

Integrating and Maintaining Berry Data

The Germinate Database and beyond

02. APRIL 2025 | ELISA SENGER

IBG-4, FORSCHUNGSZENTRUM JÜLICH, GERMANY



Advancing Berry Breeding:
From Phenotyping to Genomic Innovation
Workshop in Ancona, Italy



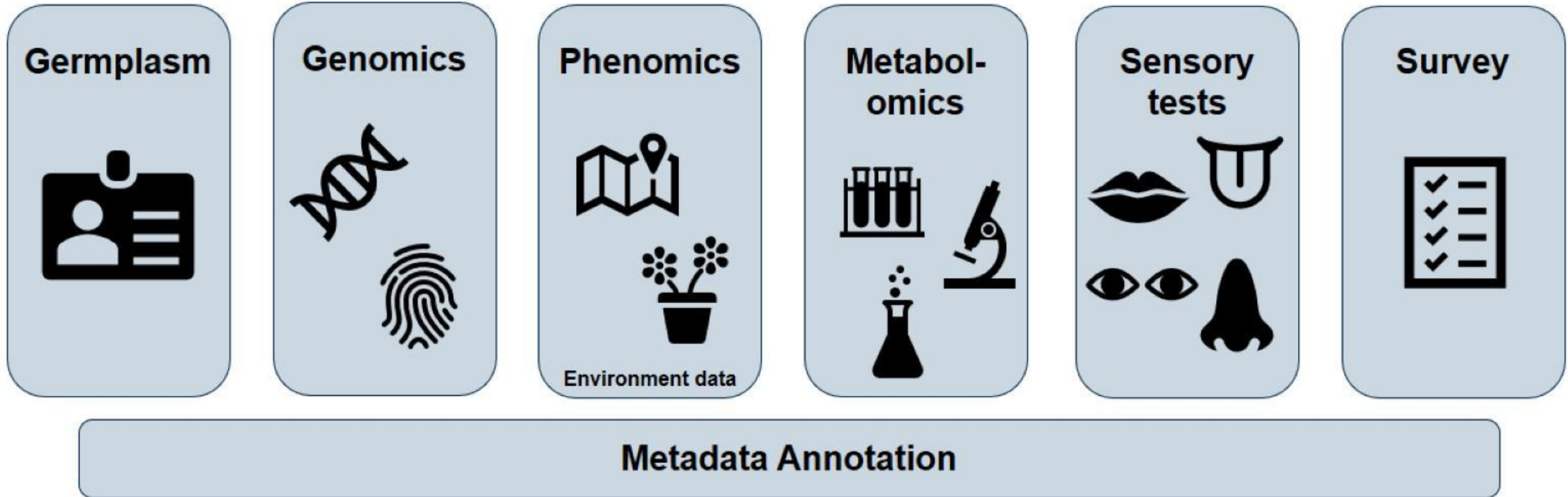
Outlook

- ▶ **Data Integration**
- ▶ **Data Repository Requirements**
- ▶ **The Germinate database**
- ▶ **BreedingValue berry data @ Germinate**



