

BGE

WP 10.1 – Reference library curation Library Curation Tool

Stephan Kühbandner, Fabian Deister

Project coordinator BIOSCAN,
ZSM/SNSB



Biodiversity Genomics Europe

Meeting details



Funded by
the European Union



UK Research
and Innovation



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra



Workflow Overview

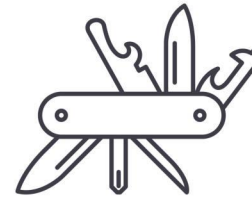


BOLD
Dataset



Automated Pipeline
(Snakemake)

- Record Selection
- Scoring against criteria
- Ranking based on criteria
- OTU clustering & phylogeny
- BAGS grading
- Representative selection
- Curation package (each family)
- Summary report



Expert Curation
(via Library Curation Tool)

- Review Records
- Correct Taxonomy
- Add Notes & Justification
- Select Final Records



Curated Reference Database
Dataset for:

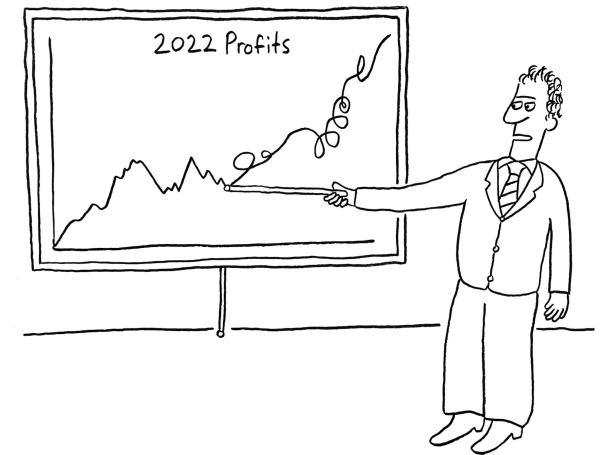
- BOLD
- Publication
- BGE Deliverable

AIM: Manual curation of pre-curated records from pipeline

Why not Excel?

- Big amount of metadata --> confusing
- no dynamic update (BIN-sharing & -splitting, BAGS)
- no standardized feedback from experts
- Reduce amount of data transferred
- 🤖

Do you know
"Advanced Excel"?



*"... and here's where I messed up
the Excel spreadsheet."*

L.M. Pazzell

The Library Curation Tool



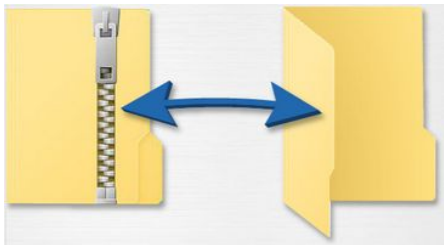
Local web application running in your browser



Download tool and database files:

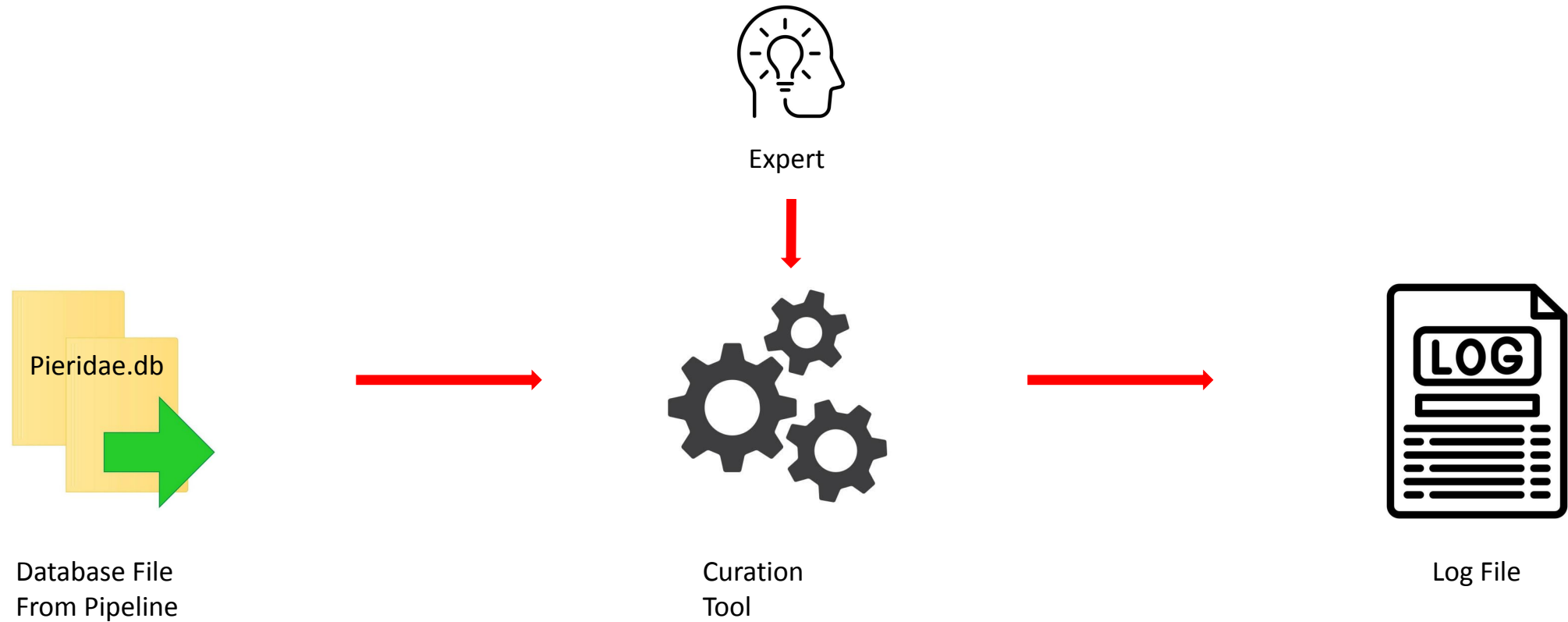
<https://bge-barcoding.github.io/manual-curation/>

(for Linux and Mac see this [GitHUB](#) repository)



Extract files and start to curate

How it works!



Steps of Manual Curation

Inspect records:



- View pre-selected Records
- Filter Data
- View BAGS, BIN Info, other metadata

Action:



- Validate / Invalidate
- Exclude (Reinclude Species)
- Change Species Name
- Choose Reason for Name Correction
- Add Curator Notes



Curation Tool - Interface

The screenshot shows the 'Library Curation Tool' interface. It includes a header with the iBOL EUROPE logo and Biodiversity Genomics Europe branding. The main area contains a 'Filter' section with search terms and types, a 'Control Panel' with buttons like 'Refresh Table', 'Show Invalid Records', 'Load Dataset', 'Stop Curation Tool', and 'Export CSV', and a 'Statistics Panel' showing counts for Total, Filtered, Species, BINs, Curated, and BIN Splitting. Below these is a 'Columns Checkbox' section for selecting columns to display. At the bottom is an 'HTML table' with columns for bin_uri, processid, identification, country_ocean, url, Ranking, BAGS, BIN Info, Status, Reason Name Correction, Correct Species Name, Curator Notes, and Action. The table contains two rows of data for Alucita hexadactyla.

URL

Filter

Control Panel

Statistics Panel

Columns Checkbox

HTML table

bin_uri	processid	identification	country_ocean	url	Ranking	BAGS	BIN Info	Status	Reason Name Correction	Correct Species Name	Curator Notes	Action
BOLD:AAD4822	AELEP007-15	Alucita hexadactyla	Norway	Link	3	A	Single BIN: BOLD:AAD4822	valid record		Alucita hexadactyl		Submit
BOLD:AAD4822	AELEP013-15	Alucita hexadactyla	Norway				Single BIN: BOLD:AAD4822			Alucita hexadactyl		Submit

Pre-Selected Records

by country_representative (records with „yes“ have a grey background)

country_representative: The best record (i.e. the record with the highest sumscore within the best rank) automatically selected for each combination of **species + country + OTU + BIN**. When ties are encountered the script picks one record. These records capture the diversity within each BIN using OTUs clustered at 99% similarity, providing representatives for each species while reducing the duplication within the dataset. Experts can select other records as representatives if they choose.

