot shows the Galaxy Australia website. The top navigation bar includes the Galaxy logo, the word "Australia", and links for Workflow, Visualize, Shared Data, Help, Log in or Register, and icons for a graduation cap, a network, and a notification. A left sidebar contains a "Tools" section with a search bar and an "Upload Data" button, followed by categories: "FILE AND META TOOLS" (Get Data, Send Data, Collection Operations), "GENERAL TEXT TOOLS" (Text Manipulation, Filter and Sort, Join, Subtract and Group), and "GENOMIC FILE MANIPULATION" (FASTA/FASTQ, FASTQ Quality Control, SAM/BAM). The main content area features a light blue banner with a flask icon and the text "The Genome Lab has just launched! This corner of the Galaxy is dedicated t...". Below this is a navigation menu with links: Home, News, People, About, Support, Docs. The central logo reads "Galaxy AUSTRALIA". A section titled "Thanks for making the Galaxy Community Conference great!" displays statistics for GCC23 Brisbane: 91 in-person participants, 40 virtual attendees, 20 countries represented, and many koalas hugged! It also lists 316 authors, 48 talks, 49 posters, and 14 training workshops. A world map highlights Australia and other participating countries. The footer includes logos for BizData, Limagrain, GalaxyWorks, GIGAF SCIENCE, GUT, Australian BioCommons, and Galaxy Australia.



Galaxy Evropa

- <https://usegalaxy.eu>



The screenshot shows the Galaxy Europe website interface. The top navigation bar is dark blue with the "Galaxy Europe" logo on the left and links for "Workflow", "Visualize", "Shared Data", "Help", and "Log in or Register" on the right. Below the navigation bar, the left sidebar contains a "Tools" section with a search bar and an "Upload Data" button. The main content area features a large blue banner for the "Call for Training Topics!" for the "GCC 2024" (Galaxy Community Conference 2024). The banner includes a QR code and a link to "Fill out the form to share your thoughts!". The background of the banner shows a stylized illustration of a city skyline.



Galaxy Training Network (GTN)

- Če se želite samostojno učiti uporabljati Galaxy, lahko uporabite spletno zbirko učnih materialov GTN: <https://training.galaxyproject.org/>
- Seznam materialov za začetnike v zbirki GTN:
<https://training.galaxyproject.org/training-material/topics/introduction/>

Your First Analysis

These are the core, foundational topics for learning how to use Galaxy.

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
A short introduction to Galaxy     español	 ▾  ▾	 ▾			
From peaks to genes   		 ▾			
Galaxy 101   		 ▾	 ▾		

Domača stran Galaxy EU

Na levi strani je stolpec z orodji in iskalnik orodij.

V sredini se pojavi obrazec za nastavitev parametrov orodja.

Na desni strani je zgodovina korakov.

The screenshot displays the Galaxy EU home page interface, which is organized into three main vertical sections:

- Left Sidebar (Tools and Navigation):** Contains icons and labels for 'Upload', 'Tools' (highlighted with a wrench icon), 'Workflows', 'Workflow Invocations', 'Visualization', 'Histories', and 'Notifications'.
- Tools List (Left Panel):** A search bar contains the text 'replace'. Below it, a list of tools is shown, including 'Replace parts of text', 'ReplaceSamHeader', 'Replace Text in a specific column', 'Replace Text in entire line', 'Replace ambiguous codons', 'Replace chromosome names', 'Replace column', and 'AddOrReplaceReadGroups'.
- Central Tool Configuration Area:** The tool 'Replace parts of text (Galaxy Version 9.3+galaxy1)' is selected. It features a 'Run Tool' button and a 'Tool Parameters' section. A message states: 'Please provide a value for this option. File to process * required'. Below this, there is a file selection interface with icons for file, folder, and a dropdown menu labeled 'accepted formats'. An orange box indicates 'No txt datasets available'. The 'Find and Replace' section includes a 'Find pattern' field (optional) and a 'Replace with' field (optional).
- Right Sidebar (History):** Titled 'History', it includes a search bar for 'search datasets'. Below it, it shows 'Unnamed history' with a message: 'This history is empty. You can load your own data or get data from an external source.'

Praktični primer izdelave in uporabe delotoka v Galaxy

- V Galaxy EU bomo (iz UniProt) uvozili aminokislinski zaporedji proteinov **Bet v 1** in **Mal d 1** v formatu FASTA.
- Z orodjem blastp bomo izračunali odstotek ujemanja med aminokislinskima zaporedjema Bet v 1 in Mal d 1.
- V vseh motivih **treonin – izolevcin** (TI) v zaporedju Mal d 1 bomo zamenjali treonin s serinom in odstranili izolevcin (**zamenjava TI v S**).
- Z blastp bomo ponovno izračunali odstotek ujemanja med Bet v 1 in “spremenjenim” Mal d 1 (Mal d 1 S).
- Iz zgodovine korakov bomo naredili delotok, ki bo računal podobnost z Bet v 1 pred in po spremembi TI – S za katerikoli vhodni protein.
- Preizkusili bomo delotok s **Pru Ar 1-like** kot poljubnim vhodnim proteinom.

Ozadje: Alergijska reakcija

- Proteini so naši gradniki, lahko so receptorji, hormoni, encimi ali komponente imunskega sistema.
- Lahko so tudi za telo škodljivi tujki, ki jih prepozna imunski sistem in sproži obrambni odziv.
- V nekaterih primerih takšni tujki sprožijo pretirani imunski odziv - to je alergijska reakcija.

Navzkrižna reaktivnost

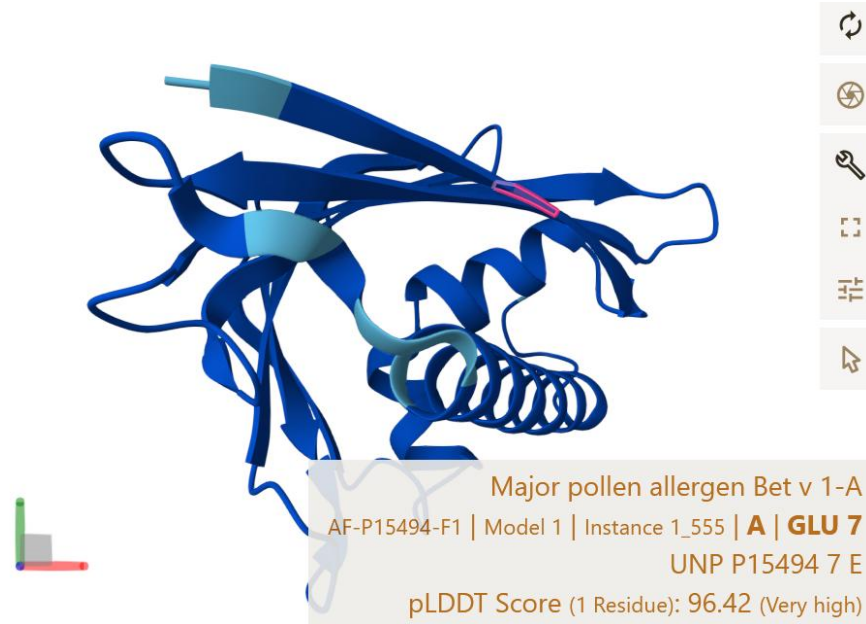
- Številni ljudje trpijo zaradi alergij in preobčutljivosti, pri alergijah pa se lahko pojavi navzkrižna reaktivnost: pri posamezniku, ki je predhodno razvil alergijsko reakcijo na nek proteinski alergen, lahko pride do alergijske reakcije ob stiku s proteinom, ki je podoben alergenemu.
- Primer: posameznik je alergičen na protein Bet v 1 iz cvetnega prahu breze, potem pa razvije alergijsko reakcijo, ko zaužije jabolko, ker jabolko vsebuje protein Mal d 1, ki je podoben Bet v 1.
- Tveganje za navzkrižno reaktivnost naj bi bilo sorazmerno s podobnostjo aminokislinskih zaporedij med nekim proteinom in znanim alergenom.

