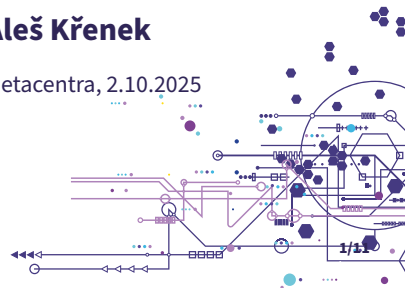


Galaxy (nejen) v Česku: mýty a fakta

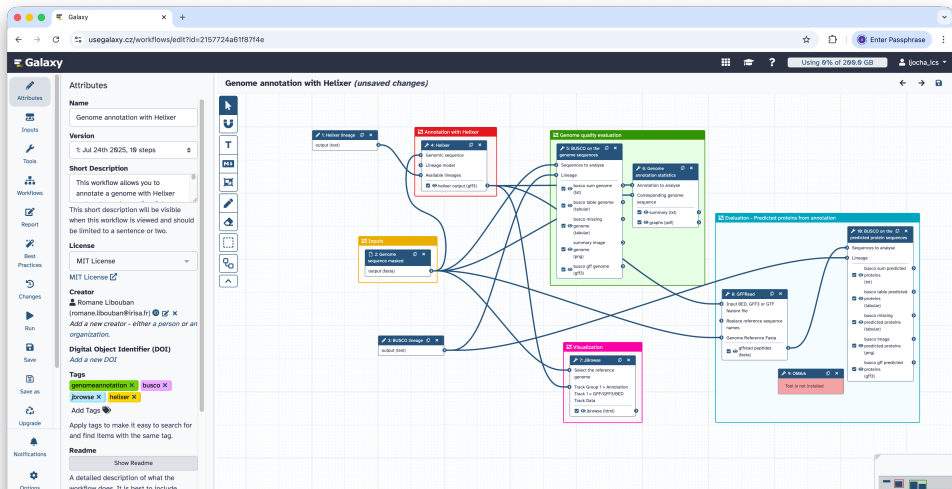
Martin Čech, Aleš Křenek

Seminář uživatelů Metacentra, 2.10.2025





Co je Galaxy?



The screenshot displays the Galaxy web interface with a workflow titled "Genome annotation with Helixer (unsaved changes)". The interface includes a left sidebar with navigation options like Inputs, Tools, Workflows, Report, Best Practices, Changes, Run, Save, Save as, Upgrade, and Notifications. The main workspace shows a workflow diagram with several steps:

- Inputs:** A step labeled "Inputs" with a dropdown menu showing "Genomic sequence" and "Available images".
- Annotation with Helixer:** A step labeled "4 Helixer" with a dropdown menu showing "Genomic sequence", "Ligase model", and "Available images".
- Genome quality evaluation:** A step labeled "5 BUSCO on the genome sequences" with a dropdown menu showing "Sequences to analyse", "Ligase", "Index run genome (nt)", "Index total genome (nt)", "Index missing (nt)", "Index summary (nt)", "Index graph (nt)", "Index genome (nt)", "Index gff genome (nt)", "Index annotation statistics", and "Corresponding genome sequences".
- Visualization:** A step labeled "7 JBrowse" with a dropdown menu showing "Select the reference genome", "Track Group 1 > Annotation", "Track 1 > GFF/GTF/BED", "Track Data", and "JBrowse (html)".
- Evaluation - Predicted proteins from annotation:** A step labeled "8 BLAST on the predicted protein sequences" with a dropdown menu showing "Sequences to analyse", "Ligase", "Index run predicted (nt)", "Index table predicted (nt)", "Index missing (nt)", "Index predicted proteins (nt)", "Index missing (nt)", "Index predicted proteins (nt)", "Index missing (nt)", "Index gff predicted (nt)", and "Index gff predicted (nt)".

The workflow is titled "Genome annotation with Helixer (unsaved changes)". The interface also shows a "Short Description" section with the text: "This workflow allows you to annotate a genome with Helixer. This short description will be visible when this workflow is viewed and should be limited to a sentence or two." The "License" is set to "MIT License". The "Creator" is "Romane Libouban (romane.libouban@irisa.fr)". The "Digital Object Identifier (DOI)" is "Add a new DOI". The "Tags" are "genomeannotation", "busco", "JBrowse", and "Helixer". The "Readme" section is visible at the bottom.

Galaxy je pouze pro práci s genomickými daty

- Ano, bylo tomu tak v prvních verzích před 20 lety
- Zcela neutrální koncept
- Datové typy, nástroje, workflows, ...
- NLP, klimatologie, astronomie, materiálové vědy, strojové učení, ...

Galaxy není užitečná pro informatiky a lidi, kteří používají nástroje přímo, hodí se jen k demonstracím a výuce

- Integrované prostředí
- Nástroje, workflow, vizualizace a analýza dat
 - Jupyter, RStudio, ...
- **Provenance a reprodukovatelnost výsledků**
- Plně programovatelná přes API a knihovny
- Pohodlí správy verzí, závislostí a prostředí

Galaxy nemá dostatečnou kapacitu pro moje nástroje a výpočty

- usegalaxy.eu
 - 9000+ jader CPU, 60+ TB RAM, 5+ PB storage
 - 65 tis. úloh denně
 - rozhraní Pulsar – směřování úloh na další zdroje
 - „anonymní“ uživatelé – malé kvóty
- usegalaxy.cz
 - potenciálně celé Metacentrum (37k+ jader, GPUs)
 - vyšší kvóty pro české uživatele
 - eventuálně přímé napojení na repozitáře NRP

Galaxy nelze použít pro data vyžadující ochranu soukromí

- Rozhraní definovaná GA4GH
- Mechanismy „Bring Your Own Compute/Storage/Data“
 - data jsou uložena a zpracovávána výhradně na zdrojích k tomu určených
 - Galaxy tato data „nevidí“ díky mechanismu *deferred transfer*
- Certifikace ISO 27001 (usegalaxy.eu)
- Dedikované instalace (např. Sciensano, BE)

Takové open-source projekty mají špatnou dokumentaci, spoustu chyb, a nespolehlivou podporu

- Stabilizovaná komunita, demokratická organizační struktura
- Vyžadovaný standard povinné dokumentace a testů
- Příspěvky institucí a projektů, dlouhodobé plány
- Oficiálně podporovaný vstupní bod do e-Infra CZ

Nástroje v Galaxy jsou zastaralé

- Silné komunity za správou nástrojů
- Propracovaná infrastruktura, automatické testy a aktualizace
- Podpora více verzí téhož nástroje
- Desítky až stovky aktualizací denně (až jednotky %)

Galaxy je černá skříňka, používá proprietární rozhraní

- Rozsáhlý, stabilní, profesionální a úspěšný open-source projekt
- veškerá funkcionalita dostupná přes API (OpenAPI skrz FastAPI)
- podpora alternativních frontend klientů
- Otevřené protokoly (TES, CWL, gxformat2, ...)

Co dál?

- Zkuste <https://usegalaxy.cz/>
- Napište na galaxy@cesnet.cz
- Čtení pro stále nepřesvědčené:
W. A. Bacon et al.,
*Ten Common Misconceptions About Galaxy
(and Why They Are Wrong!)*
<https://www.preprints.org/manuscript/202509.0823/v1>