HTML Table

bin_uri ▲	processid	identification ▲	country_ocean	url	Ranking ?	country_representative	BAGS ?	BIN Info	Status	Reason Name Correction	Correct Species Name	Curator Notes	Action
BOLD:ADY7068	BGSNL511-23	Alucita bidentata	Italy	Link	2	Yes	D	Single BIN: BOLD:ADY7068	<input type="text"/>	<input type="text"/>	Alucita bidentata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AEE3445	BGSNL512-23	Alucita canariensis	Spain	Link	2	Yes	C	BIN-splitting: BOLD:AEE3445, BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita canariensis	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AEE3445	BGSNL513-23	Alucita canariensis	Spain	Link	2	Yes	C	BIN-splitting: BOLD:AEE3445, BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita canariensis	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	GWOTV511-17	Alucita cancellata	Italy	Link	2	Yes	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	GWOTS470-17	Alucita cancellata	Italy	Link	2	Yes	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	LEASU568-18	Alucita cancellata	France	Link	2	Yes	E	BIN-sharing: Alucita cancellata BIN-splitting: BOLD:AEE3445, BOLD:AAO3227	<input type="text" value="invalid record"/>	<input type="text" value="misidentified"/>	Alucita canariensis	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	PHLSA260-11	Alucita cancellata	Spain	Link	2	Yes	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	LEASU965-18	Alucita cancellata	France	Link	2	No	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	<input type="text" value="valid record"/>	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	GBLAA2707-15	Alucita cancellata	Italy	Link	2	No	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	LEASY789-22	Alucita cancellata	France	Link	2	No	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>

Showing 1 to 10 of 196 entries

Previous 1 2 3 4 5 ... 20 Next



HTML Table

bin_uri ▲	processid ▾	identification ▲	country_ocean ▾	url ▾	Ranking ? ▾	country_representative ▾	BAGS ? ▾	BIN Info ▾	Status ▾	Reason Name Correction ▾	Correct Species Name ▾	Curator Notes ▾	Action ▾
BOLD:ADY7068	BGSNL511-23	Alucita bidentata	Italy	Link	2	pre-selected	D	Single BIN: BOLD:ADY7068	<input type="text"/>	<input type="text"/>	Alucita bidentata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AEE3445	BGSNL512-23	Alucita canariensis	Spain	Link	2	pre-selected	C	BIN-splitting: BOLD:AEE3445, BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita canariensis	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AEE3445	BGSNL513-23	Alucita canariensis	Spain	Link	2	pre-selected	C	BIN-splitting: BOLD:AEE3445, BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita canariensis	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	GWOTV511-17	Alucita cancellata	Italy	Link	2	pre-selected	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	GWOTS470-17	Alucita cancellata	Italy	Link	2	pre-selected	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	LEASU568-18	Alucita cancellata	France	Link	2	pre-selected	E	BIN-sharing: Alucita cancellata BIN-splitting: BOLD:AEE3445, BOLD:AAO3227	Invalid	misidentified ▾	Alucita canariensis	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	PHLSA260-11	Alucita cancellata	Spain	Link	2	pre-selected	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	LEASU965-18	Alucita cancellata	France	Link	2	No	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	Valid	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	GBLAA2707-15	Alucita cancellata	Italy	Link	2	No	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>
BOLD:AAO3227	LEASY789-22	Alucita cancellata	France	Link	2	No	E	BIN-sharing: Alucita canariensis Single BIN: BOLD:AAO3227	<input type="text"/>	<input type="text"/>	Alucita cancellata	<input type="text"/>	<input type="button" value="Submit"/>

Showing 1 to 10 of 196 entries

Previous 1 2 3 4 5 ... 20 Next

Support

- FAQs (Homepage <https://bge-barcoding.github.io/manual-curation/>)
- Video (Homepage <https://bge-barcoding.github.io/manual-curation/>)
- User Manual (Curation Tool – Main Folder)
- Within the tool, click on the red questionmarks 
- Contact us 



Thank you for your Attention!



Biodiversity Genomics Europe

BIODIVERSITY GENOMICS EUROPE



Funded by
the European Union



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra



UK Research
and Innovation