Protein Bet v 1

UniProt šifra: P15494

Izvor: cvetni prah breze
(lat. *Betula pendula*)

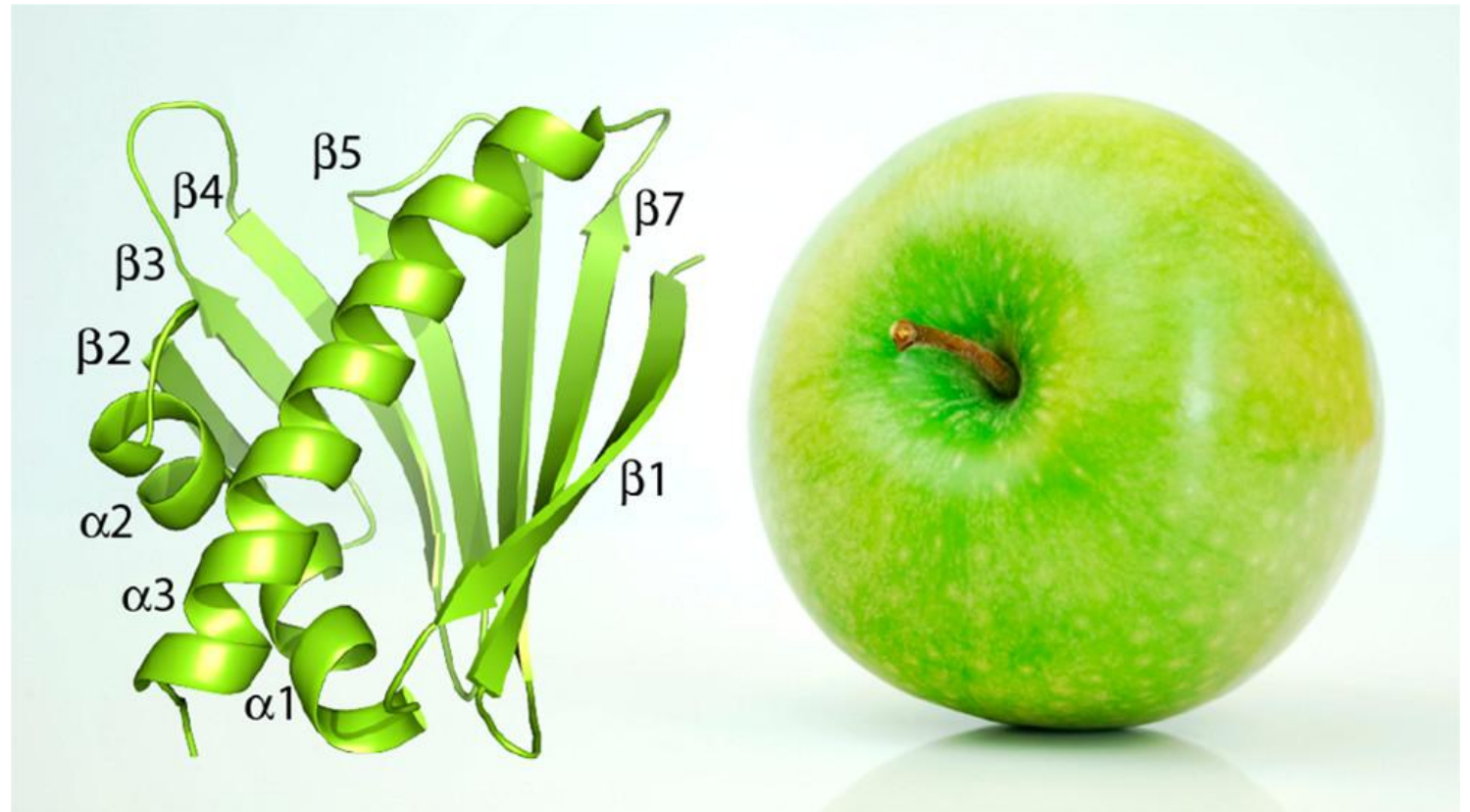
Eden od najpogostejših
povzročiteljev alergijske
reakcije.



Protein Mal d 1

UniProt šifra: Q9SYW3

Izvor: jabolko (lat. *Malus domestica*)



Aminokislinski zaporedji Bet v 1 in Mal d 1 v formatu FASTA

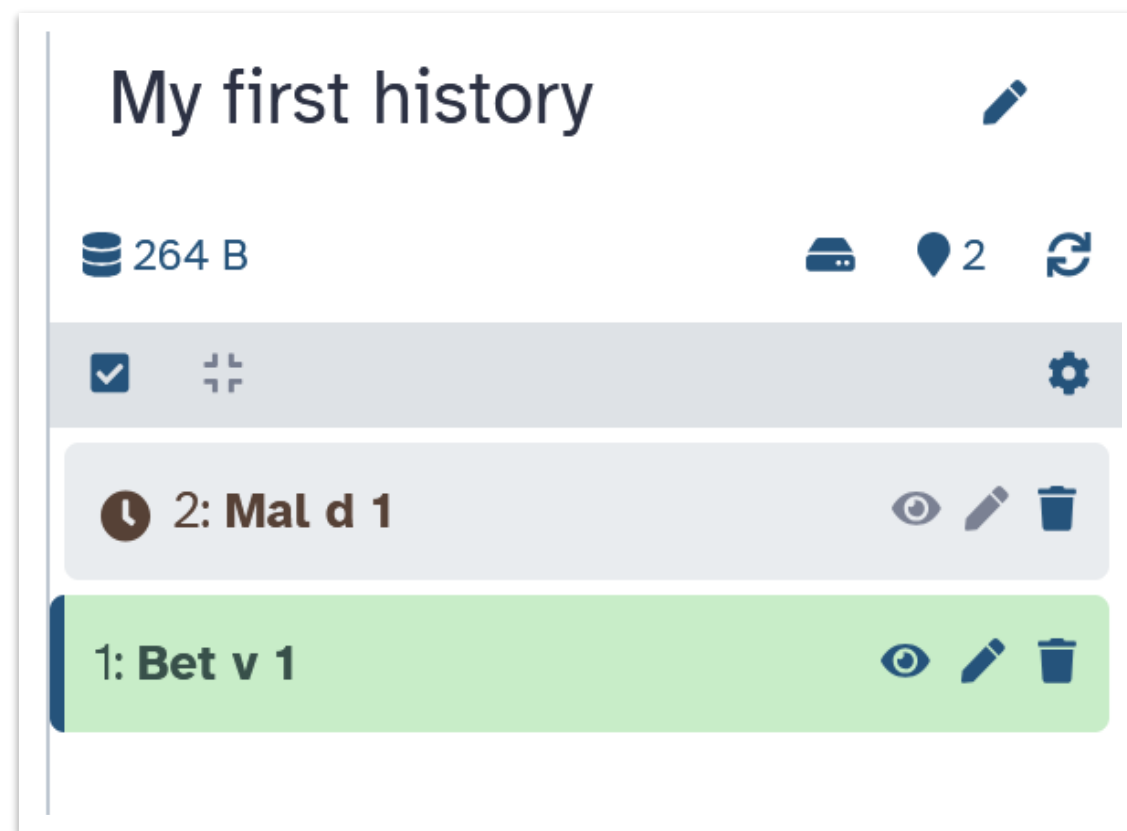
V podatkovni zbirki UniProt (<https://www.uniprot.org/>) poiščemo aminokislinski zaporedji Bet v 1 in Mal d 1 v formatu FASTA (enočrkovne oznake aminokislin).

```
>sp|P15494|BEV1A_BETPN Major pollen allergen Bet v 1-A OS=Betula pendula OX=3505 GN=BETVIA PE=1 SV=2
MGVFNYETETTSVIPAAARLFKAFILDGDNLFPKVAPQAISSVENIEGNGGPGTIKKISFP
EGFPFKYVKDRVDEVDHTNFKYNYSVIEGGPIGDTLEKISNEIKIVATPDGGSILKISNK
YHTKGDHEVKAEQVKASKEMGETLLRAVESYLLAHSDAYN
```

```
>tr|Q9SYW3|Q9SYW3_MALDO Major allergen Mal d 1 OS=Malus domestica OX=3750 GN=MALD1 PE=2 SV=1
MGVYTFENEFTSEIPPSRLFKAFVLDADNLIPKIA PQAIKQAEILEGNGGPGTIKKITFG
EGSQYGYVKHRIDSIDEASYSYSYTLIEGDALDTIEKISYETKLVACGSGSTIKSISHY
HTKGNIEIKEEHVKAGKEKAHGLFKLIESYLKDHPDAYN
```