Berry Research Data

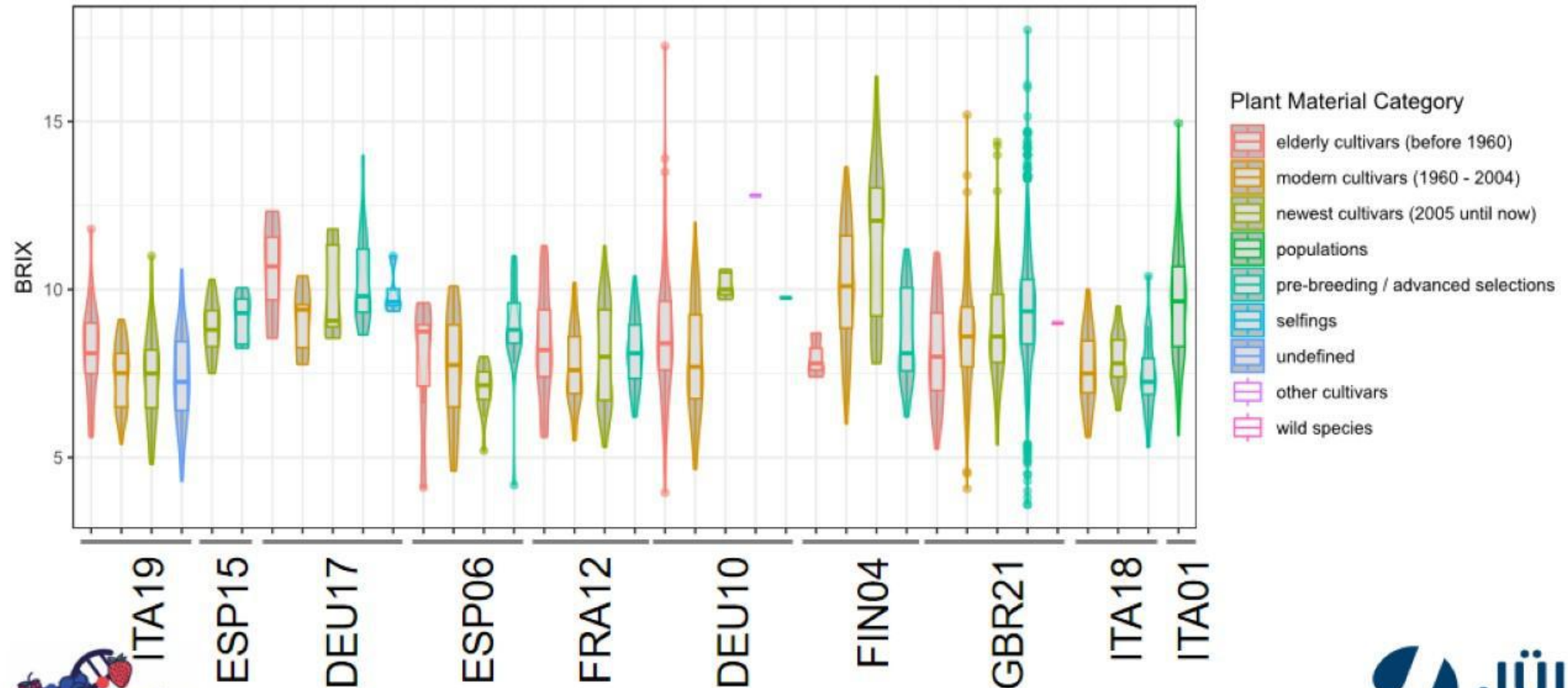
Data types



Data Integration

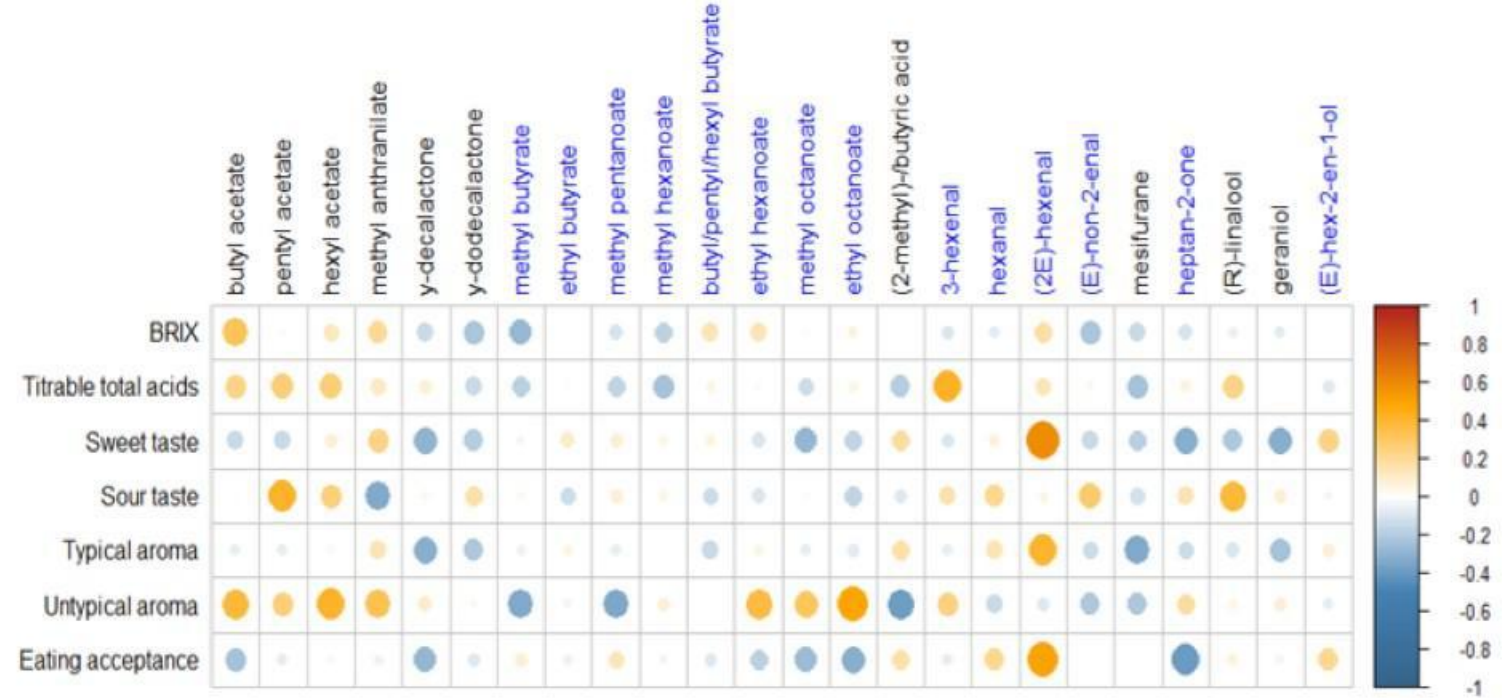
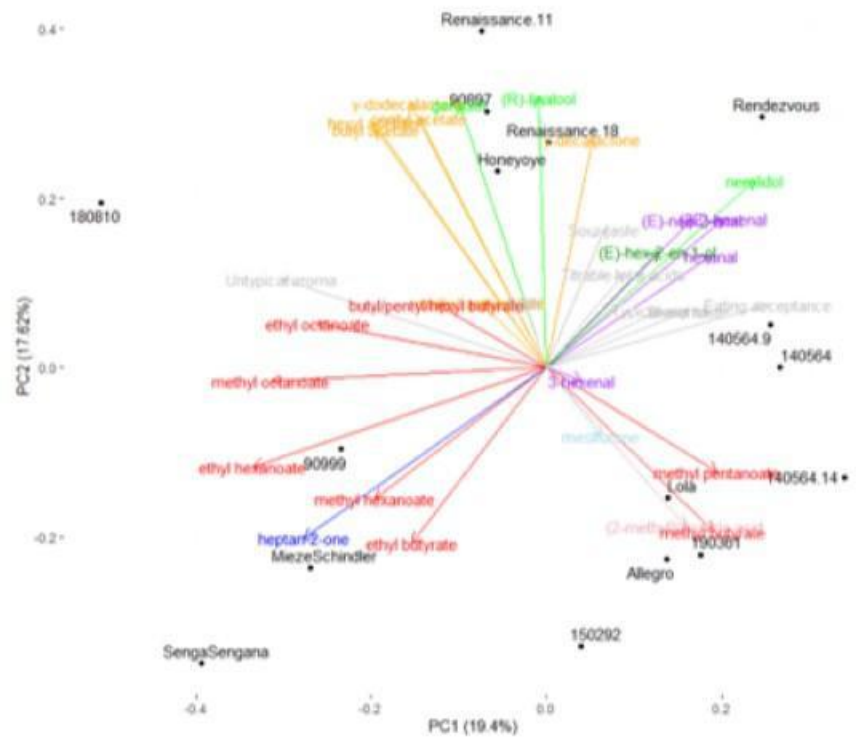
Example: Phenomics & Germplasm passport data

