



HAL
open science

Using Galaxy-Ecology to operationalize EBVs

Coline Royaux, Pauline Seguineau, Arthur Barreau, Olivier Norvez, Marie Jossé,
Jean- Baptiste Mihoub, Bjoern Gruening, Dominique Pelletier, Marc Eléaume,
Yvan Le Bras

► To cite this version:

Coline Royaux, Pauline Seguineau, Arthur Barreau, Olivier Norvez, Marie Jossé, et al.. Using Galaxy-Ecology to operationalize EBVs. Living Data 2025, GEO BON, Oct 2025, Bogota, Colombia. <hal-05565171>

HAL Id: hal-05565171

<https://hal.science/hal-05565171v1>

Submitted on 24 Mar 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Using Galaxy-Ecology to operationalize EBVs

Session : 6798685 October 23, 2025

French BON EBV operationalization pilot: Reusing existing informatics solutions and open-source communities to connect existing data ecosystem and high-performance computing

Yvan Le Bras

MNHN / PNDB-Data Terra / UAR DoHNÉE
French BON
France

Coline Royaux
MNHN / PNDB – DATA TERRA / UAR DoHNÉE / French BON
France

Marie Jossé
CNRS / – DATA TERRA / French BON
France

Pauline Segueineau
MNHN / PNDB – DATA TERRA / UAR DoHNÉE / French BON
France

Arthur Barreau
SU / PNDB – DATA TERRA / UMR ISYEB / French BON
France

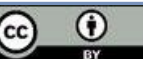
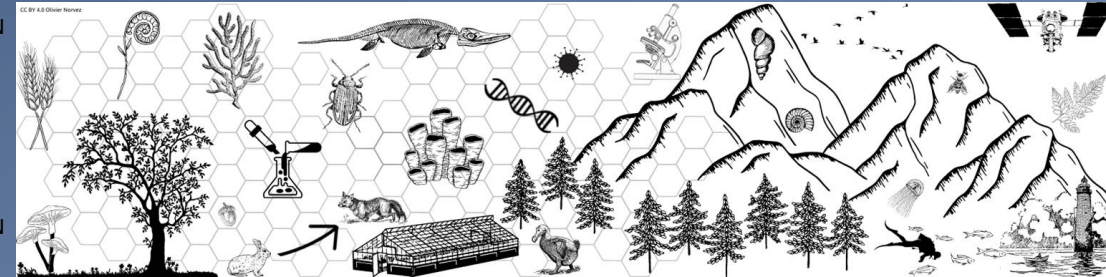
Dominique Pelletier
IFREMER / UMR DECOD / French BON
France

Olivier NORVEZ
CNRS / PNDB & Theia – DATA TERRA / French BON
France

Marc Eléaume
MNHN / UMR ISYEB / French BON
France

Björn Grüning
Freiburg Universität / Galaxy Europe
Germany

Jean-Baptiste Mihoub
SU / UMR CESCO / French BON
France



Context

Purpose depends on the origins



« Hier est derrière, demain est un mystère, mais aujourd'hui est un cadeau, C'est pour cela qu'on l'appelle présent ! »



<https://www.flickr.com/photos/27474697@N08/2556889078>



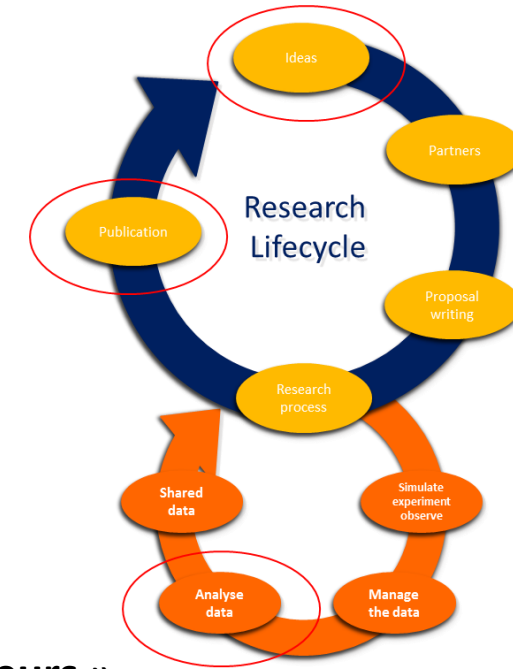
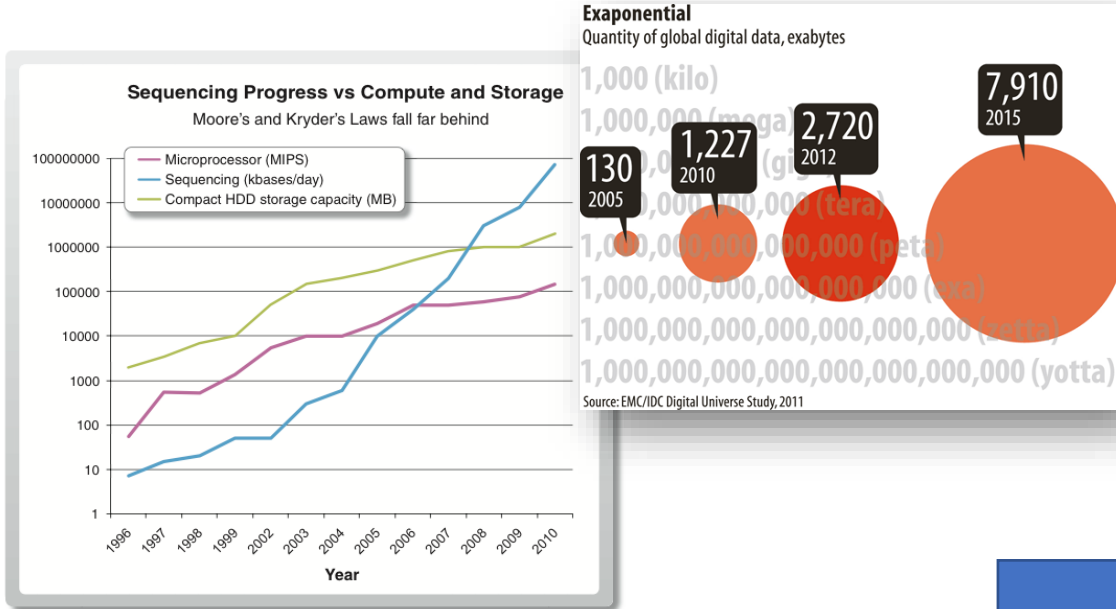
<https://www.vitav.fr/le-cercle-des-poetes-disparus-une-ode-a-la-liberte/>

« Carpediem »

Galaxy Ecology: Origins

Make a « dream » come true

From Datanami to citizen science



« 65 Millions d'observateurs »

A national project to allow citizen scientists take part to research lifecycle, not « only » as free data providers, but also to :

- Generate ideas/projects,
- Analyze data
- Generate a common knowledge

Make a « dream » come true

From Datanami to citizen science

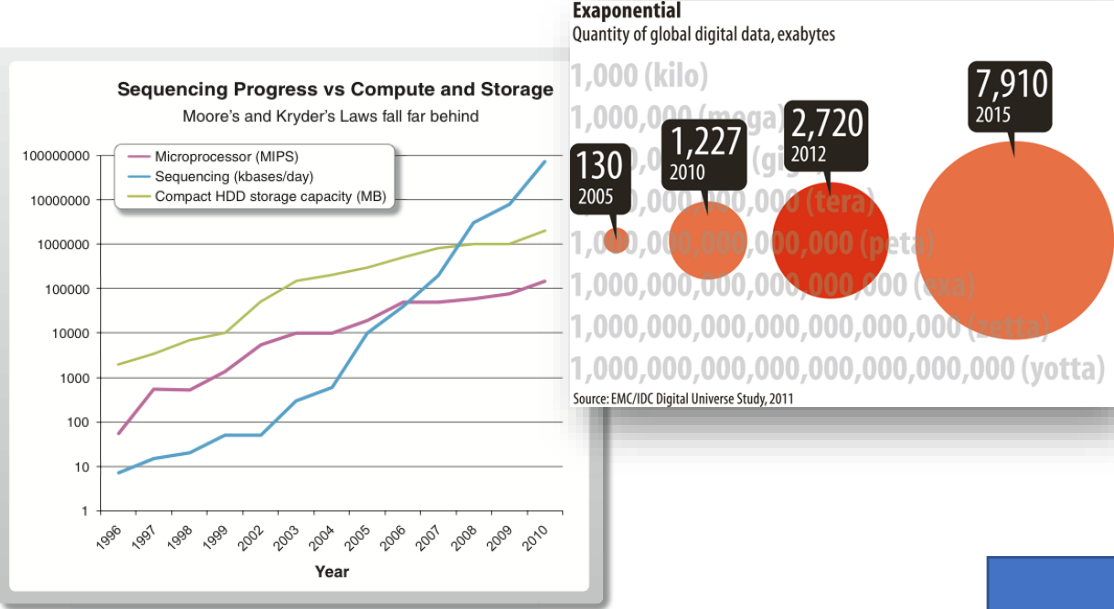
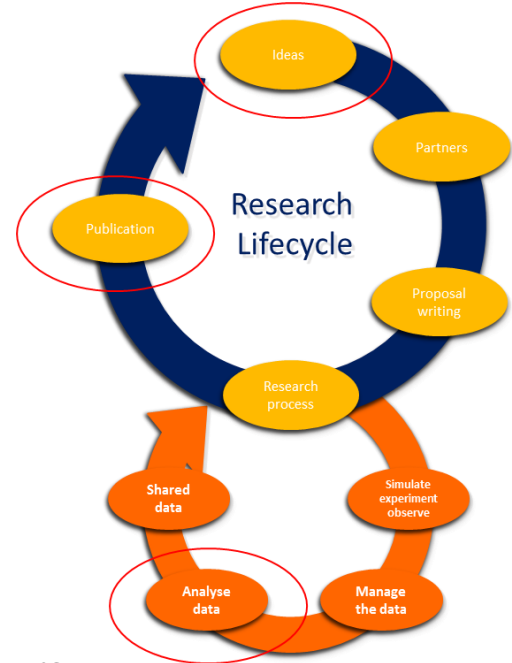


Fig. 1. A doubling of sequencing output every 9 months has outpaced and overtaken performance improvements within the disk storage and high-performance computation fields.

*“One of the **grand challenges of data-intensive science is to facilitate knowledge discovery** by assisting humans and machines in their **discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows.**”*

FORCE11 Future of Research Communication & e-Scholarship (2016)
The FAIR principles <https://www.force11.org/group/fairgroup/fairprinciples>



« 65 Millions d'observateurs »

A national project to allow citizen scientists take part to research lifecycle, not « only » as free data providers, but also to :

- Generate ideas/projects,
- **Analyse data**
- Generate a common knowledge

Make a « dream » come true

A collaborative analysis platform for citizen scientists

Goals:

- Reinforce research team with citizen to face the datanami
- Go outside the researcher's formalism
- Ameliorate global knowledge about environment to better protect

Findings:

- We need to give access to real scientific data and source codes
 - We need to capitalize on existing research e-infrastructure initiatives, tools and services
-

Make a « dream » come true

A collaborative analysis platform for citizen scientists

Goals:

- Reinforce research team with citizen to face the datanami
- Go outside the researcher's formalism
- Ameliorate global knowledge about environment to better protect

Findings:

- We need to give access to real scientific data and source codes
- We need to capitalize on existing research e-infrastructure initiatives, tools and services



2017: birth of Galaxy for Ecology initiative



Make a « dream » come true

A collaborative analysis platform for citizen scientists

Goals:

- Reinforce research team with citizen to face the datanami
- Go outside the researcher's formalism
- Ameliorate global knowledge about environment to better protect

Findings:

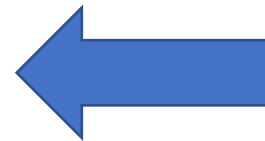
- We need to give access to real scientific data and source codes
- We need to capitalize on existing research e-infrastructure initiatives, tools and services



Essential brick of a new Biodiversity research e-infrastructure

Galaxy-E as a essential component of PNDB French Biodiversity data hub

2018: birth of PNDB french biodiversity e-infrastructure



Essential brick of a new Biodiversity research e-infrastructure

Galaxy-E as a essential component of PNDB French Biodiversity data hub

2018

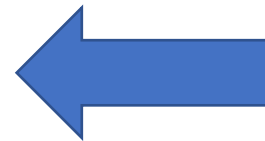
Advantages:

- Start rapidly, in TRL9, even with limited resources
- No need human resources for system administration
- No need IT resources
- « Just » develop components when needed / human resources funded
- Benefit from & contribute to an existing large community

« Duty »:

- Have a minimum quality process => best practices / FAIR principles so, also an advantage!