Uvoz podatkov iz UniProt v Galaxy

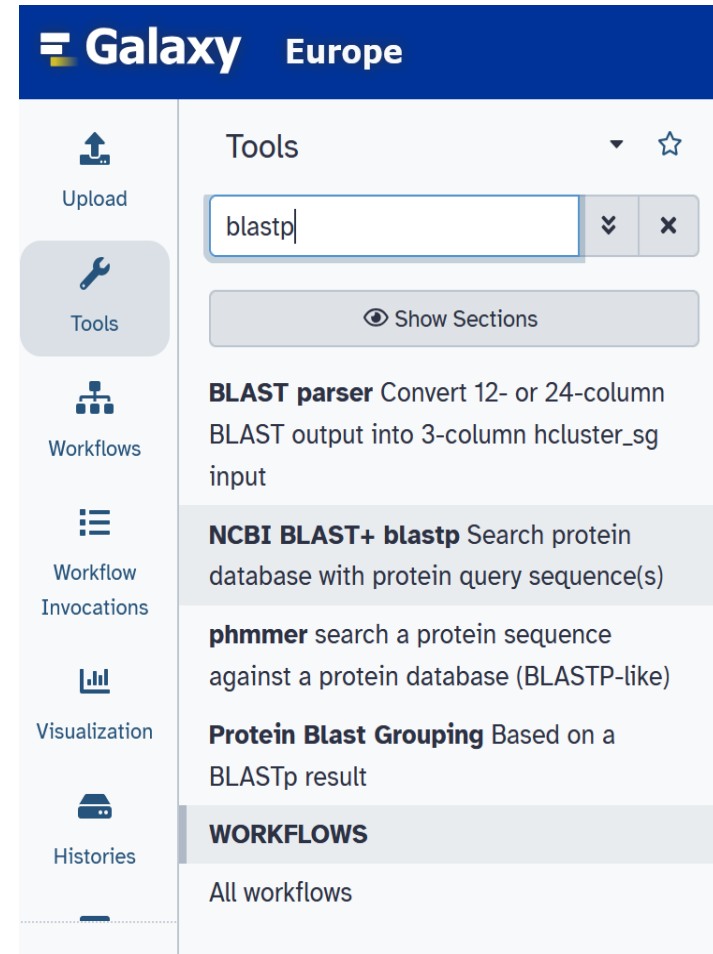
V zgodovino v Galaxy uvozimo zaporedji Bet v 1 in Mal d 1 v formatu FASTA.



Iz “orodjarne” v Galaxy izberemo orodje blastp

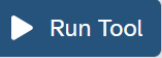
Blastp je orodje (del NCBI BLAST) za primerjavo aminokislinskih zaporedij proteinov.

1. V iskalnik orodij vpišemo “blastp”.
2. S seznama zadetkov izberemo “**NCBI BLAST+ blastp** Search protein database...”



Nastavitev orodja blastp

1. Kot “Protein query sequence(s)” izberemo **Mal d 1** (ki smo ga pred tem uvozili v našo zgodovino v Galaxy).
2. Pri “Subject database/sequences” izberemo možnost “FASTA file from your history...”.
3. Kot “Protein FASTA subject file...” izberemo **Bet v 1** (predhodni uvoz v zgodovino).
4. Pri “Type of Blast” izberemo “blastp – Traditional BLASTP...”.
5. Kliknemo “Run Tool”.

 **NCBI BLAST+ blastp**     

Search protein database with protein query sequence(s)
(Galaxy Version 2.14.1+galaxy2)

Tool Parameters

Protein query sequence(s) *

2: Mal d 1 

accepted formats 

(-query)

Subject database/sequences

FASTA file from your history (see warning note below) 

Protein FASTA subject file to use instead of a database *

1: Bet v 1 

accepted formats 

(-subject)

Type of BLAST *

☒ blastp - Traditional BLASTP to compare a protein query to a protein database

☐ blastp-short - BLASTP optimized for queries shorter than 30 residues

Pregled rezultatov blastp

1. Ko se rezultati blastp (**blastp Mal d 1 vs...**) pojavijo v zgodovini, kliknemo na ikono očesa v koraku v zgodovini in pregledamo rezultate.
2. Izpiše se stopnja podobnosti med aminokislinskima zaporedjema proteinov Mal d 1 in Bet v 1 (v odstotkih – 56,875 %)

Column 1
56.875

History

+

↔

≡

search datasets

▼

×

my-first-history

526 B

3

↺

✓

⌵

⚙

3: blastp Mal d 1 vs 'Bet v 1'

👁

✎

🗑

2: Mal d 1

Display

1: Bet v 1

👁

✎

🗑

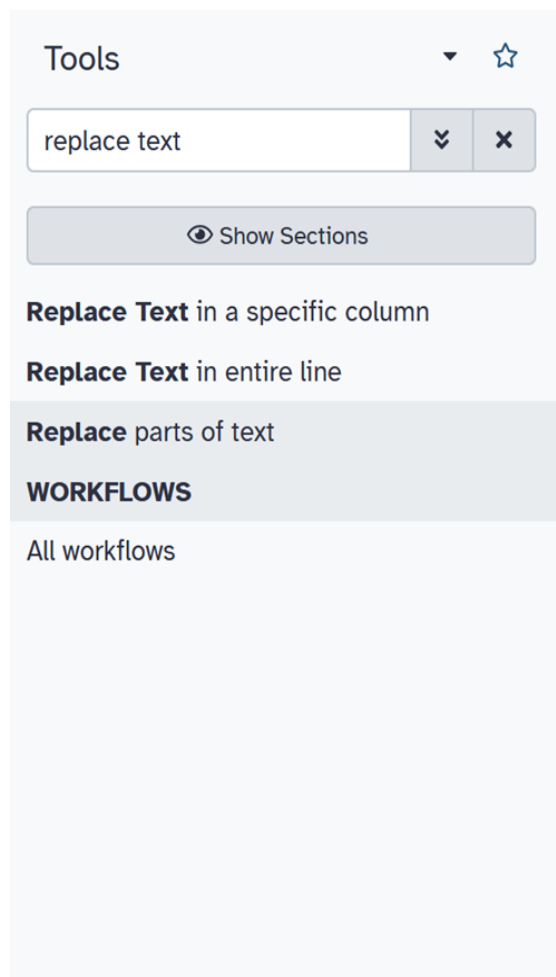
Orodje “replace text”: zamenjava TI s S

1. V iskalnik orodij vpišemo „replace“.

2. Izberemo „**Replace** parts of text“

3. S seznama izberemo “**Mal d 1**” kot vhodni podatek.

4. “**TI**” zamenjamo s “**S**” (enočrkovne oznake aminokislin v zapisu FASTA).