- How are accessions characterized? example of strawberry BRIX values per plant material category at different sites



Example: Metabolomics & Sensory tests

- Which metabolites constitute taste, flavor, aroma perception? example of strawberry volatiles, correlation & PCA



Data Integration

Example: Genomics & Metabolomics

https://www.plabipd.de/helixer_main.html



- Which are the genes? Structural annotation with 'Helixer'

HELIXER

about

gene annotation

Job submission

- upload a file with nucleotide sequences (one or many records in a valid FASTA format)
 - minimum sequence length of a single record: 25 kbp
 - maximum file size (including all records): 1 GByte
 - optionally, compress the file ('.gz', '.zip' and '.bz2' are supported)
- select the organism lineage 'land plants', 'vertebrates', 'invertebrates' or 'fungi' (required for lineage-specific annotation)
- optionally, provide an email address (where a link to the GFF3-formatted gene annotation results will be sent)
- optionally, specify a label (used as prefix in the GFF3-formatted gene annotation results to specify the source of the nucleotide sequence)

Helixer version (v0.3.4)

Please note:

- Helixer is being actively maintained and improved.
- please do not submit the same genome to be processed multiple times. The results will be unchanged and will use resources unnecessarily.
- we currently support compressed (.gz .zip .bz2) and uncompressed input files.

Data Integration

Example: Genomics & Metabolomics

- Which are the genes? Structural annotation with 'Helixer'

https://www.plabipd.de/helixer_main.html

A screenshot of the Helixer web interface. The interface has a light blue header with the "HELIXER" logo. On the left, there is a sidebar with links for "about" and "gene annotation". The main content area contains several input fields and buttons. At the top, there is a section for "Upload nucleotide sequence file (FASTA format)" with a "Durchsuchen..." button and the text "Keine Datei ausgewählt.". Below this is a checkbox labeled "Use demo file - Arabidopsis_lyrata.v.1.0.dna.chromosome.8.fa.gz". Further down is a "Select Lineage-specific mode" section with a dropdown menu currently showing "Land plant". Below that is an "Enter label for GFF feature naming" section with a text input field containing "gff_label_name" and an information icon. At the bottom of this section is an "Email address" field with the placeholder "email address" and an information icon. A blue "Submit Job" button is located at the bottom center. The footer includes the European Union flag, the "Value" logo, and the "JÜLICH Forschungszentrum" logo.