2018: birth of PNDB french biodiversity e-infrastructure



Essential brick of a new Biodiversity research e-infrastructure

Galaxy-E as a essential component of PNDB French Biodiversity data hub

2018



2018: birth of PNDB french biodiversity e-infrastructure



Essential brick of a new Biodiversity research e-infrastructure

Galaxy-E as a essential component of PNDB French Biodiversity data hub

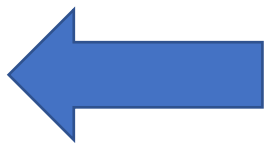
2018



First BON dev session for me
Just coming from Galaxy conference in USA
Discovering Bon in a box with Maria Cecilia, Mike, Henrique

2018: birth of PNDB french biodiversity e-infrastructure

=> Push forward the idea to have a subpart of Bon in a box giving access to tools, tutorials and computational workflows user can execute to test / reuse



Essential brick of a new Biodiversity research e-infrastructure

Galaxy-E as a essential component of PNDB French Biodiversity data hub

2018



Thank you!



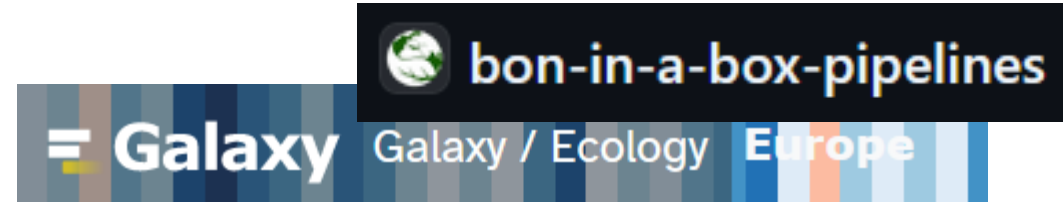
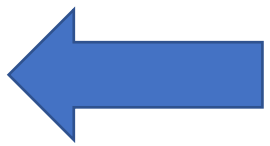
First BON dev session for me

Just coming from Galaxy conference in USA

Discovering Bon in a box with Maria Cecilia, Mike, Henrique

=> Push forward the idea to have a subpart of Bon in a box giving access to tools, tutorials and computational workflows user can execute to test / reuse

2018: birth of PNDB french biodiversity e-infrastructure



Galaxy Ecology meet EBV

Needs for supporting KMGBF



Representativeness

Sensitivity

Comparability

Integration

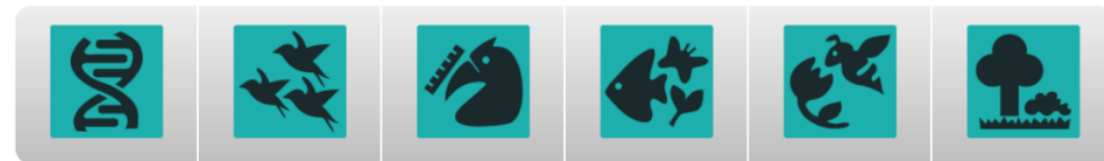
Accessibility
Automation
Reproducibility
Rigor



Needs for supporting KMGBF



Representativeness
Sensitivity
Comparability
Integration



Essential Biodiversity Variables

Accessibility
Automation
Reproducibility
Rigor



Needs for supporting KMGBF

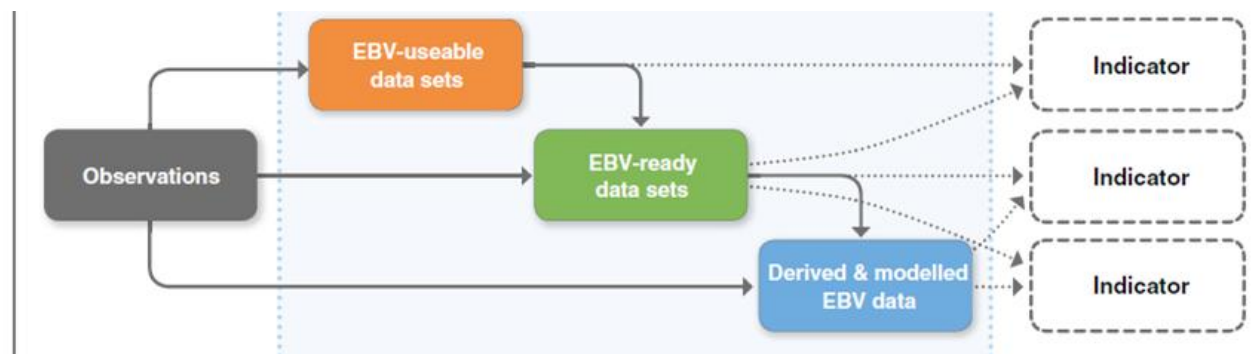


Representativeness
Sensitivity
Comparability
Integration

Kissling et al. 2018 Biol. Reviews

“for consistently **producing** and **replicating EBV data products** it would be advantageous to **develop** and **preserve all data access, integration and processing steps as an open-source and freely available service in a workflow-oriented e-infrastructure that supports curation, sharing and collaboration**”

EBV workflow e-infrastructure



Accessibility
Automation
Reproducibility
Rigor



Kissling et al. 2018 Biol. Reviews

Needs for supporting KMGBF

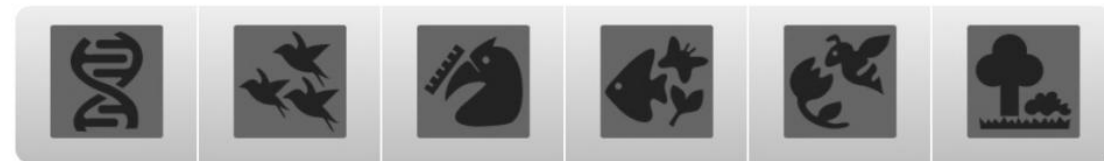


Representativeness

Sensitivity

Comparability

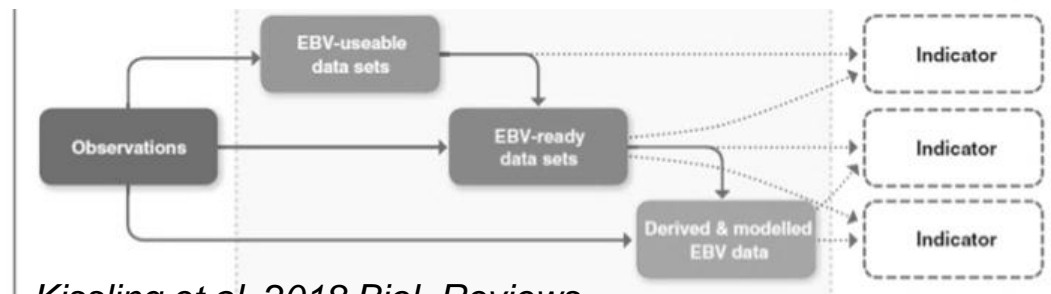
Integration



Kissling et al. 2018 Biol. Reviews

*“for consistently **producing** and **replicating EBV data products** it would be advantageous to **develop** and **preserve all data access, integration and processing steps as an open-source and freely available service in a workflow-oriented e-infrastructure that supports curation, sharing and collaboration**”*

 **Galaxy**
PROJECT



Kissling et al. 2018 Biol. Reviews

Accessibility
Automation
Reproducibility
Rigor



Galaxy



Web platform to share and process research data

Permit easy access to cloud analysis and High Performance Computing through interfacing of many programming languages

Four key principles

✓ Accessibility ✓ Reproducibility ✓ Transparency ✓ Peer review

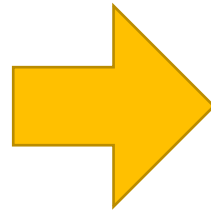
Galaxy



```


      ,direction="wide")
tab[is.na(tab)] <- 0
# filename <- "trouverUnNom"
# chemin <- paste(rep,filename,sep="/")
# write.table(tab, chemin)
colnames(tab) <- sub("nombre.", "", colnames(tab))
return(tab)
}
## sous jeux de donnees si choix d espece d annee ou d un pourcentage de carres
makeSousTab <- function(tab,vecSp=NULL,echantillon=1,
  methodeEchantillon="carre",vecannees=NULL) {
  cat(" -- Fabrication du sous jeu de donnees --\n")
  flush.console()
  ## reduction de la table à certaine espèces
  if(!is.null(vecSp)) {
    cat(" selection",length(vecSp),"espece(s):\n -> ")
    cat(vecSp)
    cat("\n")
    tab <- data.frame(carre = tab$carre,annee = tab$annee,tab[,vecSp])
    colnames(tab) <- c("carre","annee",vecSp)
  }
  ## reduction de la table pour certaines annees
  if(!is.null(vecannees)) {
    tab <- subset(tab,annee>=vecannees[1] & annee <= vecannees[2])
  }
  ## reduction de la table par une proportion de carre suivie
  if(echantillon != 1) {
    if(echantillon < 1 & echantillon >0) {
      nbinit <- nrow(tab)
      if(methodeEchantillon == "global") {
        nb <- round(nrow(tab)*echantillon)
        cat(" echantillonage",echantillon*100,
          "% des donnees par la methode",methodeEchantillon,"\n")
        cat("-> conservation de",nb,"lignes sur",nbinit,"\n")
        flush.console()
        tab <- tab[sample(1:nrow(tab))[1:nb],]
      } else {
        if (methodeEchantillon == "carre") {
          cat(" echantillonage",echantillon*100,
            "% des donnees par la methode",methodeEchantillon,"\n")
          flush.console()
          tab <- tab[sample(1:nrow(tab))[1:nb],]
        }
      }
    }
  }
}

```



Calculate community metrics calculate community metrics from abundance data (Galaxy Version 0.0.1)

Input file

 No tabular dataset available.

Observation data file, with location, year, species and abundance.

Choose the community metrics you want to compute

Select/Unselect all

Presence/absence, Species richness, Simpson and Shannon index are systematically computed.

Calculate community metrics from abundance data

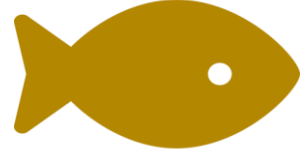
Galaxy Ecology: From a platform to best practices in Biodiversity data analysis

Applying "Galaxy-E" best practices to reuse!