Tools

replace text

Show Sections

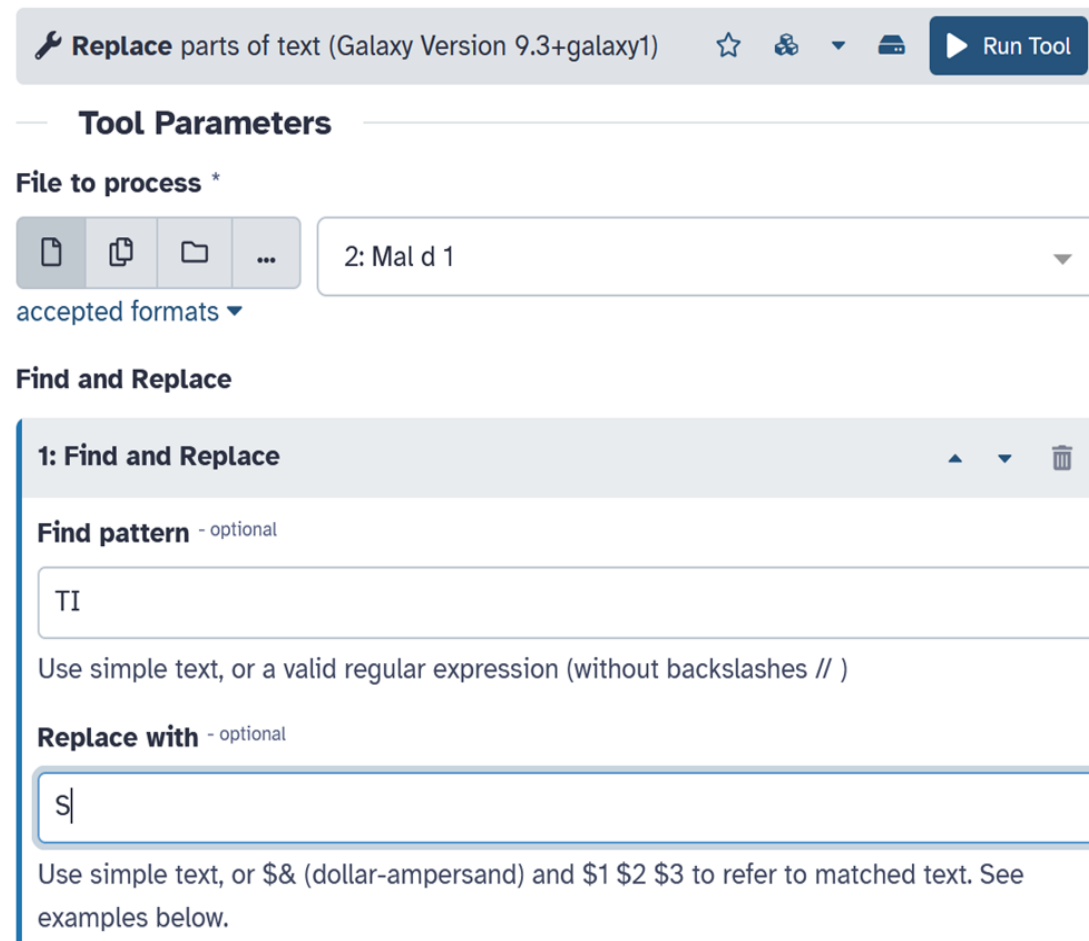
Replace Text in a specific column

Replace Text in entire line

Replace parts of text

WORKFLOWS

All workflows



Replace parts of text (Galaxy Version 9.3+galaxy1)

Run Tool

Tool Parameters

File to process *

2: Mal d 1

accepted formats

Find and Replace

1: Find and Replace

Find pattern - optional

TI

Use simple text, or a valid regular expression (without backslashes //)

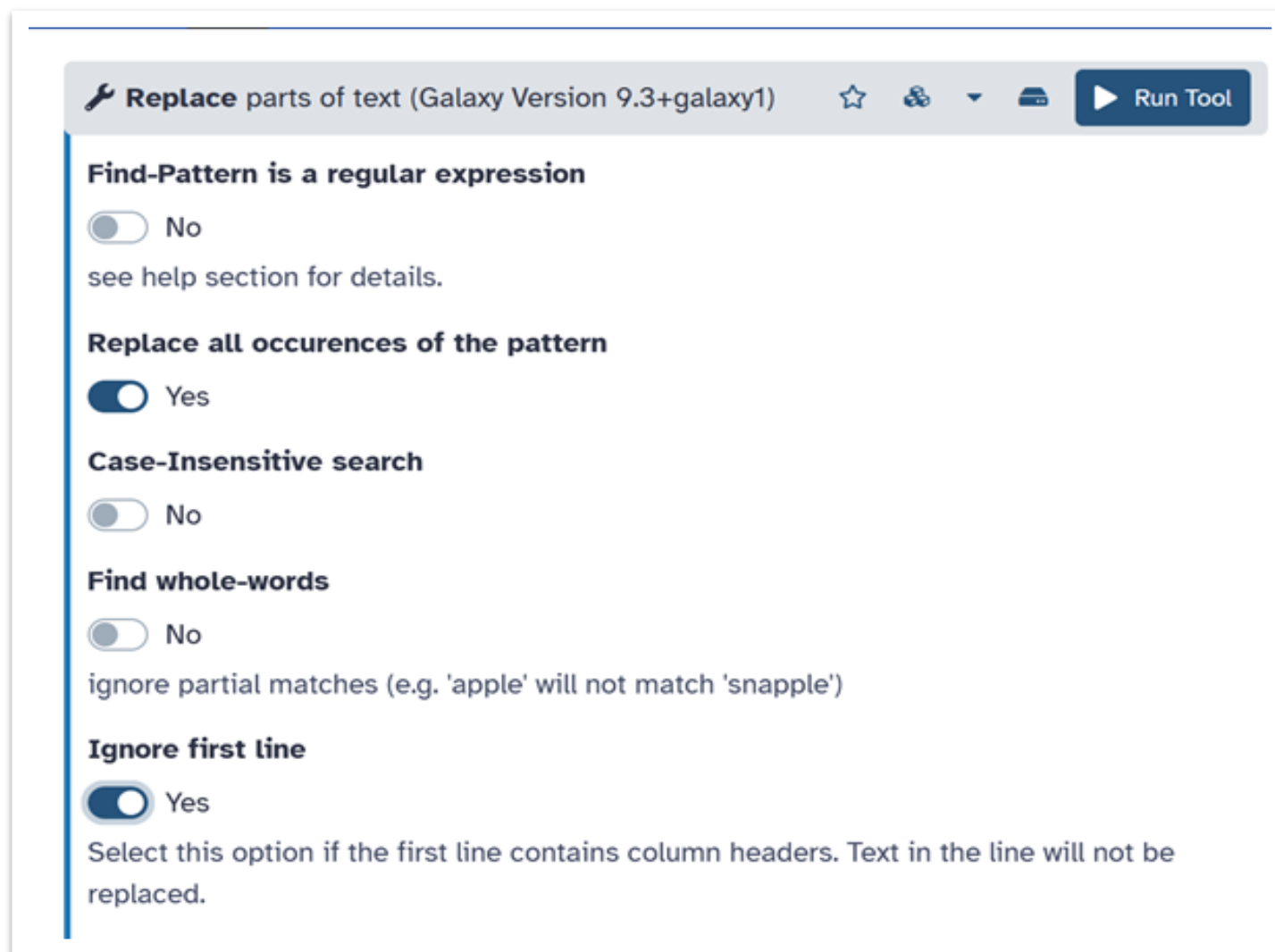
Replace with - optional

S

Use simple text, or \$& (dollar-ampersand) and \$1 \$2 \$3 to refer to matched text. See examples below.

Orodje “replace text” (nadaljevanje)

5. Izberemo možnost “Replace all occurrences of the pattern”
6. Izberemo možnost “Ignore first line”, ker ne želimo, da bi se TI zamenjal s S tudi v naslovni vrstici, kjer so opisni podatki (metapodatki).
7. Kliknemo “Run Tool”.



Preimenovanje rezultatov v “Mal d 1 S”

1. Kliknemo na ikono svinčnika v koraku v zgodovini.

2. Rezultate (Replace on data...) preimenujemo v “**Mal d 1 S**”.

3. Kliknemo “Save”.

The screenshot displays the 'Edit Dataset Attributes' interface. The main panel has three tabs: 'Attributes' (selected), 'Datatypes', and 'Permissions'. The 'Name' field is set to 'Mal d 1 S'. Below it are 'Info' and 'Annotation' (optional) text areas. At the bottom is a 'Database/Build' (optional) dropdown menu set to 'unspecified (?)'. Two buttons are at the bottom: 'Save' and 'Auto-detect'.

On the right, a 'History' panel shows a list of dataset operations. The top entry is 'my-first-history' with a size of 778 B. Below it are four entries in a green bar: '4: Replace on data 2', '3: blastp Mal d 1 vs 'Bet v 1'', '2: Mal d 1', and '1: Bet v 1'. Each entry has icons for view, edit, and delete. An 'Edit attributes' button is visible next to the third entry.