Data Integration

Example: Genomics & Metabolomics

- What are the gene functions? Functional annotation with 'Mercator'

https://www.plabipd.de/mercator_main.html



MERCATOR4

[about](#)
[version history](#)
[FAQs](#)
[de.NBI feedback survey](#)

▼ Job submission for protein annotation

▶ Result of protein annotation

▶ Visualize result in tree viewer

▶ Visualize result in heatmap viewer

Mercator4 v7.0

[Job submission](#) [Result tree viewer](#) [Result Heatmap viewer](#)

Sequence type ☒ Protein ☐ DNA ⓘ

[Job submission](#) [Result tree viewer](#) [Result Heatmap viewer](#)

Sequence type ☒ Protein ☐ DNA ⓘ

☐ Include Prot-scriber annotations (Beta version) ⓘ

☐ Include Swissprot annotations ⓘ

Upload FASTA file Keine Datei ausgewählt.

☐ Use demo FASTA file

Job name ⓘ

Email address ⓘ

Data Integration

Example: Genomics & Metabolomics

- What are the gene functions? Functional annotation with 'Mercator'

https://www.plabipd.de/mercator_main.html



Hierarchical framework
consisting of 31 top-level
categories (also called *BINs*)

- 1 Photosynthesis
- 2 Cellular respiration
- 3 Carbohydrate metabolism
- 4 Amino acid metabolism
- 5 Lipid metabolism
- 6 Nucleotide metabolism
- 7 Coenzyme metabolism
- 8 Polyamine metabolism
- 9 Secondary metabolism
- 10 Redox homeostasis
- 11 Phytohormone action
- 12 Chromatin organisation
- 13 Cell division
- 14 DNA damage response
- 15 RNA biosynthesis
- 16 RNA processing
- 17 Protein biosynthesis
- 18 Protein modification
- 19 Protein homeostasis
- 20 Cytoskeleton organisation
- 21 Cell wall organisation
- 22 Vesicle trafficking
- 23 Protein translocation
- 24 Solute transport
- 25 Nutrient uptake
- 26 External stimuli response
- 27 Multi-process regulation
- 28 Plant reproduction
- 30 Clade-specific metabolism
- 50 Enzyme classification



11 Phytohormone action

- 11.1 abscisic acid
 - 11.1.1 biosynthesis
 - 11.1.1.1 zeaxanthin epoxidase *(ABA1)
 - 11.1.1.2 neoxanthin synthase *(ABA4)
 - 11.1.1.3 neoxanthin biosynthesis cofactor *(NCD1)
 - 11.1.1.4 9-cis-epoxycarotenoid dioxygenase *(NCED)
 - 11.1.1.5 xanthoxin oxidase *(ABA2)
 - 11.1.1.6 xanthoxin oxidase molybdopterin sulfurase *(ABA3)
 - 11.1.1.7 abscisic aldehyde oxidase *(AAO)
 - 11.1.2 perception and signalling
 - 11.1.3 conjugation and degradation
 - 11.1.4 transport
- 11.2 auxin
- 11.3 brassinosteroid
- 11.4 cytokinin
- 11.5 ethylene
- 11.6 gibberellin
- 11.7 jasmonic acid
- 11.8 salicylic acid
- 11.9 strigolactone
- 11.10 karikins

Data Integration

Example: Genomics & Metabolomics

- What are the gene functions? Functional annotation with 'Mercator'

https://www.plabipd.de/mercator_main.html



▼ 9 | Secondary metabolism

▼ 9.1 | terpenoids

└ mevalonate (MVA) pathway

- 3 2 2 2 acetyl-CoA C-acyltransferase *(ACAT1/2)
- 3 1 1 1 3-hydroxy-3-methylglutaryl-CoA synthase *(HMGS)
- 3 2 3 1 3-hydroxy-3-methylglutaryl-CoA reductase *(HMGR)
- 0 1 1 1 mevalonate kinase *(MVK)
- 2 1 1 1 phosphomevalonate kinase *(PMK)
- 1 2 1 1 mevalonate diphosphate decarboxylase *(MVD1/2)
- 1 2 2 1 isopentenyl diphosphate isomerase *(IDI1/2)

▼ methylerythritol phosphate (MEP) pathway

└ 1-deoxy-d-xylulose 5-phosphate import

- 1 1 1 1 D-xylulose kinase
- 1 1 1 0 D-xylulose 5-phosphate transporter
- 7 3 4 2 1-deoxy-D-xylulose 5-phosphate synthase *(DXS)
- 2 1 1 2 1-deoxy-D-xylulose 5-phosphate reductase *(DXR)

page 10

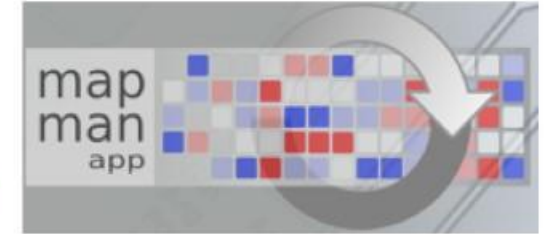


Data Integration

https://www.plabipd.de/mapman_main.html

Example: Genomics & Metabolomics

- How do the gene functions embed in pathways? Visualization with MapMan



Protein function annotation



Mapman4 category (BIN)	protein name	protein description
BINCODE	NAME	IDENTIFIER
"1.3.1"	"Phosphoglycolate phosphatase"	"ab0g12345"
"2.3.1"	"Citrate synthase"	"ab0g23456"
"18.9.1"	"Protein modification, hydroxylation, prolyl hydroxylase"	"ab0g34567"