Atomization & Generalization to reuse & transpose



GigaScience, 2025, 14, 1–12
DOI: 10.1093/gigascience/giae122
Review



PAMPA

Guidance framework to apply best practices in ecological data analysis: lessons learned from building Galaxy-Ecology

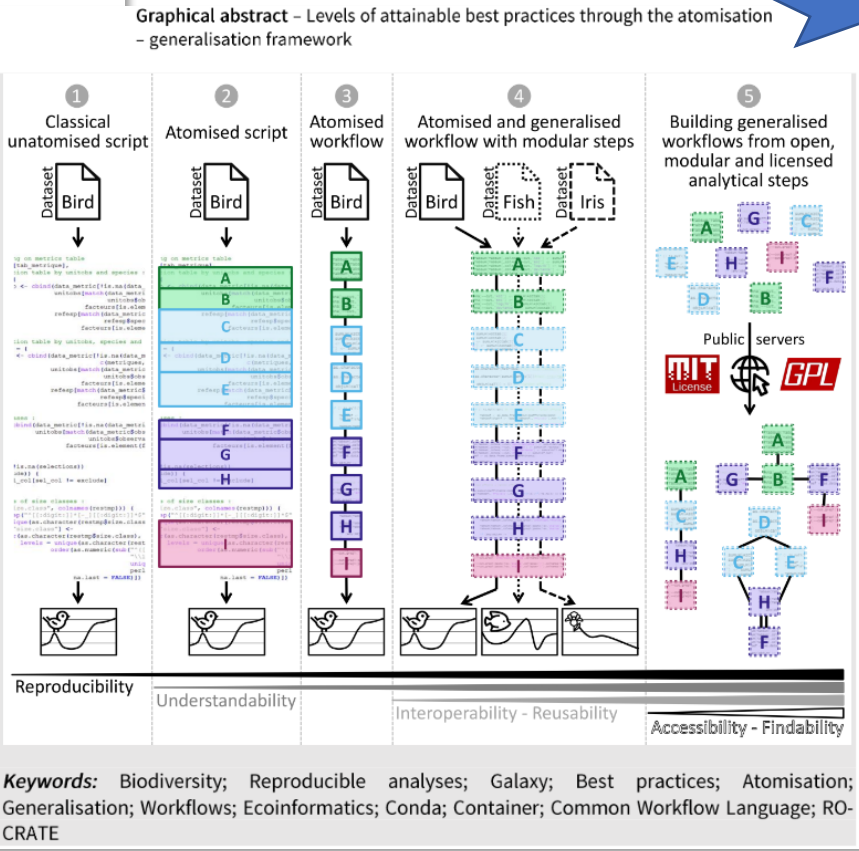
Coline Royaux^{1,2,*}, Jean-Baptiste Mihoub³, Marie Jossé⁴, Dominique Pelletier⁵, Olivier Norvez⁶, Yves Anne Fouilloux⁹, Helena Rasche¹⁰, Saskia Hiltmann¹¹, Bérénice Batut^{12,13}, Eléaume Marc^{14,15}, Pauli Guillaume Massé¹⁶, Alan Amossé¹⁷, Claire Bissery^{8,18}, Romain Lorrilliere³, Alexis Martin¹⁹, Yves Bas^{3,20}, Thimothée Virgoulay^{21,22}, Valentin Chambon¹⁷, Elie Arnaud², Elisa Michon²³, Clara Urfer^{2,24}, Eloïse Trigodet²¹, Gregoire Lois³, Romain Julliard³, Björn Grüning²⁵, Yvan Le Bras²⁶, and The Galaxy-E community

Guidance framework to apply best practices in ecological data analysis: Lessons learned from building Galaxy-Ecology

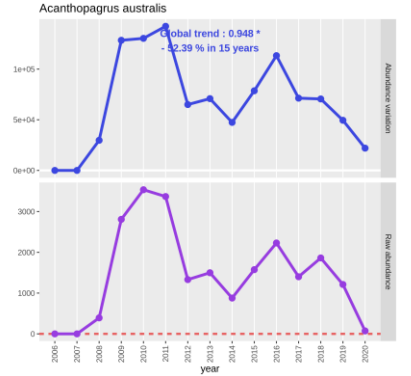


PCI recommendation: <https://doi.org/10.24072/pci.ecology.100694>

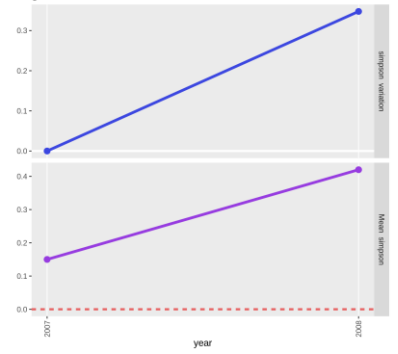
Royaux Coline^{1,2,*}, Mihoub Jean-Baptiste³, Jossé Marie⁴, Pelletier Dominique⁵, Norvez Olivier⁶, Reecht Yves^{7,8}, Fouilloux Anne⁹, Rasche Helena¹⁰, Hiltmann Saskia¹¹, Batut Bérénice^{12,13}, Eléaume Marc^{14,15}, Segueineau Pauline^{14,15}, Massé Guillaume¹⁶, Amossé Alan¹⁷, Bissery Claire^{8,18}, Lorrilliere Romain³, Martin Alexis¹⁹, Bas Yves^{3,20}, Virgoulay Thimothée^{21,22}, Chambon Valentin¹⁷, Arnaud Elie², Michon Elisa²³, Urfer Clara^{2,24}, Trigodet Eloïse^{21,24}, Delannoy Marie³, Lois Gregoire³, Julliard Romain³, Grüning Björn²⁵, The Galaxy-E community, Le Bras Yvan²



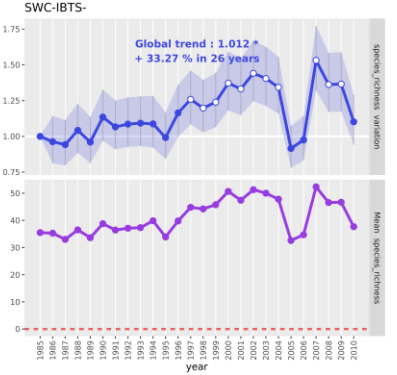
REEF LIFE SURVEY fishes data



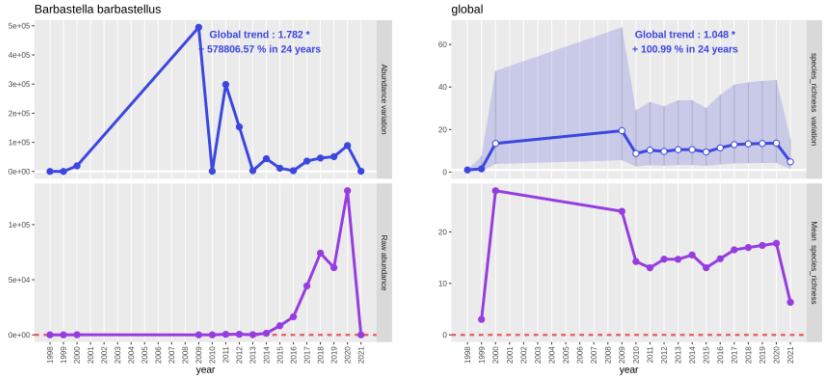
Marine Protected Areas (PAMPA) fishes data



DATRAS fishing data



VIGIE CHIRO bats data



Applying "Galaxy-E" best practices to reuse!

Different data, same primary variables, same workflow

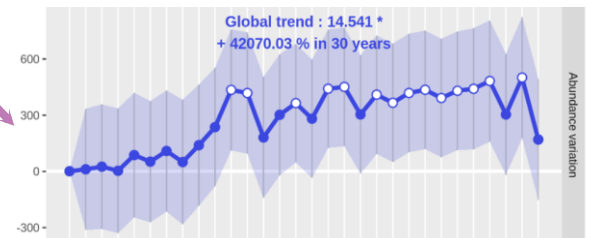
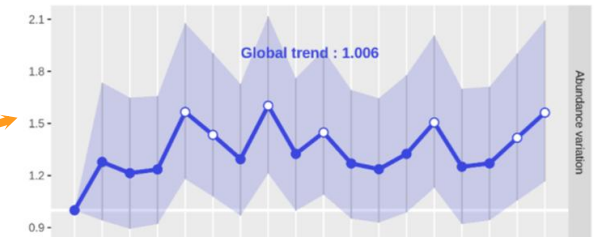
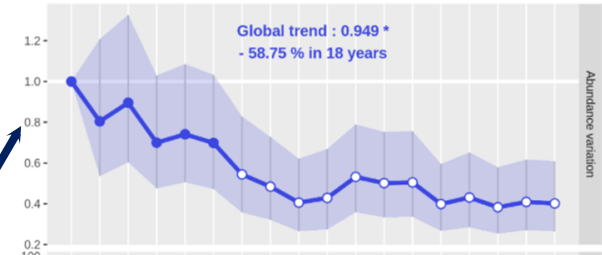
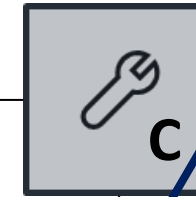
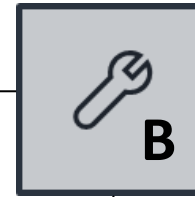
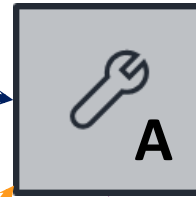
 Location

 Year

 Species

 Occurrence

	carre	annee	espece	abond
1	2	2016	ACCGEN	0
2	2	2017	ACCGEN	0
3	2	2018	ACCGEN	0
4	2	2019	ACCGEN	0
5	183	2016	ACCGEN	0
6	183	2017	ACCGEN	0



Survey	Year	Quarter	Area	AphiaID	Species	LngtClass	CPUE_number_per_hour
1	1991	1	22	126281	Anguilla anguilla	0	0.000000
2	1991	1	22	126281	Anguilla anguilla	720	0.009160
3	1991	1	22	126417	Clupea harengus	0	0.000000
4	1991	1	22	126417	Clupea harengus	80	0.075785
5	1991	1	22	126417	Clupea harengus	85	0.000000
6	1991	1	22	126417	Clupea harengus	95	0.000000
7	1991	1	22	126417	Clupea harengus	100	0.000000
8	1991	1	22	126417	Clupea harengus	105	0.012492
9	1991	1	22	126417	Clupea harenqus	110	0.618357



Unit	bs	rotation	codeSp	sexe	taille	classe_taille	poids	nb_ind
1	AS140155	3	Hemifasc	-999	-999	P	-999	1
2	AS140159	1	Nasosp.	-999	-999	P	-999	3
3	AS140159	3	Gompvari	-999	-999	P	-999	1
4	AS140160	3	Gompvari	-999	-999	P	-999	1



Applying “Galaxy-E” best practices to reuse!

Different data, same primary variables, same workflow

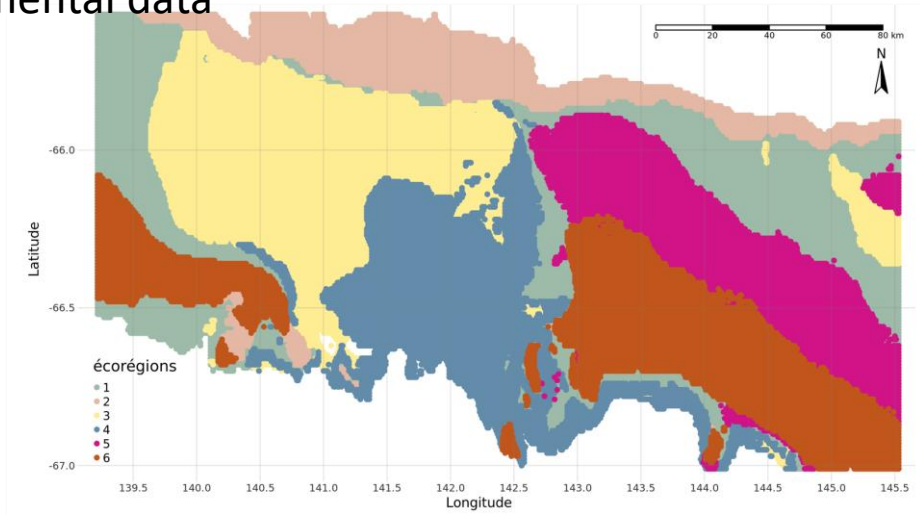
2024

Transpose to other taxonomic and/or geographical and/or temporal context

From Antarctica to North East Atlantic

From Invertebrates to Fishes

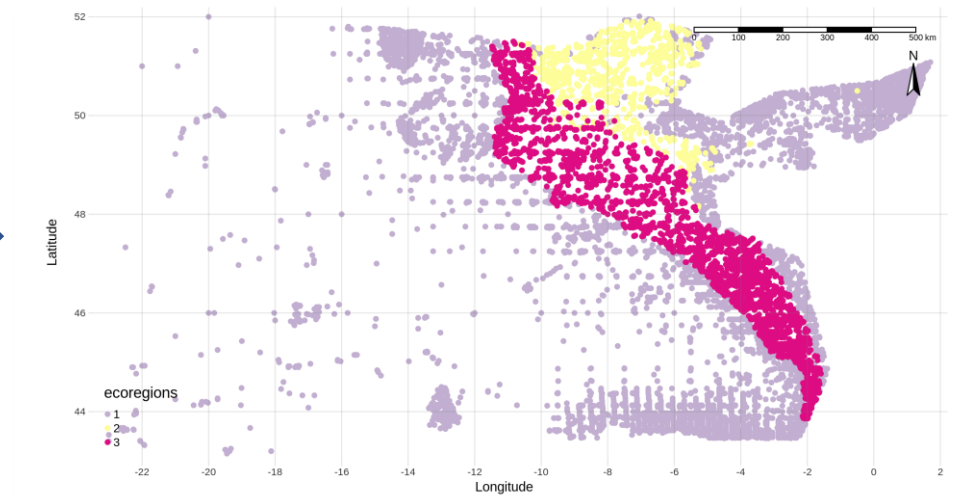
Project data (invertebrates)
Project environmental data



North Atlantic

OBIS marine occurrence data (fishes)
OBIS marine environmental data (all)

c1: decimallongitude x c2: decimallatitude x c5: shoredistance x c6: bathymetry x c7: sst x c8: sss x



Applying “Galaxy-E” best practices to reuse!