Ponovimo blastp (z Mal d 1 S kot vhodnim podatkom)

- Podobnost med Mal d 1 S in Bet v 1 je 54,375 %, kar je manj kot pri nespremenjenem Mal d 1 (56, 875 %).

The screenshot displays the Galaxy web interface for running the NCBI BLAST+ blastp tool. The interface is divided into several sections:

- Tools:** A sidebar on the left lists various tools. The 'blastp' tool is selected, showing its description: 'Search protein database with protein query sequence(s) (Galaxy Version 2.14.1+galaxy2)'. Other tools listed include 'BLAST parser', 'phmmer', and 'Protein Blast Grouping'.
- Tool Parameters:** The main configuration area for the 'blastp' tool. It includes:
 - Protein query sequence(s) *:** A dropdown menu showing '4: Mal d 1 S'.
 - Subject database/sequences:** A dropdown menu showing 'FASTA file from your history (see warning note below)'.
 - Protein FASTA subject file to use instead of a database *:** A dropdown menu showing '1: Bet v 1'.
 - Type of BLAST *:** A radio button selection for 'blastp - Traditional BLASTP to compare a protein query to a protein database'.
- History:** A sidebar on the right showing a list of previous jobs. The jobs are numbered 1 to 5, with job 5 being the current job: '5: blastp Mal d 1 S vs 'Bet v 1''. The jobs are listed in a table with columns for job name, size, and actions (view, edit, delete).

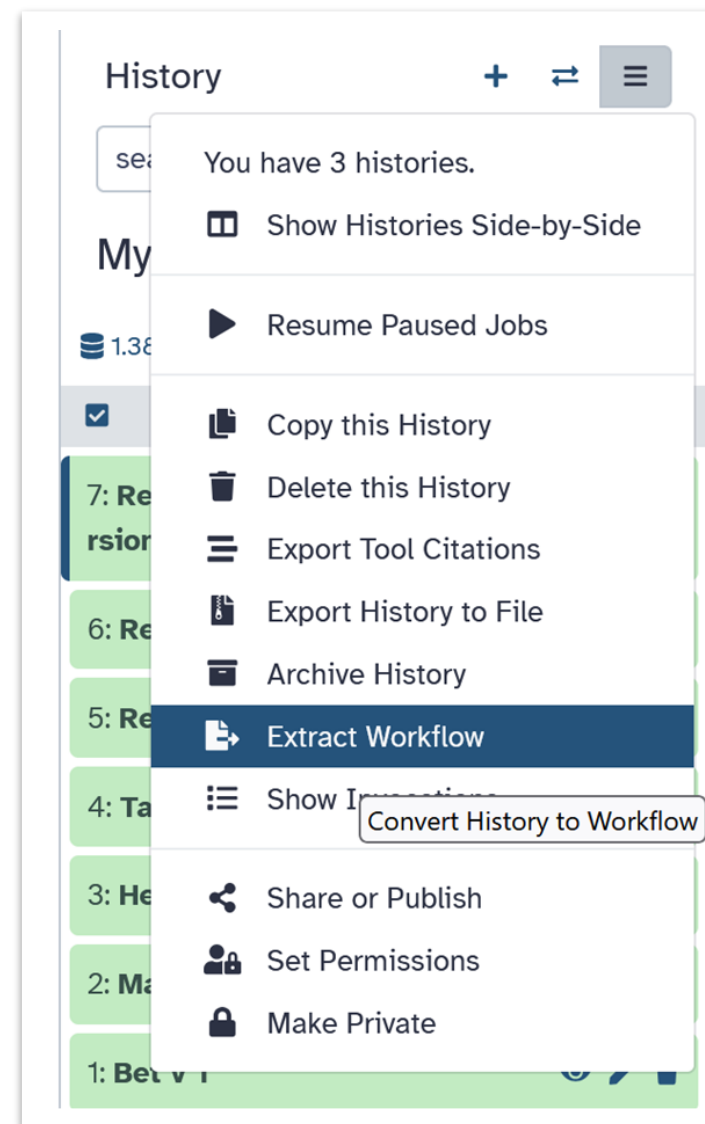
The output of the tool is displayed in a table with the following data:

Column 1
54.375

Izdelava delotoka iz zgodovine

Korake v zgodovini želimo povezati v delotok, ki ga lahko uporabimo za kombinacijo Bet v 1 s katerim koli proteinom.

1. S spustnega seznama, ki se pojavi po kliku na ikono s tremi črtami, izberemo možnost **“Extract Workflow”**.



Izdelava delotoka iz zgodovine (nadaljevanje)

2. Delotok poimenujemo:

“Sequence similarity between Bet v 1 and random protein”.

3. Kliknemo na “Create Workflow”.

Tools which cannot be run interactively and thus cannot be incorporated into a workflow will be shown in gray.

Workflow name

Sequence similarity between Bet v1 and random protein

Create Workflow

Check all

Uncheck all

Tool

History items created

Data Fetch

This tool cannot be used in workflows



1 Bet v 1

☒ Treat as input dataset

Bet v 1

Data Fetch

This tool cannot be used in workflows



2 Mal d 1

☒ Treat as input dataset

Mal d 1

my-first-history

785 B

5



5: blastp Mal d 1 S vs 'Bet v 1'



4: Mal d 1 S



3: blastp Mal d 1 vs 'Bet v 1'