Run Mercator on genome assembly

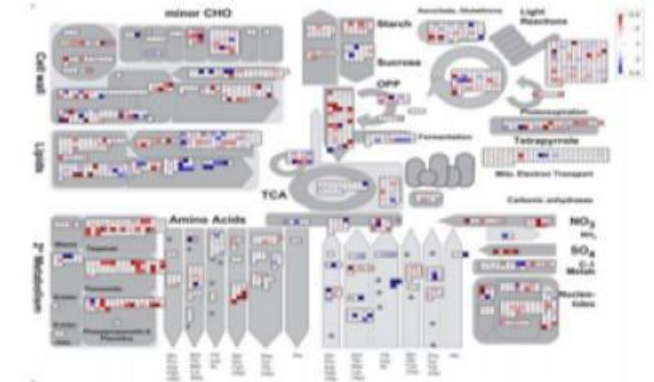
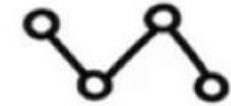
Gene expression data



	condition 1		condition 2	
	expression change	significance	expression change	significance
	log2 fold		log2 fold	
example	"cond1 vs. normal"	"cond1 vs. normal"	"cond2 vs. normal"	"cond2 vs. normal"
"ab0g12345"	-0.25	0	0.85	1
"ab0g23456"	0.63	1	1.25	0
"ab0g34567"	0.46	0	1.62	1