Different data, same primary variables, same workflow

2024

Test several scenarios

From Antarctica to North East Atlantic

From Invertebrates to Fishes

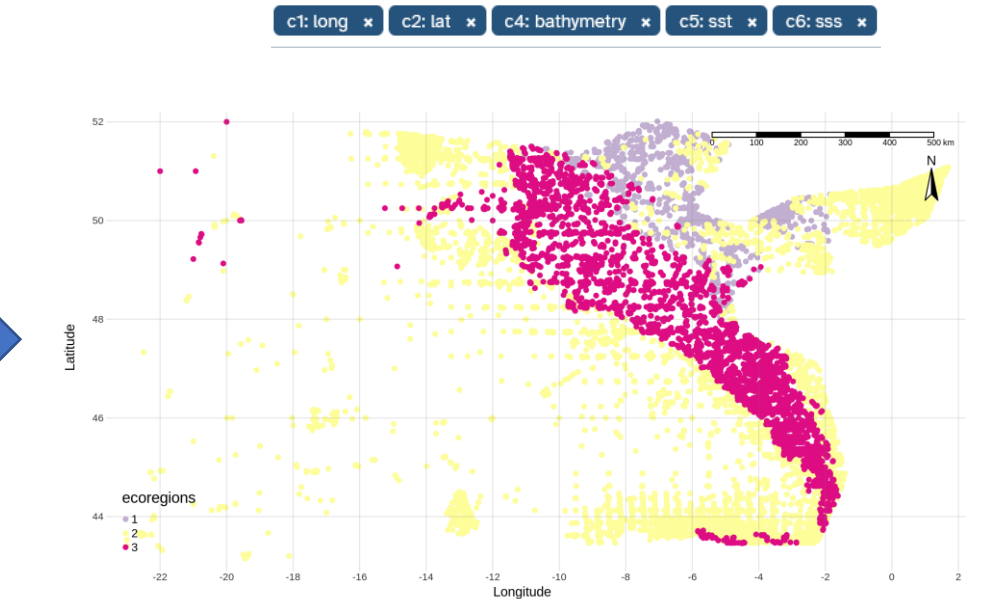
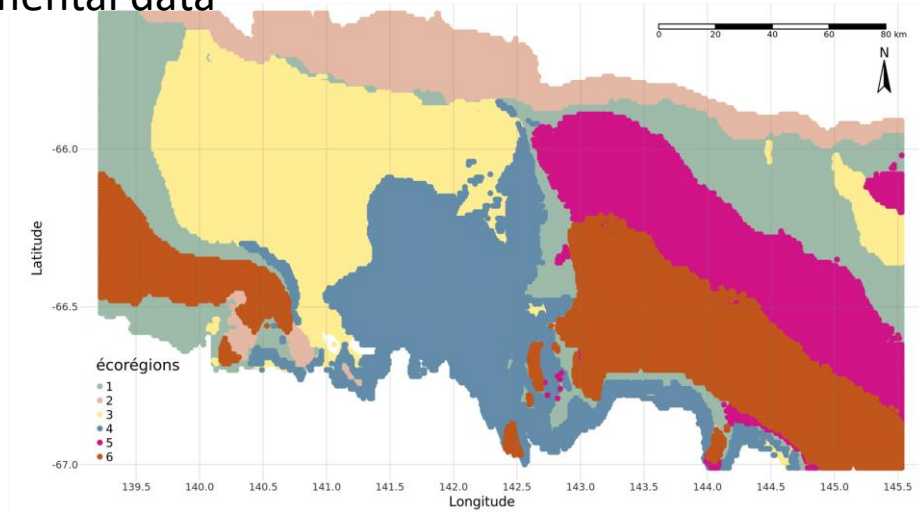
Modifying modelling parameters

North Atlantic

OBIS marine occurrence data (fishes)

OBIS marine environmental data (sst / bathy / sss)

Project data (invertebrates)
Project environmental data



Training tutorials for Ecology



Data and Metadata Management

These tutorials are focusing on data and metadata management in Ecology.

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
Cleaning GBIF data using OpenRefine biodiversity					
Creating FAIR Quality assessment reports and draft of Data Papers from EML metadata with MetaShRIMPS Metadata EML FAIR Data Paper biodiversity					
Creating metadata using Ecological Metadata Language (EML) standard with EML Assembly Line functionalities Metadata EML Ecology Biodiversity FAIR Data Paper					
Data submission using ENA upload Tool					

Data access

These lessons focus on ways to access data classically used in Ecology

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
QGIS Web Feature Services earth-system GIS Geographical Information System WFS Spatial data Maps OGC biodiversity					

Data preprocessing

These lessons focus on manners to preprocess data used in Ecology

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
Biodiversity data exploration taxonomic data data quality biodiversity					
Checking expected species and contamination in bacterial isolate illumina bacteria microgalaxy					
Cleaning GBIF data for the use in Ecology gbif data management data cleaning biodiversity					

Data visualization

These tutorials showcase data visualization in Ecology

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
Visualization of Climate Data using NetCDF xarray Map Plotting paragon					
Visualize EBV cube data with Panoply netCDF viewer interactive-tools EBV cube Data visualization					

Data analysis

These lessons focus on ways to analyse data in Ecology

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
Champs blocs indicators Ecosystem EBV class EBV dataset EBV workflow Marine ecosystems biodiversity					
Compute and analyze biodiversity metrics with PAMPA toolsuite Species population EBV class Community composition EBV class EBV dataset EBV workflow modeling biodiversity					
Ecoregionalization workflow tutorial ecology taxonomic data EBV workflow modeling gbif ocean earth-system interactive-tools biodiversity					
From NDVI data with OpenEO to time series visualisation with Holoviews earth-system land degradation NDVI copernicus holoviews					
Life Traits Ecoregionalization workflow biodiversity					
Marine Omics identifying biosynthetic gene clusters earth-system ocean marine omics biodiversity					
Metabarcoding/eDNA through Obftools Genetic composition EBV class Community composition EBV class EBV dataset EBV workflow eDNA Metabarcoding biodiversity					
Obis marine indicators earth-system ocean marine omics biodiversity					
Phylodiversity analysis quick tutorial ecology taxonomic data					
Preparing genomic data for phylogeny reconstruction phylogeny data handling functional annotation biodiversity					
RAD-Seq Reference-based data analysis RAD-seq Genetic composition EBV class Species population EBV class EBV dataset EBV workflow biodiversity					
RAD-Seq de-novo data analysis RAD-seq Genetic composition EBV class Species population EBV class EBV dataset EBV workflow biodiversity					
RAD-Seq to construct genetic maps RAD-seq Genetic composition EBV class EBV dataset EBV workflow biodiversity					
Regional GAM Species population EBV class Species traits EBV class EBV dataset EBV workflow biodiversity					
Sentinel 2 biodiversity Remote sensing biodiversity					
Species distribution modeling interactive-tools modeling gbif species population EBV class biodiversity					
Taxonomic Analysis of eDNA biodiversity					

**French BON EBV
operationalization pilot**

Indicators need transparency, accessibility, traceability and reproducibility

Galaxy Ecology

Invoked Workflow: Workflow for representing State of the Environment in Antarctic (Version: 145)

112 of 128 steps successfully scheduled...

22 of 32 jobs complete (total number of jobs will change until all steps fully scheduled).

History

Workflow for representing State of the Environment in Antarctic

- 47: Interactive JupyterLab Notebook on data 39
- 46: Executed Jupyter Notebook
- 45: Jupyter output collection
- 44: API.txt
- 43: Executed Jupyter Notebook
- 42: Jupyter output collection
- 41: Collapse Collection on data 3
- 40: Collapse Collection on data 3
- 39: Collapse Collection on data 2
- 38: Collapse Collection on data 2

Galaxy Ecology

Create a plot from GLM data as temporal trend (Galaxy Version 0.0.2)

Tool Parameters

Input glm results file *

56: GLM - Results from your population analysis on data 36 and data 29

Metrics data table used for glm *

29: Calculate presence absence table on data 24

Unitobs table used for glm *

36: Regex Find And Replace on data 35

Additional Options

Email notification

Attempt to re-use jobs with identical parameters?

History: Compute and analyze biodiversity metrics with PAMPA toolsuite

From <https://training.galaxyproject.org/training-material/topics/ecology/tutoria...>

56: GLM - Results from your population analysis on data 36 and data 29

#Concatenate x #EVHOC x #unitobs x

6 lines

format tabular, génome de référence ?

Error in .Call("FreeADFunObject", ptr, PACKAGE = DLL): "FreeADFunObject" not available for .Call() for

1	2	3
"analysis"	"interest.var"	"distr"
"Gadus morhua"	"number"	"gauss"
"Leptidionthombus whiffiagonis"	"number"	"gauss"
"Lopholissius budgassus"	"number"	"gauss"
"Lopholissius piscatorius"	"number"	"gauss"

45: Report

Sabina Leonelli, Living Data 2025 22th october:

« It is essential to track data provenance and data processing supporting knowledge »

« we are continuing creating a lot of tools to go very fast but unfortunately not to go very far »

> [Nucleic Acids Res. 2024 Jul 5;52\(W1\):W83-W94. doi: 10.1093/nar/gkac410.](https://doi.org/10.1093/nar/gkac410)

The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update

Galaxy Community

Collaborators + expand

PMID: 38769056 PMID: [PMC11223835](https://pubmed.ncbi.nlm.nih.gov/38769056/) DOI: [10.1093/nar/gkac410](https://doi.org/10.1093/nar/gkac410)

Indicators need transparency, accessibility, traceability and reproducibility

Galaxy Ecology interface showing a workflow titled "Workflow for representing State of the Environment in Antarctic (Version: 145)". The workflow is displayed as a series of interconnected steps in a grid. The top bar indicates "Invoked Workflow: Workflow for representing State of the Environment in Antarctic (Version: 145)" and "Invoked 3 minutes ago". A progress bar shows "112 of 128 steps successfully scheduled..." and "22 of 32 jobs complete (total number of jobs will change until all steps fully complete)". The interface includes tabs for Overview, Steps, Inputs, Outputs, Report, Export, and Metrics. A sidebar on the left contains navigation options like Interactive Tools, Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History Multiview, Datasets, Pages, and Libraries.

Galaxy Ecology interface showing the "Compute GLM on population data" tool parameters. The tool is identified as "Compute a GLM of your choice on population data (Galaxy Version 0.0.2)". The "Tool Parameters" section includes: "Input metrics file" (27: Calculate presence absence table on data 22), "Unitobs informations file" (36: Regex Find And Replace on data 35), "Response variable from metrics file" (c4: "number"), "Explanatory variables" (Year, Site), and "Random effect" (Site). A "Specify advanced parameters" section is at the bottom with the option "No, use program defaults." A dropdown menu shows "Selected 0.0.2" and "Switch to 0.0.1".