2: Mal d 1

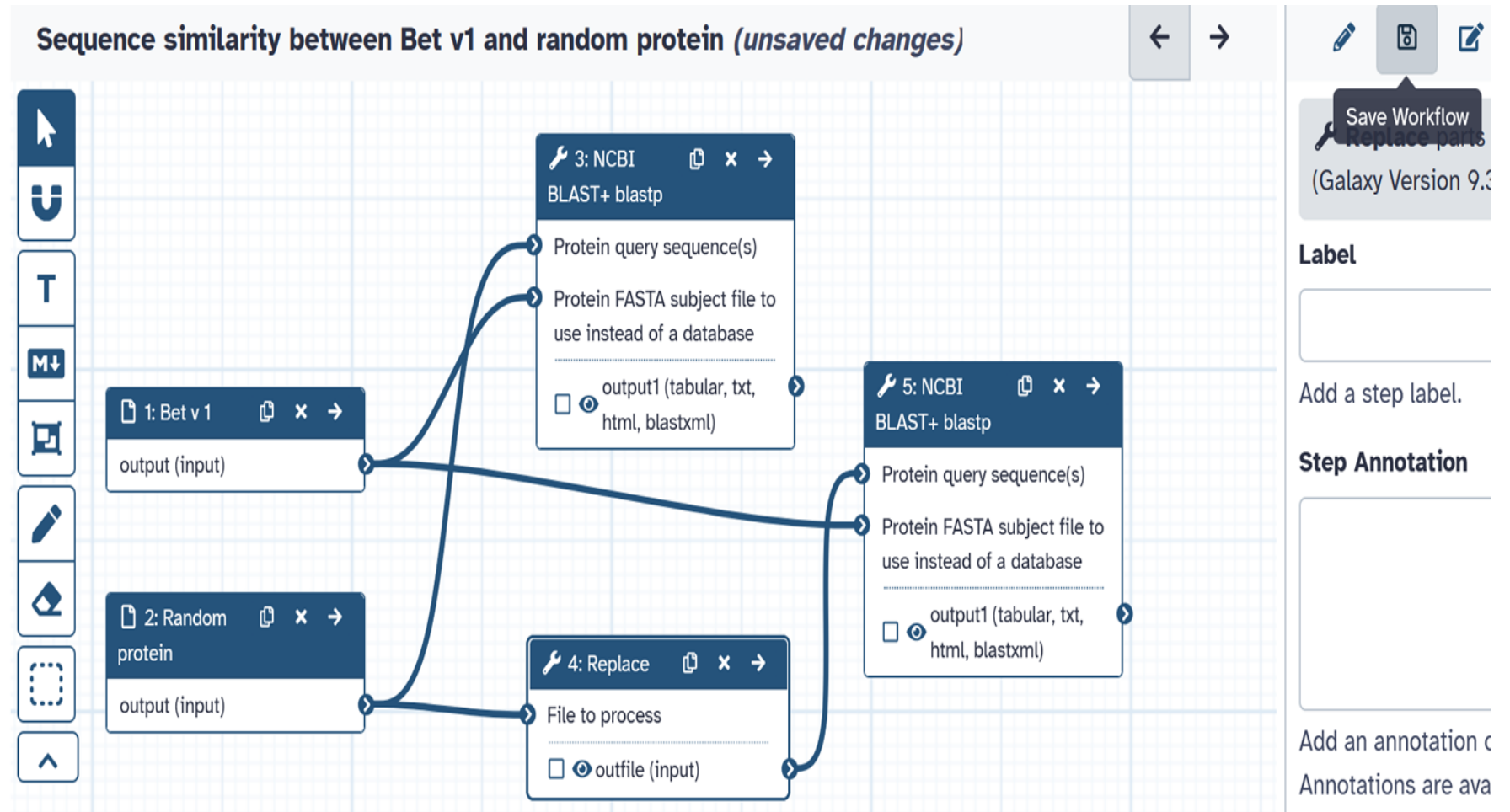


1: Bet v 1



Uredimo shemo delotoka

Spremembe v delotoku shranimo s klikom na “Save Workflow”.

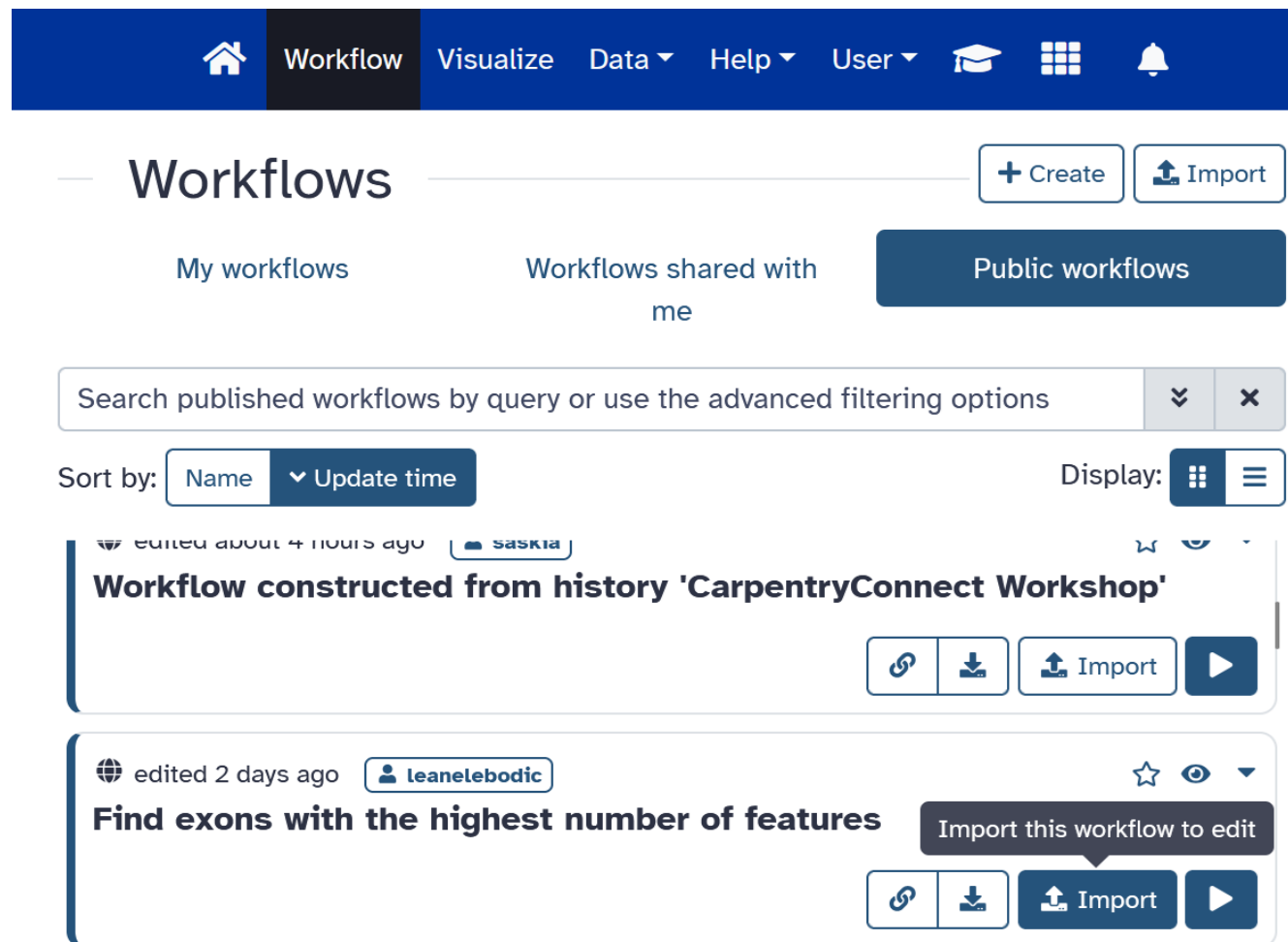


Različni načini za izdelavo delotoka

Izdelava delotoka iz zgodovine ni edina možnost.

Galaxy omogoča izdelavo delotokov na tri načine:

1. Iz zgodovine
2. Uvoz že pripravljenih delotokov
3. Ustvarjanje delotokov v urejevalniku (brez zgodovine)



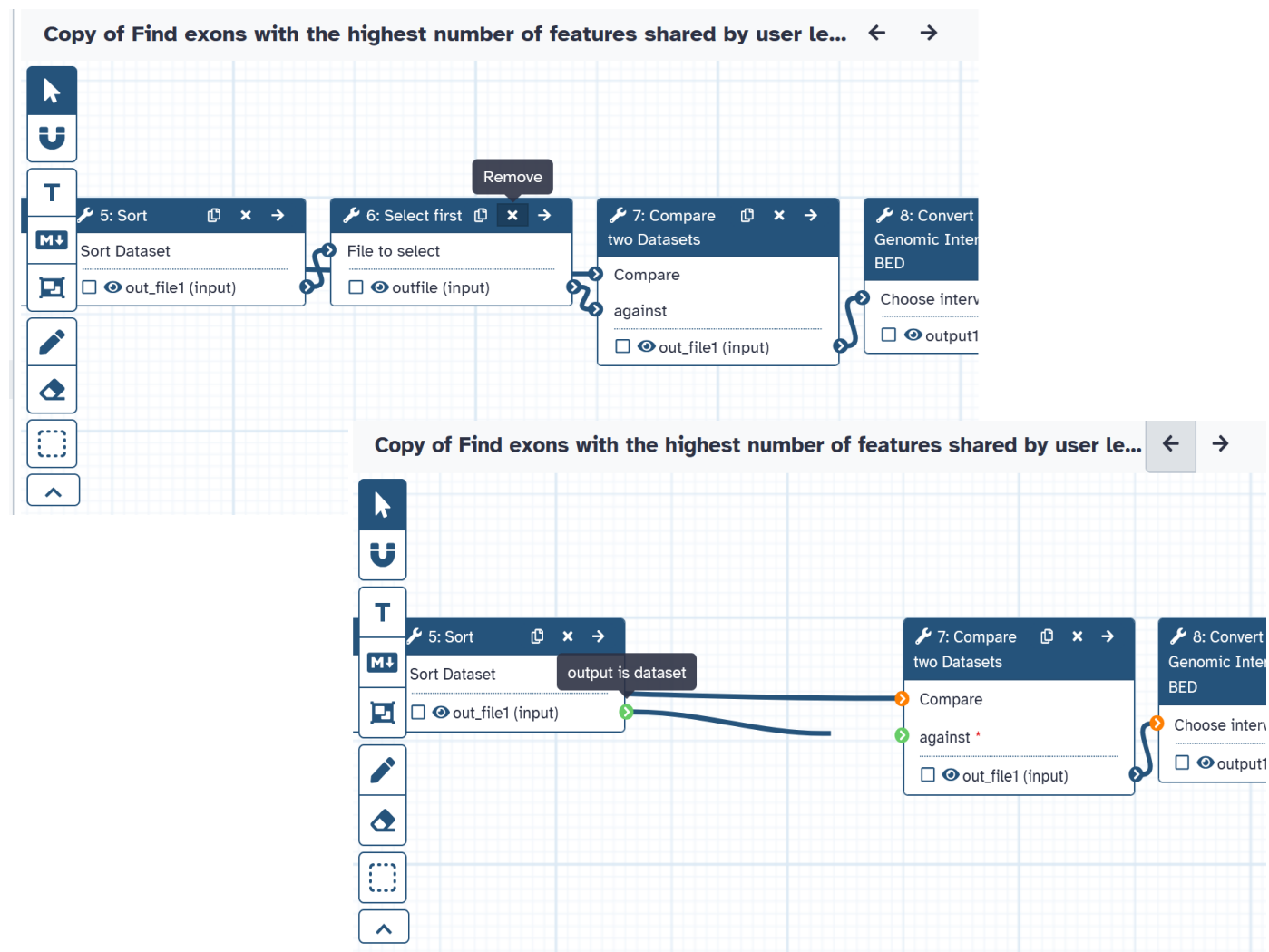
The screenshot displays the Galaxy Workflows interface. At the top is a navigation bar with icons for Home, Workflow (selected), Visualize, Data, Help, User, and a grid icon. Below the navigation bar, the 'Workflows' section is active, showing tabs for 'My workflows', 'Workflows shared with me', and 'Public workflows' (which is selected). A search bar is present with the text 'Search published workflows by query or use the advanced filtering options'. Below the search bar, there are sorting options ('Sort by: Name', 'Update time') and a display toggle ('Display: Grid', 'List'). Two workflow cards are visible. The first card is titled 'Workflow constructed from history 'CarpentryConnect Workshop'' and was edited about 4 hours ago by 'saskia'. It has buttons for 'Link', 'Download', 'Import', and 'Run'. The second card is titled 'Find exons with the highest number of features' and was edited 2 days ago by 'leanelebodice'. It also has buttons for 'Link', 'Download', 'Import', and 'Run'. A tooltip 'Import this workflow to edit' is shown over the 'Import' button of the second card.