Differential gene expression, metabolites, proteins

Pathway diagram

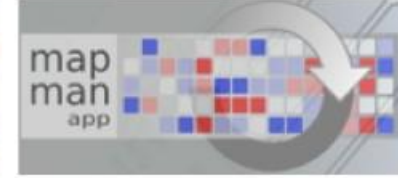


Pathway of interest, available on the plabipd website

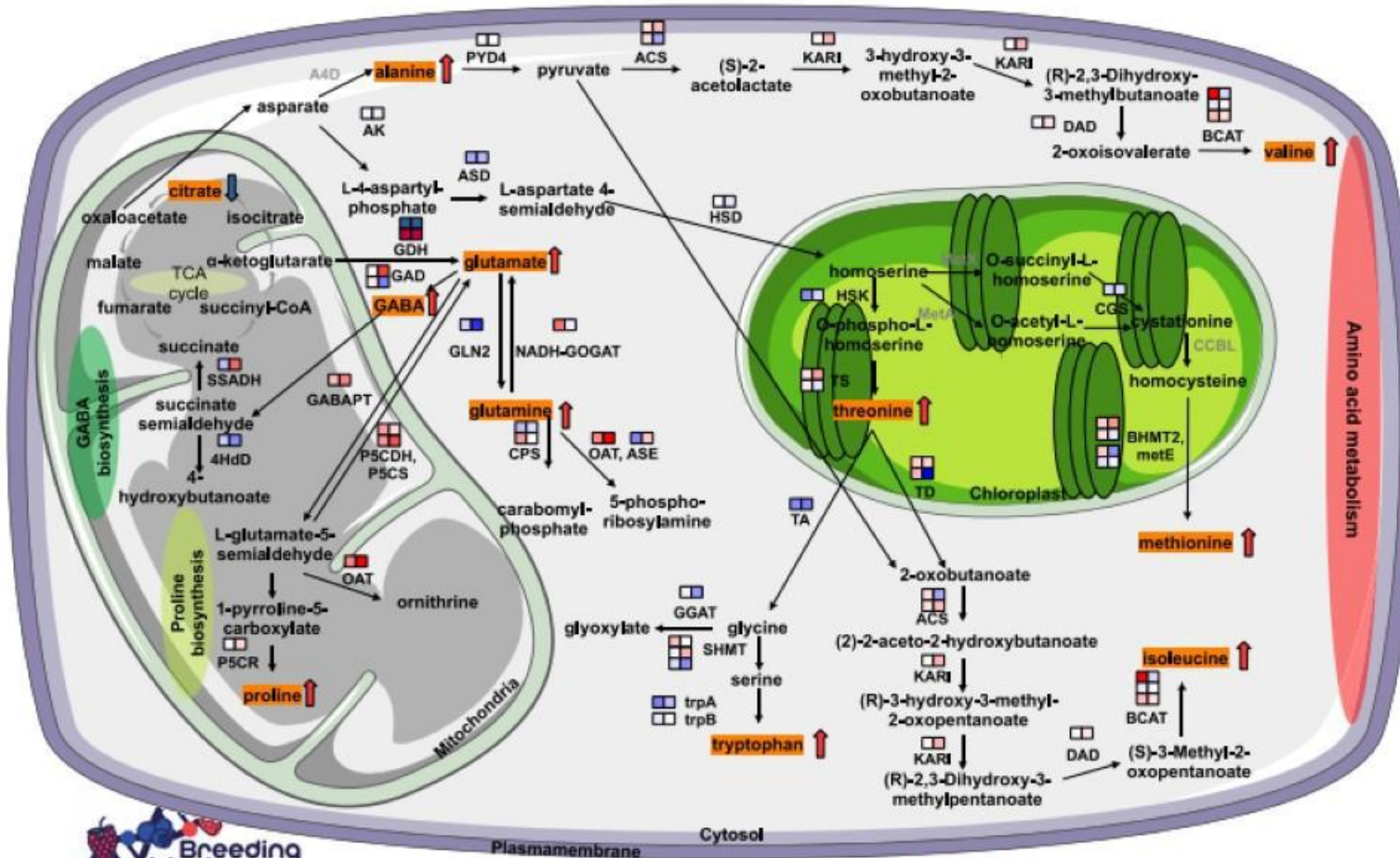


Data Integration

Example: Genomics & Metabolomics



- Which genes are active? Differential expression analysis and embedding into pathways



Ziegler et al. (2025)
A full genome assembly reveals drought stress effects on gene expression and metabolite profiles in blackcurrant (*Ribes nigrum* L.)

<https://doi.org/10.1093/hr/uhae313>



Data Repository

General Requirements

- Databases shall host and curate data mid- to long-term
- Databases shall provide tools to explore the data online
- FAIR¹ and open data
 - “desirable” or mandatory under public funding
 - increases value and utility of the data (repetitive and broader applications)
 - knowledge becomes more accessible for exploitation
- Large specialized databases existing for single or few data types
- Few integrative plant databases existing but limited for integrated² multi-omics data

¹ Findable + Accessible + Interoperable + Re-useable

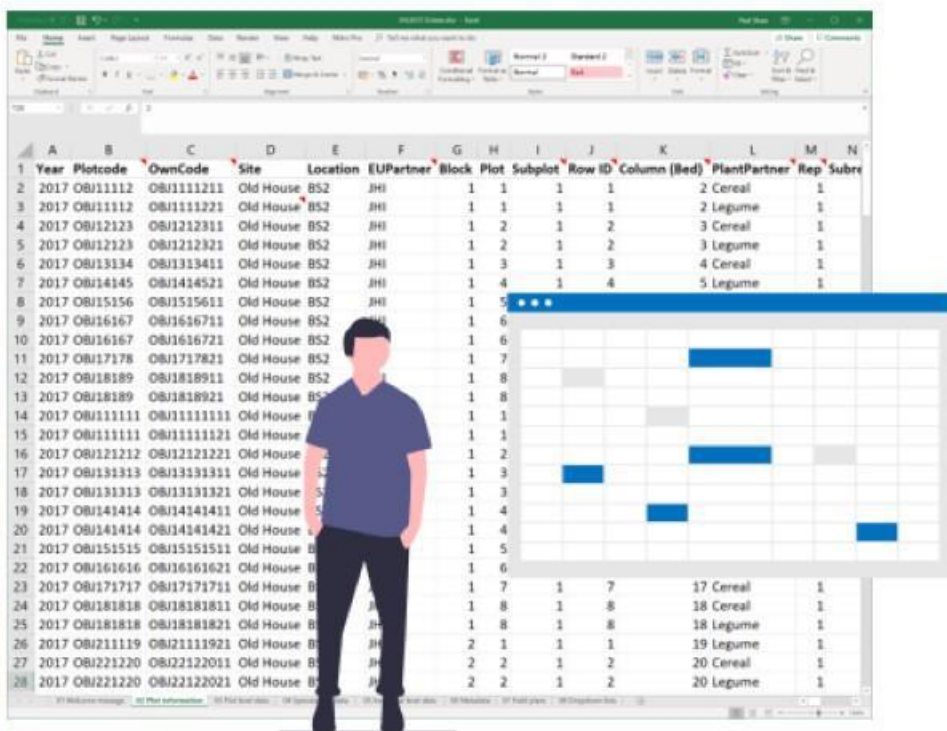
² data types connected to each other within the database (e.g. genotypic, phenotypic, metabolomics, climate, passport)



Germinate (developed by JHI)