Photograph of a presentation slide titled "All Hands 2018" with the GEO BON logo. The slide content includes:

- Traceability / provenance of source data and EBV / documentation of production protocols
- Feedback from WGs and BONs about how to optimize EBV reporting
- Brainstorm on Minimum Information Standards components
- Endorsement process for streaming EBVs via the data portal

A water bottle is visible in the bottom right corner of the slide image.

Partial screenshot of a Galaxy interface sidebar showing the user "yebbras" and various tool options like "diversity metri...", "LM a 29", "unitobs", "Run Job Again", and "ce ?".

« we are continuing creating a lot of tools to go very fast but unfortunately not to go very far »

> Nucleic Acids Res. 2024 Jul 5;52(W1):W83-W94. doi: 10.1093/nar/gkae410.

The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update

Galaxy Community

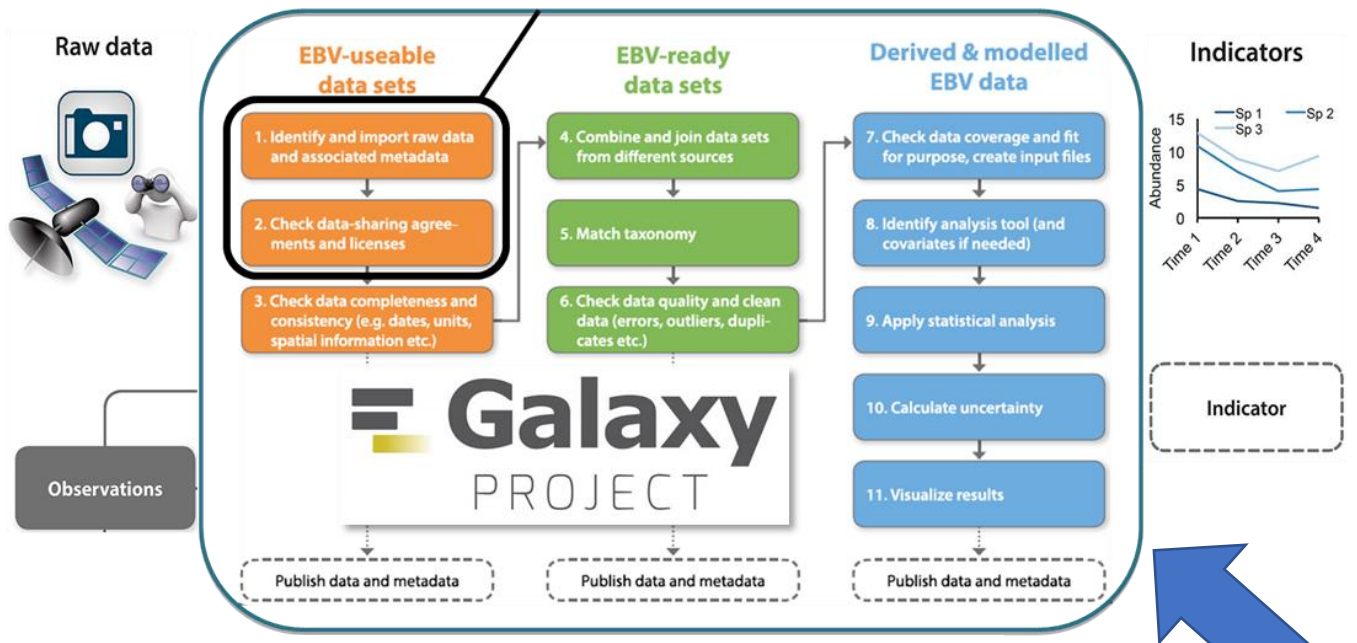
Collaborators + expand

PMID: 38769056 PMID: PMC11223835 DOI: 10.1093/nar/gkae410

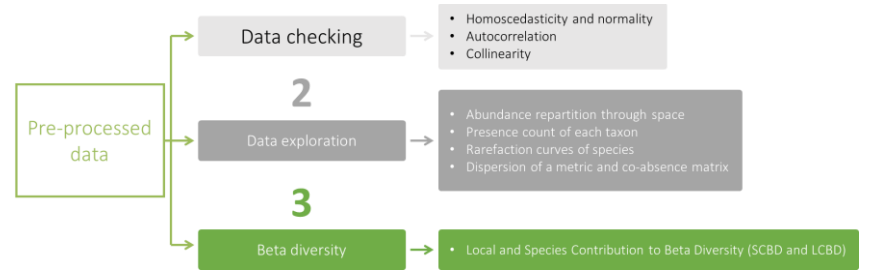
Essential brick to take part to sizing initiatives

From PNDDB to french participation to GEO BON

2019



French BON EBV operationalization pilot



Adapted from Kissling *et al.* 2018 Biol. Reviews

Biodiversity data exploration

Authors: Olivier Norvez, Marie Josse, Coline Royaux, Yan Le Bras

Questions:

- How to explore biodiversity data?
- How to look at Homoscedasticity, normality or collinearity of presences-absence or abundance data?
- How to compare beta diversity taking into account space, time and species components?

Objectives:

- Explore Biodiversity data with taxonomic, temporal and geographical informations
- Have an idea about quality content of the data regarding statistical tests like normality or homoscedasticity and coverage like temporal or geographical coverage

Requirements:

- Introduction to Galaxy Analyses

Time estimation: 1 hour

Supporting Materials: Datasets, Workflows, Recordings, Available on these Galaxies

Last modifications: Oct 28, 2022

License: Tutorial Content is licensed under Creative Commons Attribution 4.0 International License The GTN Framework is licensed under MIT



Galaxy Training! Ecology Learning Pathways Help Extras Search Tutorials

Compute and analyze biodiversity metrics with PAMPA toolsuite

Authors: Coline Royaux, Yan Le Bras, Jean-Baptiste Mihoub

Questions:

- How to evaluate properly populations and communities biological state with abundance data?
- How does trawl exploited populations of Baltic sea, Southern Atlantic and Scotland are doing over time?
- How to compute and analyze biodiversity metrics from abundance data?

Objectives:

- Upload data from DATRAS portal of ICES
- Pre-process population data with Galaxy
- Learning how to use an ecological analysis workflow from raw data to graphical representations
- Learning how to construct a Generalized Linear (Mixed) Model from a usual ecological question
- Learning how to interpret a Generalized Linear (Mixed) Model

Essential brick to take part to sizing initiatives

From PNDB to french participation to GEO BON

- Nucleotidic diversity (π)
- Frequency of the prevalent allele in the population (P)
- Polymorphic sites, variant sites and polymorphic loci percentage
- Haplotypic diversity (H)
- Gene diversity
- Haplotype count
- Genotype count
- Maximum percentage of the most frequent genotype
- Genotype frequency
- Inbreeding coefficient for each loci (F)
- Between-individuals differentiation coefficient
- Genetic differentiation between sub-populations (F_{st})
- Orthogonal measure of locus differentiation (D_{est})
- Absolute measure of locus differentiation (D_{xy})

Genetic
composition



Coline Royaux



About This Workflow

Intraspecific genetic diversity, genetic differentiation and inbreeding EBV workflow - Version 6

Author

coline_royaux



All published Workflows by coline_royaux

Creators

Coline Royaux

Description

EBV workflow corresponding to intraspecific genetic diversity, genetic differentiation estimation in a population and a population map

Tags

#EBV #GeneticComp

License

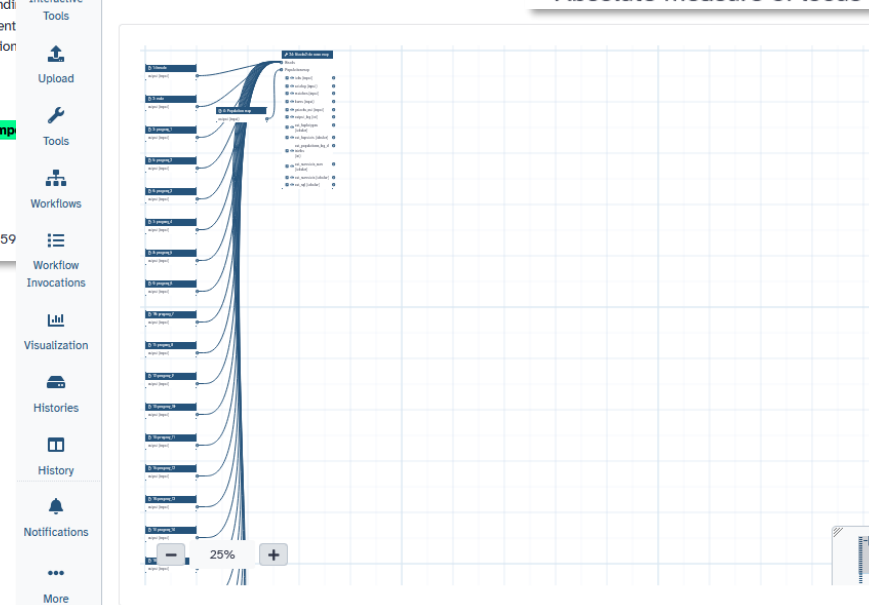
MIT License

Last Updated

Monday May 27th 12:12:59

Galaxy Ecology

Workflow Preview



About This Workflow

Intraspecific genetic diversity and inbreeding EBV workflow - Version 6

Author

coline_royaux



All published Workflows by coline_royaux

Creators

Coline Royaux

Description

EBV workflow corresponding to intraspecific genetic diversity and inbreeding estimation in a population based on sequencing data and a population map

Tags

#EBV #GeneticComposition

License

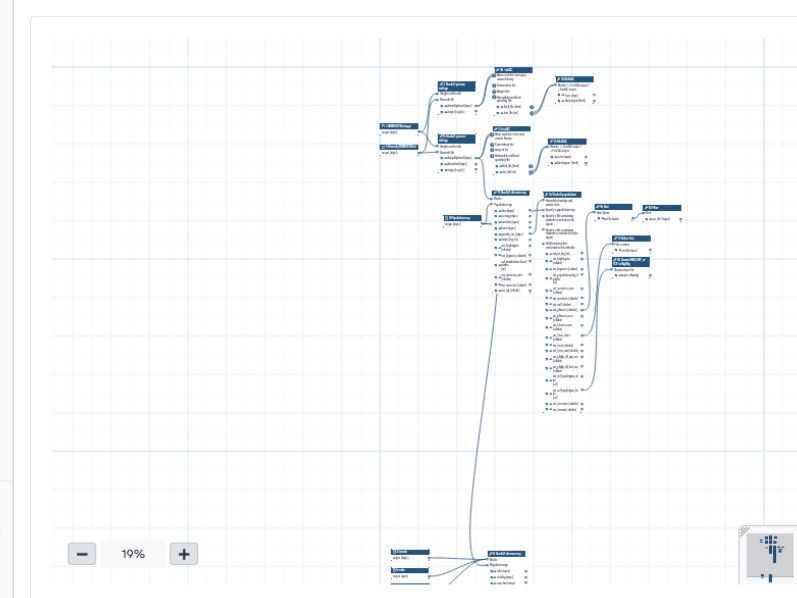
MIT License

Last Updated

Monday May 27th 12:42:11 2024 GMT+2

Galaxy Ecology

Workflow Preview



<https://ecology.usegalaxy.eu/published/workflow?id=b193dcf37dcc8a84>

Other tool list

- Amova
- Homova
- Mantel

<https://ecology.usegalaxy.eu/published/workflow?id=e98615eca9019d6f>

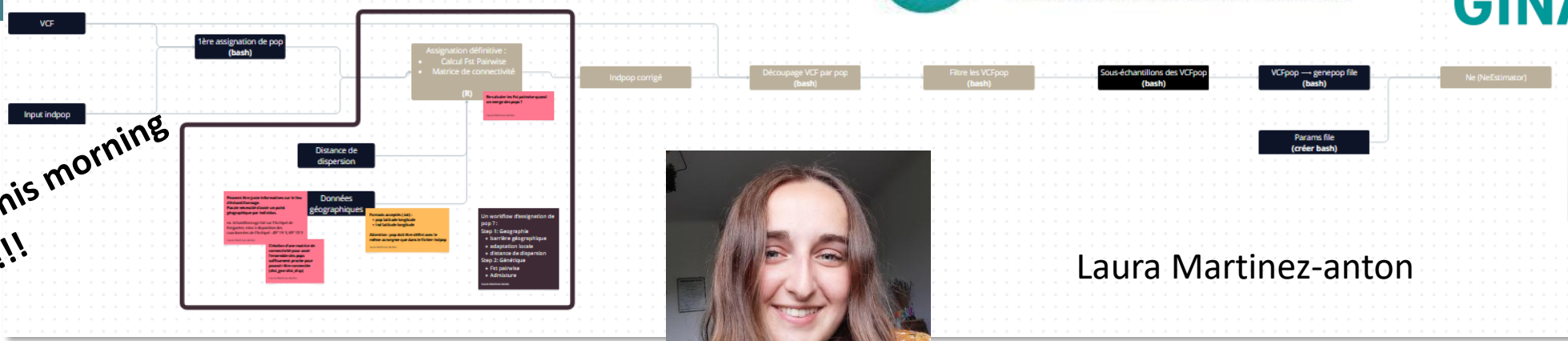
Essential brick to take part to sizing initiatives