Urejanje uvoženega delotoka

Primer urejanja
uvoženega delotoka:

Odstranimo korak
“Select first” (ker želimo,
da se naslednje operacije
v delotoku izvedejo na
celotnem obsegu
vhodnih podatkov in ne
samo na prvi vrstici).

Po odstranitvi koraka
delotok ponovno
povežemo

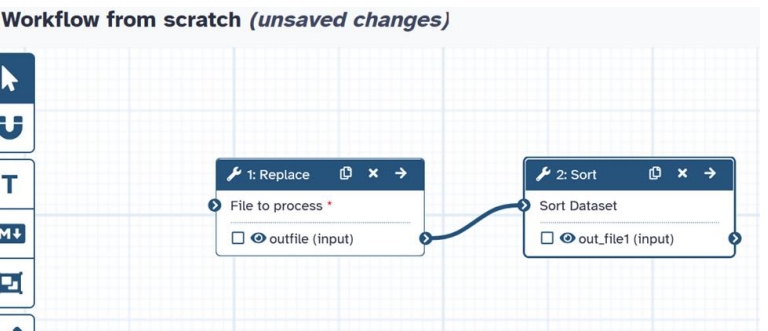
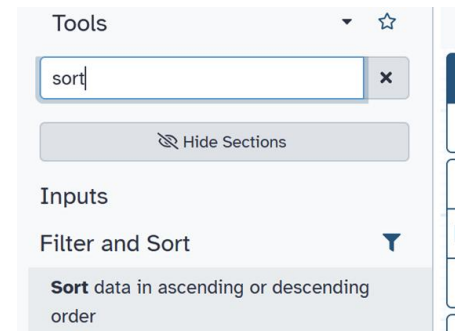
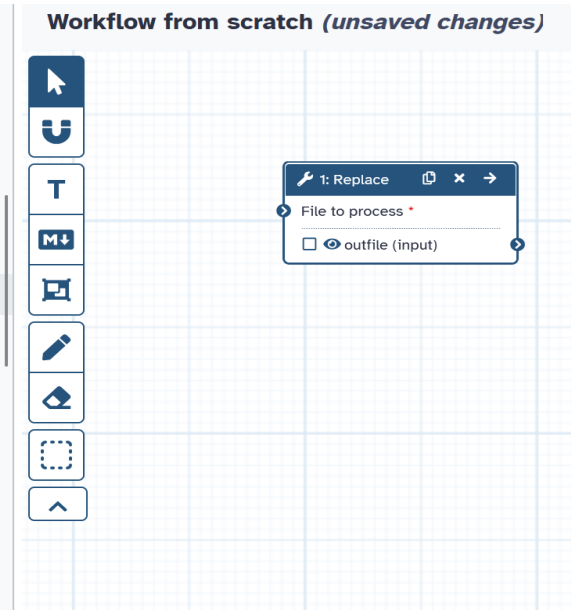
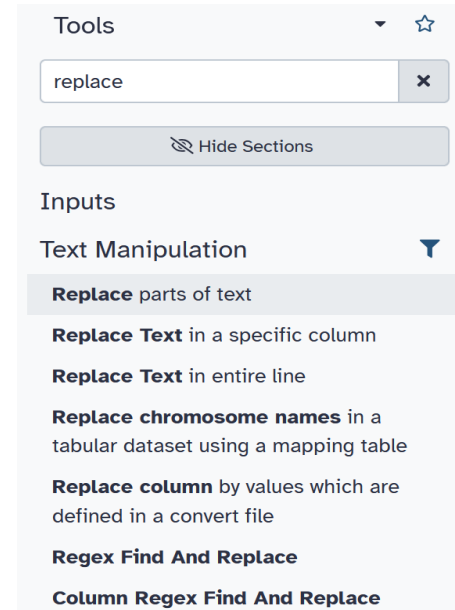
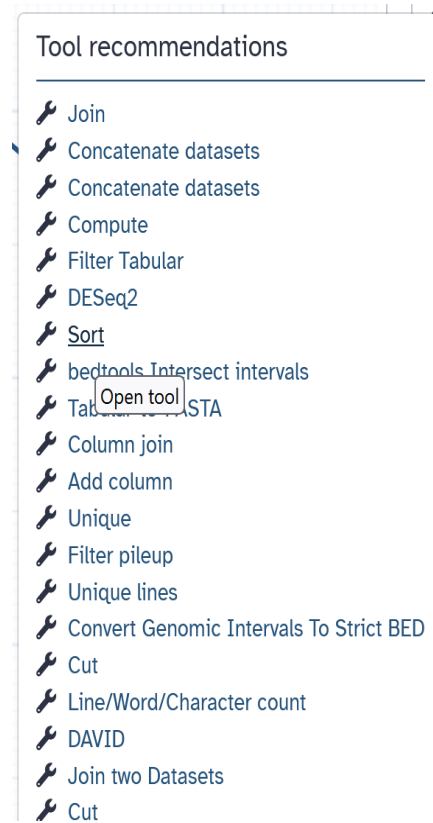


Izdelava delotoka v urejevalniku

Iz orodjarne izberemo orodje, ki ga želimo uporabiti v delotoku.

Okvirček, ki predstavlja to orodje, se pojavi v urejevalniku.

Ko ustvarimo okvirčke za vsa orodja, moramo okvirčke povezati v smiselno shemo delotoka.



Protein Pru Ar 1-like

V Galaxy uvozimo aminokislinsko zaporedje proteina **Pru Ar 1-like (UniProt šifra: Q6QHU2)** v formatu FASTA (iz UniProt). To je protein iz češnje (lat. *Prunus avium*).



Major cherry allergen Pru av 1.0202
AF-Q6QHU2-F1 | Model 1 | Instance 1_555 | **A** | **SER 41**
Confidence score: 93.44 (High)



Zagon delotoka “Sequence similarity...”