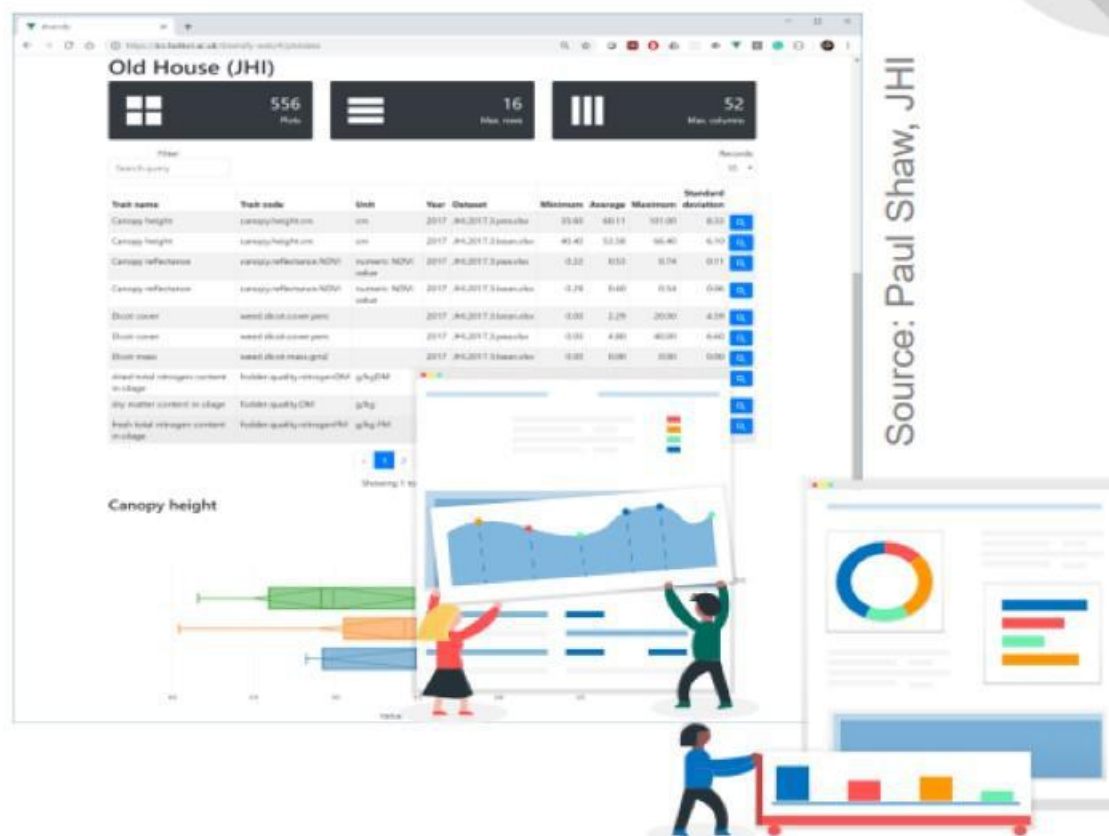
In a nutshell...

From this...