From PNDP to french participation to GEO BON

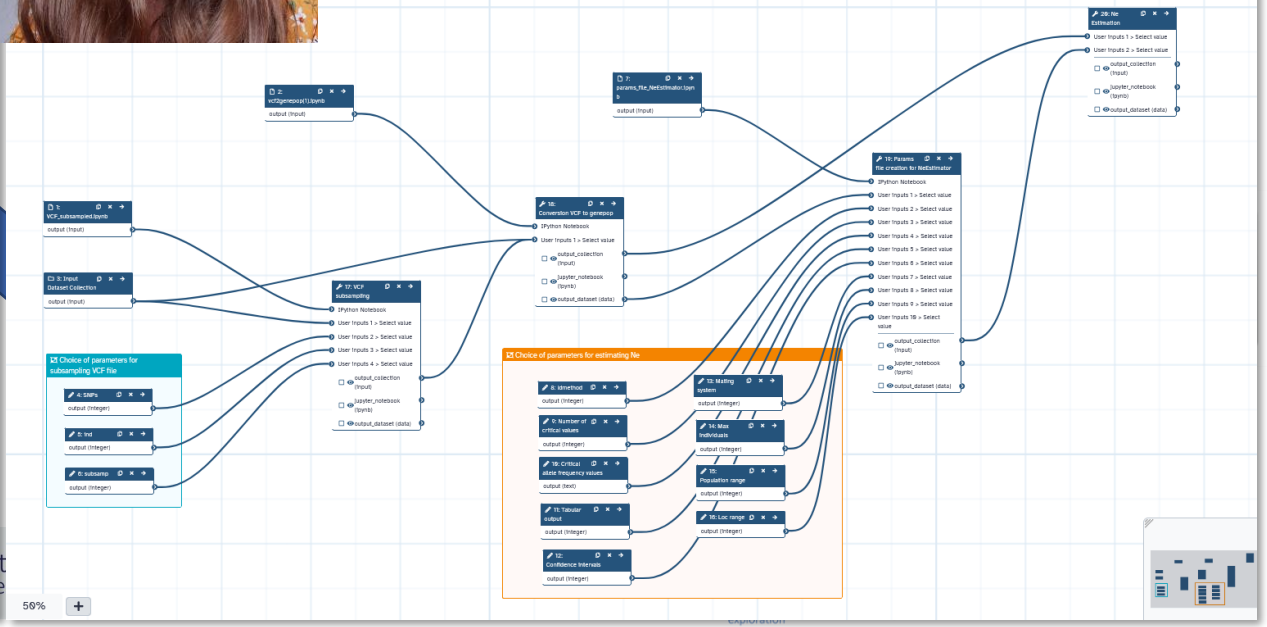
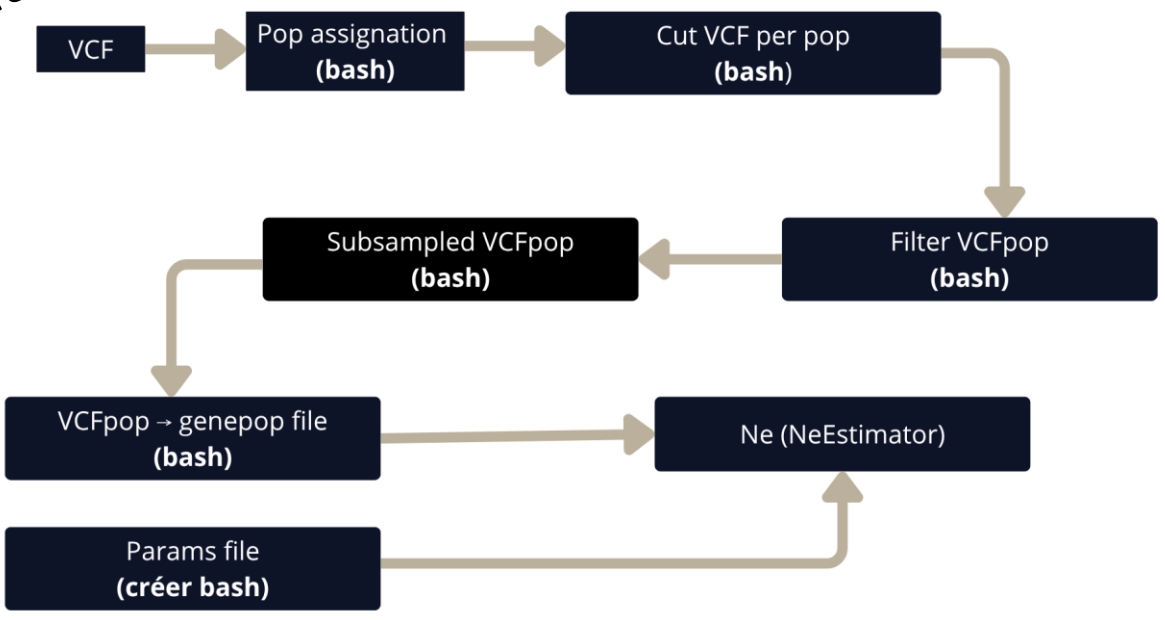
NE500

Breaking news (from this morning meeting)!!!



Laura Martinez-anton

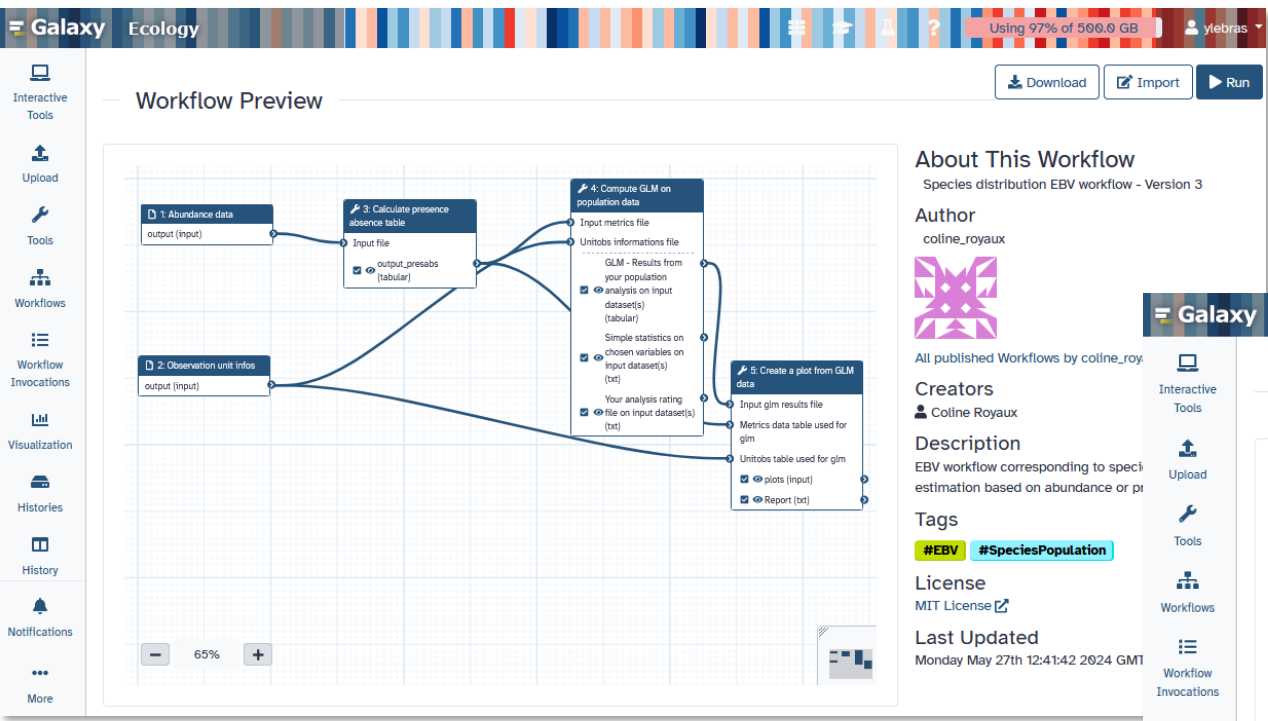
Genetic composition



Essential brick to take part to sizing initiatives

From PNDB to french participation to GEO BON

- Presence-absence
- Abundance



Workflow Preview

1. Abundance data (output: input)

2. Observation unit infos (output: input)

3. Calculate presence absence table (input: input file, output: output_presabs (tabular))

4. Compute GLM on population data (input: input metrics file, Untobis Informations file, GLM - Results from your population analysis on input dataset(s) (tabular), Simple statistics on chosen variables on input dataset(s) (txt), Your analysis rating file on input dataset(s) (txt), output: input glm results file, Metrics data table used for glm, Untobis table used for glm, plots (input), Report (txt))

5. Create a plot from GLM data (input: input glm results file, Metrics data table used for glm, Untobis table used for glm, plots (input), Report (txt))