1. Kliknemo na zavihek Workflow.
2. Izberemo delotok “Sequence similarity...” in kliknemo na ikono “Run Workflow”.

The screenshot displays the 'Workflows' management interface. The top navigation bar includes 'Workflow', 'Visualize', 'Data', 'Help', 'User', and a 'Using 0%' indicator. The main section is titled 'Workflows' and features tabs for 'My workflows', 'Workflows shared with me', and 'Public workflows'. A search bar allows filtering workflows by query. Below the search bar, there are options to sort by 'Name' or 'Update time' and to filter by 'Show deleted' or 'Show bookmarked'. The workflow list shows two entries: 'Sequence similarity between Bet v1 and random protein' (edited 18 minutes ago, never run) and 'Test workflow' (edited 1 day ago, 1 workflow run). Each entry has a 'Run workflow' button. On the right, a 'History' panel shows a list of workflow runs, including '6: Pru Ar 1-like', '5: blastp Mal d 1 S vs 'Bet v 1'', '4: Mal d 1 S', '3: blastp Mal d 1 vs 'Bet v 1'', '2: Mal d 1', and '1: Bet v 1'.

Zagon delotoka – izberemo Pru Ar 1-like kot vhodni podatek

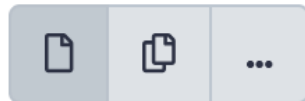
3. Izberemo prvi vhodni podatek: **Bet v 1**
4. Izberemo drugi vhodni podatek (“Random protein”): v tem primeru Pru Ar 1-like.
5. Kliknemo “Run Workflow”.

Workflow: Sequence similarity between Bet v1 and random protein
(version: 3)



▶ Run Workflow

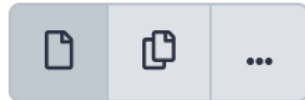
Bet v 1 *



accepted formats ▼

1: Bet v 1 ▼

Random protein *



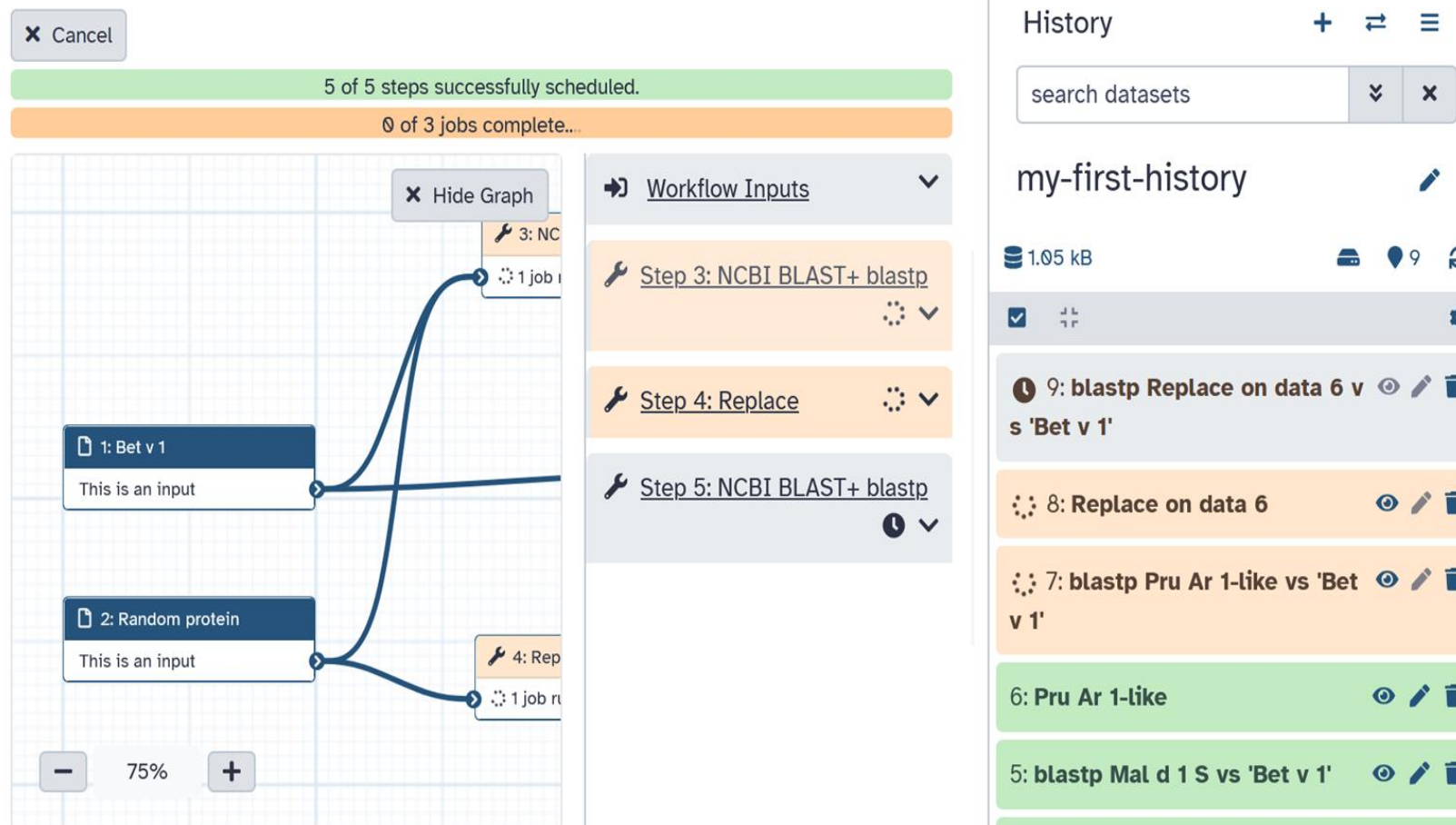
accepted formats ▼

6: Pru Ar 1-like ▼

[Expand to full workflow form.](#)

Spremljanje izvajanja delotoka

V zgodovini se pojavijo vsi koraki, ki smo jih prej delali ročno korak za korakom.



Rezultati delotoka

Podobnost med Pru Ar 1-like in Bet v 1: 56,875 %.

Column 1
56.875

History + ↔ ≡

search datasets ▼ ✕

my-first-history ✎

1.33 kB 📁 🔒 9 🔄

☒ 📄 🔍 ⚙️

9: blastp Replace on data 6 vs 'Bet v 1' 👁 ✎ 🗑

8: Replace on data 6 👁 ✎ 🗑

7: blastp Pru Ar 1-like vs 'Bet v 1' 👁 ✎ 🗑

Display

Podobnost med Pru Ar 1-like **S** in Bet v 1: 55,625 %.

Column 1
55.625

History + ↔ ≡

search datasets ▼ ✕

my-first-history ✎

1.33 kB 📁 🔒 9 🔄

☒ 📄 🔍 ⚙️

9: blastp Replace on data 6 vs 'Bet v 1' 👁 ✎ 🗑

Display

Zaključek

- virtualizacija, vsebniki, programska okolja, delotoki
- olajšajo analize in poenostavijo ponovljivost
- izziv je poenotenje / standardizacija
 - interoperabilnost vsebnikov
 - izmenljivost delotokov
 - skupni jezik za delotoke



snakemake



COMMON
WORKFLOW
LANGUAGE

nextflow



Galaxy
PROJECT



MINI CONDA



CONDA



Vprašanja ?



Univerza v Ljubljani, Medicinska fakulteta
University of Ljubljana, Faculty of Medicine
Vrazov trg 2, SI-1000 Ljubljana
Slovenija / Slovenia

Ekipa IBMI/Središča ELIXIR-SI:

dr. B. Leskošek
dr. P. Ferk
dr. M. Vidak
dr. B. Kverh
J. Gojznikar

Nacionalni inštitut za biologijo (NIB)
Oddelek za biotehnologijo in sistemsko biologijo
Večna pot 121, 1000 Ljubljana
Slovenija

Ekipa NIB:

dr. Maja Zagorščak
dr. Nadja Nolte
dr. Carissa Bleker
dr. Marko Petek
dr. Anže Županič