About This Workflow
Species distribution EBV workflow - Version 3

Author
coline_royaux

License
MIT License

Last Updated
Monday May 27th 12:41:42 2024 GMT

Coline Royaux Dominique Pelletier

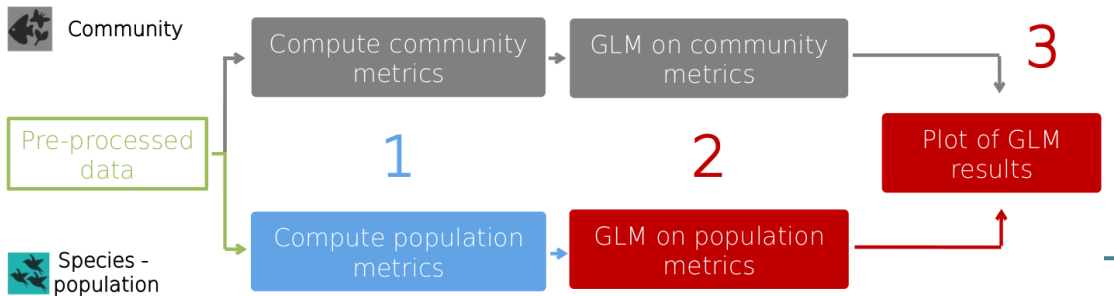
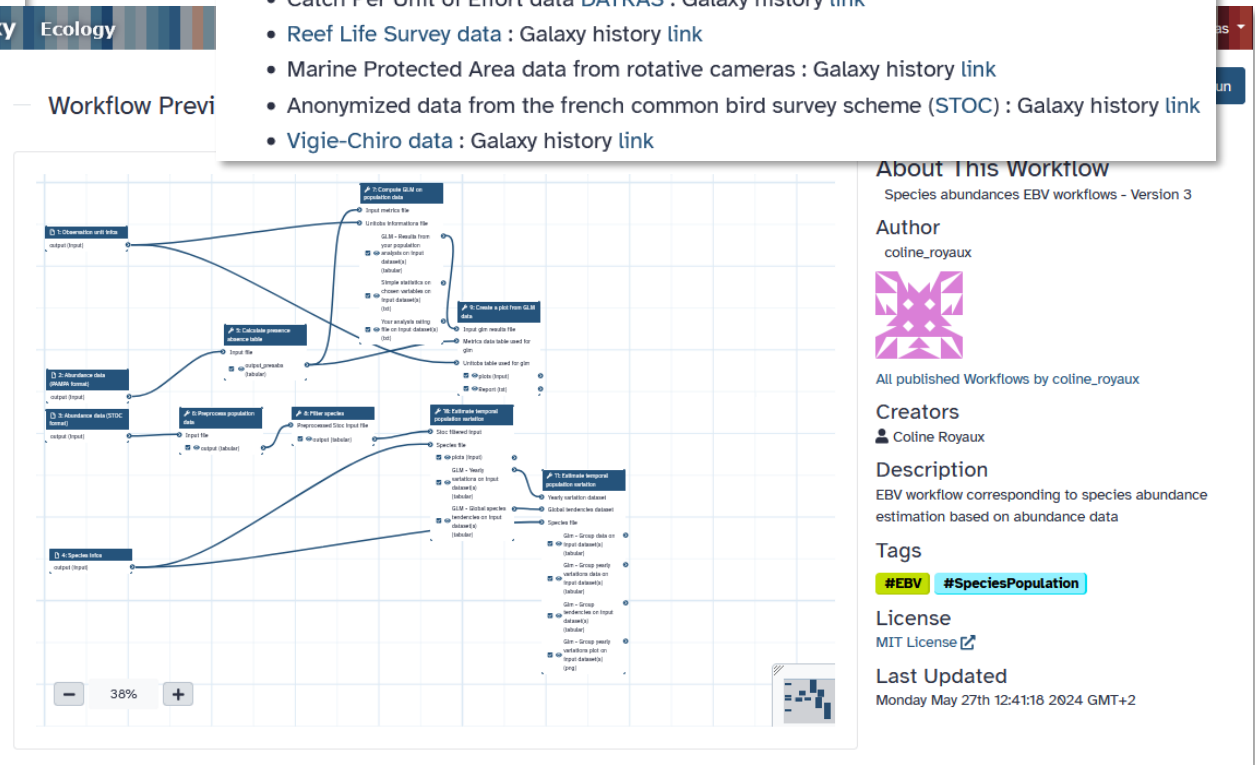


- Species distribution
- Species abundances

Use-cases of the workflow on several data types

- Catch Per Unit of Effort data DATRAS : Galaxy history link
- Reef Life Survey data : Galaxy history link
- Marine Protected Area data from rotative cameras : Galaxy history link
- Anonymized data from the french common bird survey scheme (STOC) : Galaxy history link
- Vigie-Chiro data : Galaxy history link

<https://ecology.usegalaxy.eu/published/workflow?id=4337a20517b47787>

Workflow Preview

1. Abundance data (output: input)

2. Observation unit infos (output: input)

3. Calculate presence absence table (input: input file, output: output_presabs (tabular))

4. Compute GLM on population data (input: input metrics file, Untobis Informations file, GLM - Results from your population analysis on input dataset(s) (tabular), Simple statistics on chosen variables on input dataset(s) (txt), Your analysis rating file on input dataset(s) (txt), output: input glm results file, Metrics data table used for glm, Untobis table used for glm, plots (input), Report (txt))

5. Create a plot from GLM data (input: input glm results file, Metrics data table used for glm, Untobis table used for glm, plots (input), Report (txt))

About This Workflow
Species abundances EBV workflows - Version 3

Author
coline_royaux

License
MIT License

Last Updated
Monday May 27th 12:41:18 2024 GMT+2

<https://ecology.usegalaxy.eu/published/workflow?id=179ffb0aaf189739>

Essential brick to take part to sizing initiatives

From PNDDB to french participation to GEO BON

- Presence-absence
- Abundance

The screenshot shows the Galaxy Ecology interface with a workflow preview. The workflow consists of several steps: 1. data_FrenchBBS.tabular, 2. coordCarreSTOCfaux.tabular, 3. tabSpecies.csv, 4. species_indicateur_fonctionne.Ltabular, 5. Preprocess population data, 6. Filter species, 7. Temporal trend indicator, and 8. Estimate temporal population evolution. The interface includes a sidebar with navigation options like Interactive Tools, Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, and History. The main area shows the workflow steps connected by arrows, with input and output files for each step. On the right, there is an 'About This Workflow' section with details about the author (yiebras), description, tags, license, last updated date, and sharing options.

<https://ecology.usegalaxy.eu/published/workflow?id=01274dfdbdfcaa88>

Benjamin Yguel



- Species distribution
- Species abundances

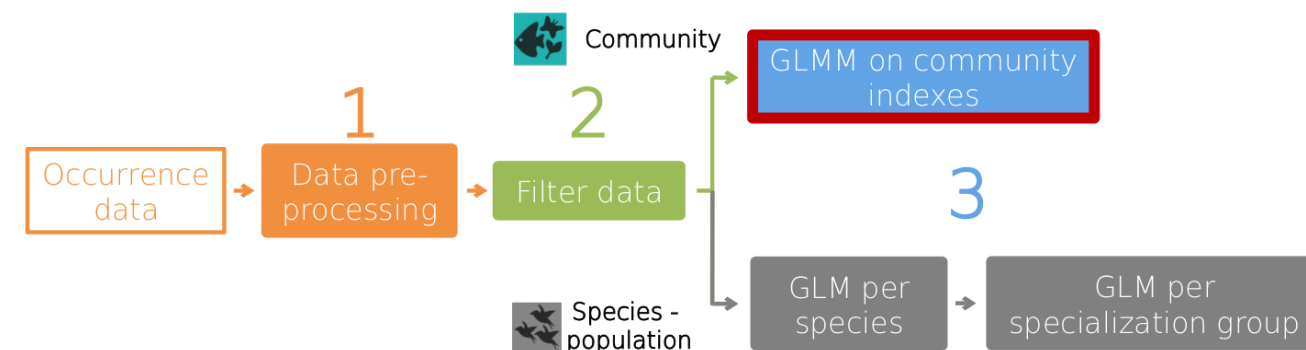
Species populations



STOC workflow

This workflow permits, from occurrence data of one or several species population(s) and diverse informations on species, to compute a Generalized Linear Model (GLM) estimating the effects of year and site on the abundance of each species and for each specialization group (generalist, forestry, ...). It also permits to get a plot of the temporal variation of the number of sites where the species has been observed and of the abundance of the species (raw and estimated by the model for each species and specialization group).

This branch of the main workflow



Essential brick to take part to sizing initiatives

From PNDB to french participation to GEO BON

Galaxy Ecology

Workflow Preview

1: Butterfly abundance data (output: input)

2: Flight curve (output: tabular)

3: Abundance index (output: tabular)

4: Scatterplot w ggplot2 (output: png)

5: Expected trend (output: png)

About This Workflow
Phenology EBV workflow - Version 3

Author
coline_royaux

Creators
Coline Royaux

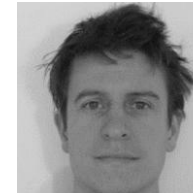
Description
EBV workflow corresponding to butterfly flight phenology estimation based on abundance data

Tags
#EBV #SpeciesTraits

License
MIT License

Last Updated
Monday May 27th 12:45:13 2024 GMT+2

Simon Bénateau



Elisa Michon

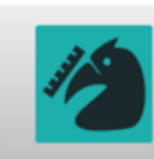


Clara Urfer



- Morphology
- Physiology
- Phenology :
- Mobility

Species traits



Regional GAM workflow | Tutorial

This workflow permits, from occurrence data of a single species population, to compute an abundance index that will be used to modelize and represent in a plot the flight phenology of a butterfly species through time.

Regional GAM

Author(s) Clara Urfer, Elisa Michon, Yan Le Bras, Bérénice Batut, Simon Benateau

Reviewers

Overview

Questions:

- What are abundance and trends of a butterfly species?

Objectives:

- Obtain and filter/manipulate occurrence data
- Compute and visualize phenology of a species through the years
- Compute temporal abundance trends

Requirements:

- Introduction to Galaxy Analyses

Time estimation: 2 hours 30 minutes

Supporting Materials:

Datasets, Workflows, FAQs

Published: Jul 4, 2019

Last modification: Jun 3, 2025

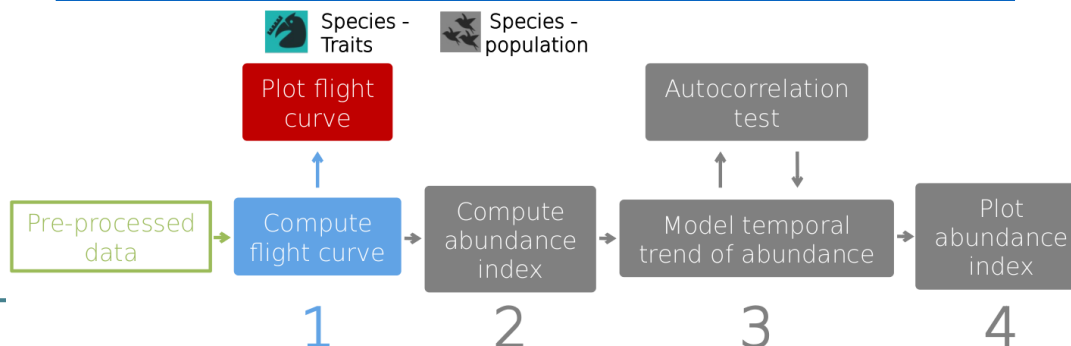
License: Tutorial Content is licensed under Creative Commons Attribution 4.0 International License. The GTN Framework is licensed under MIT

PURL: <https://gxy.io/GTN:T00134>

Rating: 5.0 (0 recent ratings, 1 all time)

Revision: 18

<https://ecology.usegalaxy.eu/published/workflow?id=5708413da63185ed>



<https://training.galaxyproject.org/training-material/topics/ecology/tutorials/regionalGAM/tutorial.html>

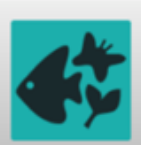
Essential brick to take part to sizing initiatives



From PNDP to french participation to GEO BON

- alpha and gamma diversity
 - Species richness
 - Shannon index
 - Simpson index
 - Hill index
- Pielou index

Community composition



- Community abundance : Galaxy Wor
- Taxonomic / phylogenetic diversity :
- Traits diversity : Galaxy Workflow
- Interactions diversity : Galaxy Workf

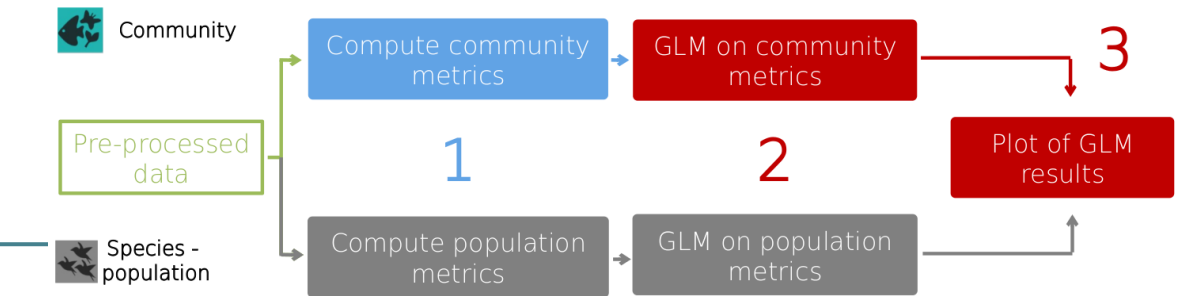
Coline Royaux



Use-cases of the workflow on several data types

- Catch Per Unit of Effort data DATRAS : Galaxy history link
- Reef Life Survey data : Galaxy history link
- Marine Protected Area data from rotative cameras : Galaxy history link
- Anonymized data from the french common bird survey scheme (STOC) : Galaxy history link
- Vigie-Chiro data : Galaxy history link

<https://ecology.usegalaxy.eu/published/workflow?id=37a5a355b699de55>



<https://ecology.usegalaxy.eu/published/workflow?id=2977073db769d29b>

Essential brick to take part to sizing initiatives



From PNDB to french participation to GEO BON

- alpha and gamma diversity
 - Species richness
 - Shannon index
 - Simpson index
 - Hill index
- Pielou index

Community composition



- Community abundance : Galaxy Wor
- Taxonomic / phylogenetic diversity :
- Traits diversity : Galaxy Workflow
- Interactions diversity : Galaxy Workf

Coline Royaux



Several OTU workflows from genetics data

This workflow permits, from diverse Operational Taxonomic Units (OTU) data files (phylogenetic tree, sequence count, taxonomic map, ...), to compute metrics of the EBV type 'Taxonomic / phylogenetic diversity' :

- alpha and gamma diversity
 - Species richness
 - Shannon index
 - Simpson index
 - Fisher index
 - Unique branch length
- beta diversity
 - alpha/gamma matrix
 - Pearson index
 - Spearman index

Tool list (non-exhaustive) :

- Phylo.diversity
- Collect.single
- Collect.shared
- Rarefaction.single
- Rarefaction.shared
- Run QIIME diversity analyses
- beta_diversity_through_plots
- OTU alpha diversity
- OTU alpha rarefaction
- Vegan diversity, fisher, rarefaction

Moreover, it permits to plot a rarefaction curve of alpha diversity and OTUs.

The screenshot shows a Galaxy workflow titled 'Workflow Preview'. It consists of eight steps: 1. data_FrenchBBS.tabular (output: Input), 2. coordCarreSTOCfaux.tabular (output: Input), 3. tabSpecies.csv (output: Input), 4. species_indicateur_fonctionneL.tabular (output: Input), 5. Preprocess population data (input: Input file, output: output (tabular)), 6. Filter species (input: Preprocessed Stoc Input file, output: output (tabular)), 7. Temporal trend indicator (input: Stoc filtered Input, Coordinates of sites, Species file, output: Stoc filtered Input, Coordinates of sites, Species file), 8. Estimate temporal population evolution (input: Stoc filtered Input, Species file, output: plots (input), yearly_variations (tabular), global_tendencies (tabular), lot (png), output_tab_gam_smoos (tabular), thed (tabular)).

About This Workflow
Workflow constructed from history 'STOC CSI' - Version 1

Author
ylebras

Description
This Workflow has no description.

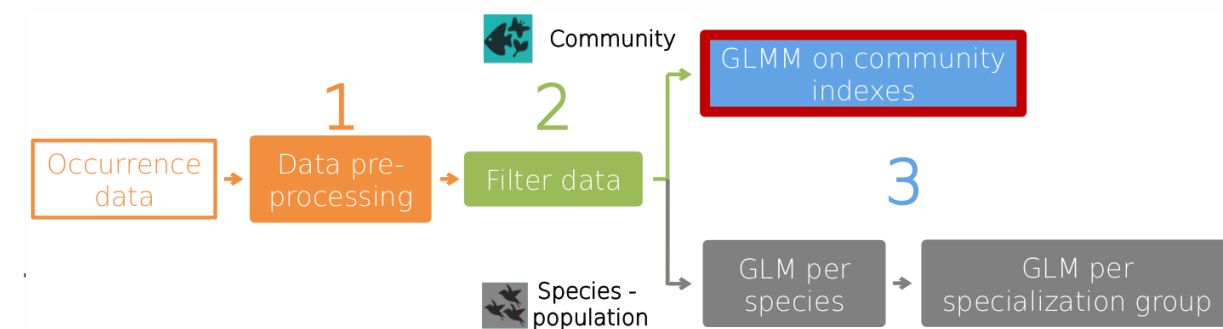
Tags

License
No License specified

Last Updated
Wednesday Jan 6th 17:11:25 2021 GMT+1

Sharing
Use the following link to share preview of this workflow:
<https://ecology.usegalaxy.eu/published/workflow?id=01274dfdbdfcaa88>
id=01274dfdbdfcaa88 Manage sharing settings here.

<https://ecology.usegalaxy.eu/published/workflow?id=01274dfdbdfcaa88>



Essential brick to take part to sizing initiatives



From PNDB to french participation to GEO BON

Workflow Preview

Données d'entrée

- 1: @_RESULTATS_BOIS_BIODIV.csv
- 2: Carte des Grandes régions écologiques (GRECO) de France

Traitement des données

- 3: Analyse des données
- 4: Traitement des sorties d'analyse

Production des visualisations

- 6: Création du diagramme d'évolution du volume de bois
- 5: Création de la carte du taux d'évolution du volume de bois

About This Workflow
 Workflow 'ONB: Création de l'indicateur bois mort' - Version 7

Author
ylebras

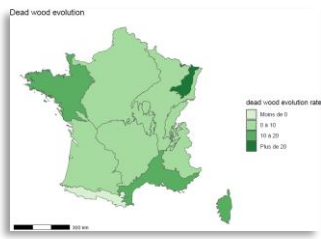
Creators
PNDB

Description
 Workflow créé par le PNDB dans le cadre d'une tâche commune PNDB-SIB afin de présenter une première application de l'approche mise en place dans le cadre du pilote d'opérationnalisation des variables essentielles de biodiversité mené par le PNDB pour GEO BON à travers l'utilisation de la plateforme Galaxy pour partager les codes sources permettant de produire des indicateurs. Lien vers description du pilote d'opérationnalisation des variables essentielles de biodiversité : <https://www.pndb.fr/fr/activites/projets-techniques-et-scientifiques/ebvoc>

Tags
ONB EBV Ecosystem

License
MIT License

<https://ecology.usegalaxy.eu/published/workflow?id=b5f8fa88823d5a16>



Pauline Seguineau



Marie Jossé



Ecosystem functioning

- Primary production
- Ecosystem phenology
- Ecosystem disturbances



Workflow Preview

About This Workflow
 Remote sensing Sentinel 2 data analysis to produce biodiversity metrics - Version 1

Author
ylebras

Description
 This Workflow has no description.

Tags

License
 No License specified

Last Updated
 Thursday Nov 9th 13:51:31 2023 GMT+1

Sharing
 Use the following link to share preview of this workflow: <https://ecology.usegalaxy.eu/published/workflow?id=da68448ab15d167d>. Manage sharing settings here.

<https://ecology.usegalaxy.eu/u/ylebras/w/remote-sensing-sentinel-2-data-analysis-to-produce-biodiversity-metrics>

Essential brick to take part to sizing initiatives



From PNDP to french participation to GEO BON

Ecosystem functioning



- Primary production
- Ecosystem phenology
- Ecosystem disturbances

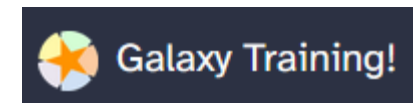
Marie Jossé



Champs blocs indicators

Author(s) Marie Josse Yvan Le Bras

Reviewers



Overview

Questions:

- Which biological questions are addressed by the tutorial?
- Which bioinformatics techniques are important to know for this type of data?

Objectives:

- Calculate IVR (Visual Rollover Indicator)
- Compute dissimilarity indice
- Visualize some diversity indices

Requirements:

- [Introduction to Galaxy Analyses](#)

Time estimation: 1 hour

Supporting Materials:

[Workflows](#)

Published: Jan 4, 2023

Last modification: Jun 3, 2025

License: Tutorial Content is licensed under [Creative Commons Attribution 4.0 International License](#). The GTN Framework is licensed under [MIT](#)

PURL: <https://gxy.io/GTN:T00127>

Revision: 11



<https://gxy.io/GTN:T00127>

<https://ecology.usegalaxy.eu/published/workflow?id=a3ad27817f77061b>

<https://gxy.io/GTN:T00127>

Essential brick to take part to sizing initiatives



From PNDB to french participation to GEO BON

Workflow Preview

Using 97% of 500.0 GB | ytebras

About This Workflow
Ecoregionalization workflow (Part 1) - Version 1

Author
pauline-seguineau

Creators
Seguineau Pauline

Description
Part 1 of ecoregionalization

Tags
Ecology earth-system ecology #Ecology #OccurrencesData

License
Creative Commons Attribution 4.0 International

Last Updated
Wednesday Jul 16th 11:16:04 2025 GMT+2

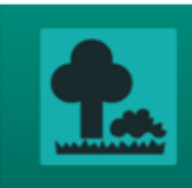
Sharing
Use the following link to share preview of this workflow:
<https://ecology.usegalaxy.eu/published/workflow?id=c5de0b794849e1b1>

Pauline Seguineau



Ecosystem structure

- Live cover fraction
- Ecosystem distribution
- Ecosystem Vertical Profile



Workflow Preview

Using 97% of 500.0 GB | ytebras

About This Workflow
Ecoregionalization workflow (Part 2) - Version 2

Author
pauline-seguineau

Creators
Seguineau Pauline

Description
Second part of the ecoregionalization

Tags
#Ecology Ecology earth-system #OccurrencesData ecology

License
Creative Commons Attribution 4.0 International

Last Updated
Wednesday Jul 16th 10:58:58 2025 GMT+2

Sharing
Use the following link to share preview of this workflow:
<https://ecology.usegalaxy.eu/published/workflow?id=b727b64f37238fff>

Ecoregionalization workflow(s) (in 2 parts)

<https://ecology.usegalaxy.eu/published/workflow?id=b727b64f37238fff>

Essential brick to take part to sizing initiatives

From PND B to french participation to GEO BON

Arthur Barreau



About This Workflow
GEO BON headline indicator KM-GBF A-1 red list ecosystem - Version 3

Author
ylebras

Description
Galaxy Ecology workflow allowing to compute Red List Index Ecosystem indicators.

Tags
EBV KM-GBF Red_List A-1 Ecosystem

License
MIT License

Last Updated
Saturday Sep 20th 11:27:21 2025 GMT+2

Sharing
Use the following link to share preview of this workflow:
<https://ecology.usegalaxy.eu/published/workflow?id=5f7a767d96844c24>

<https://ecology.usegalaxy.eu/u/ylebras/w/geo-bon-headline-indicator-km-gbf-a-1-red-list-ecosystem>

45: output_RLIE ok

total_weight	total_count	upper	lower	RLIE	criteria
total_weight	total_count	upper	lower	RLIE	criteria
955	567	1	0.2	0.6232741617357	criterion_A

JupyterTool output collection
a list with 5 tsv datasets

- output_RLIE
- output_data
- output_one_grouping
- output_two_groupings
- output_weights

Index overview

The Red List Index of ecosystems (RLIE) measures trends in ecosystem collapse risk. It uses the risk categories defined based on IUCN Red List of Ecosystems risk assessments. The index complements the Red List Index of species survival, providing comparable information about ecosystems risk. It is calculated for the overall risk category assigned to each ecosystem and separately for each criterion.

This information sheet provides the code used to calculate the index and an example of each step.

Reference:

Rowland, J. A., Bland, L. M., Keith, D. A., Bignoli, D. J., Burgman, M., Etter, A., Ferrer-Paris, J. R., Miller, R. M. and Nicholson, E. (2020) Ecosystem indices to support global biodiversity conservation. Conservation Letters. e12680

Galaxy Ecology initiative: A lot of young contributors!

Claire Dussin



Elouan Le Mestric



Jean Le Cras



Alan Amossé

Visit, test, follow tutos,
and come back to us!

=> <https://ecology.usegalaxy.eu/>



Olivier Norvez Elie Arnaud

Yassine Ankerl



Molène Mahé



Kévin Payet



Najat Amoukou



Arthur Barreau



Triskell Cumunel



Clara Urfer

Elisa Michon

Timothée Virgoulay

Eloïse Trigodet

Valentin Chambon

Simon Bénateau



Benjamin Yguel



Coline Royaux



Marie Jossé



Pauline Segueineau



2018

2019

2020

2021

2022

2023

2024

2025