Source: Paul Shaw, JHI

...to this



Germinate

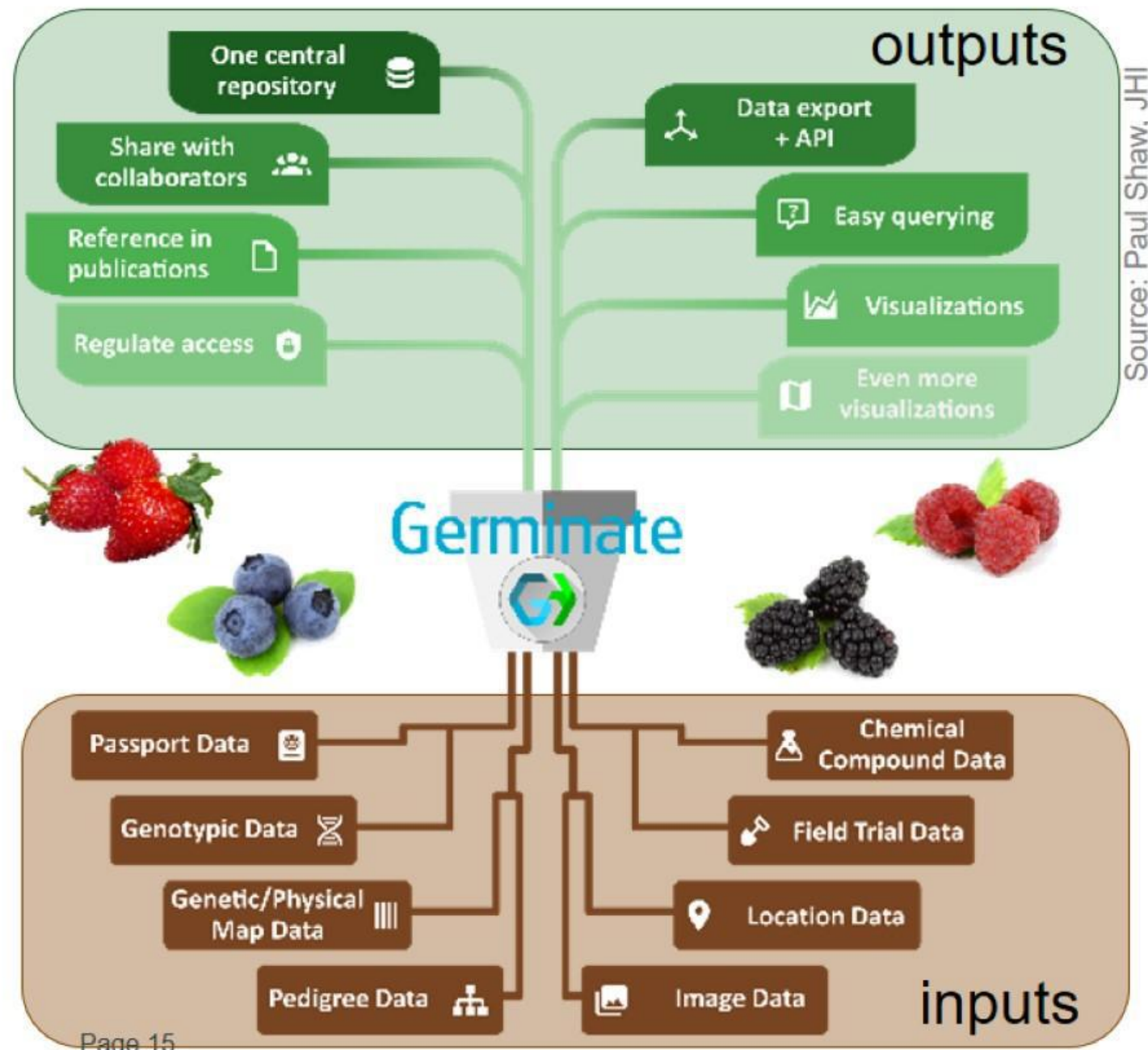
Data of plant genetic resources

- Connecting different data types
- Public and private data
- Visualizing + exploring
- Compatible with MCPD + Dublin Core
- BrAPI implementation currently being extended
- Implementation of ontologies ongoing

Publications:

Shaw et al. (2017) Germinate 3
doi:10.2135/cropsci2016.09.0814

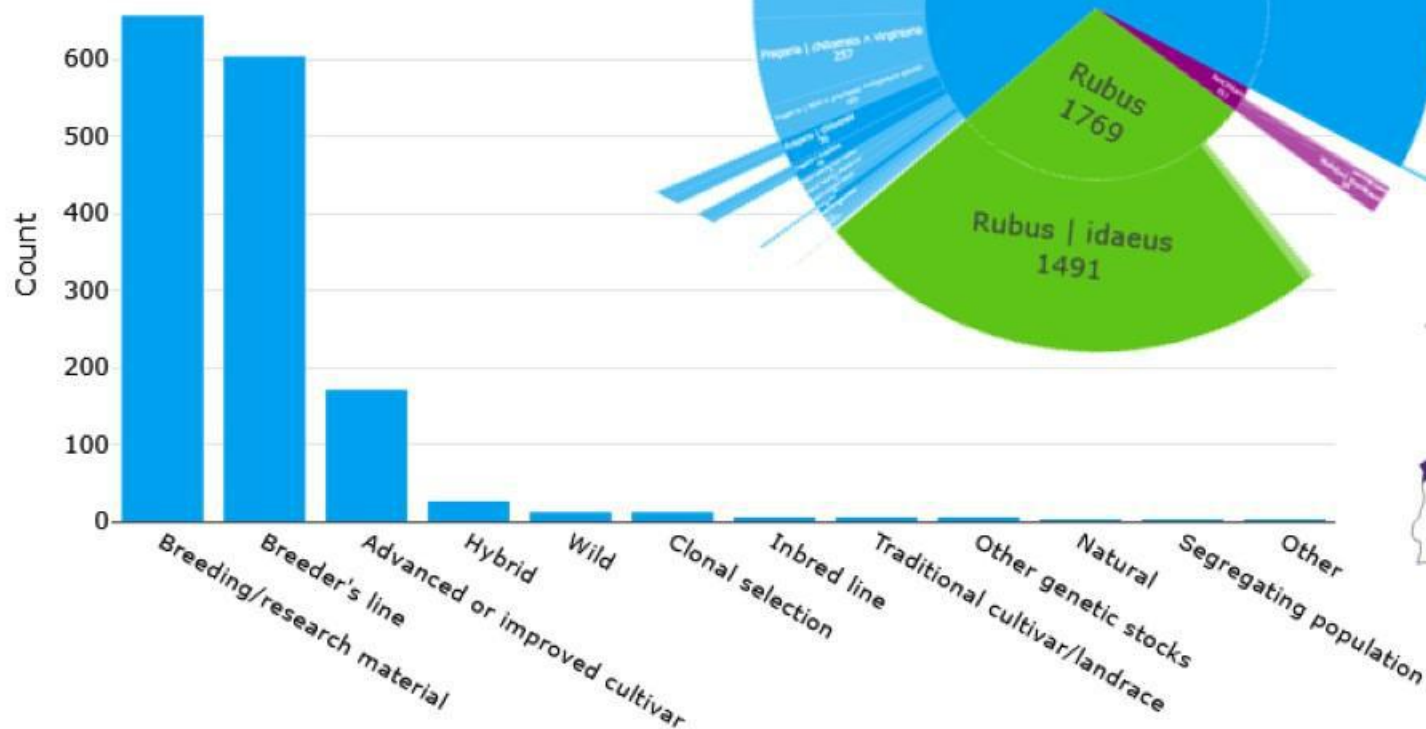
Raubach et al. (2020) From bits to bites
doi:10.1002/csc2.20248



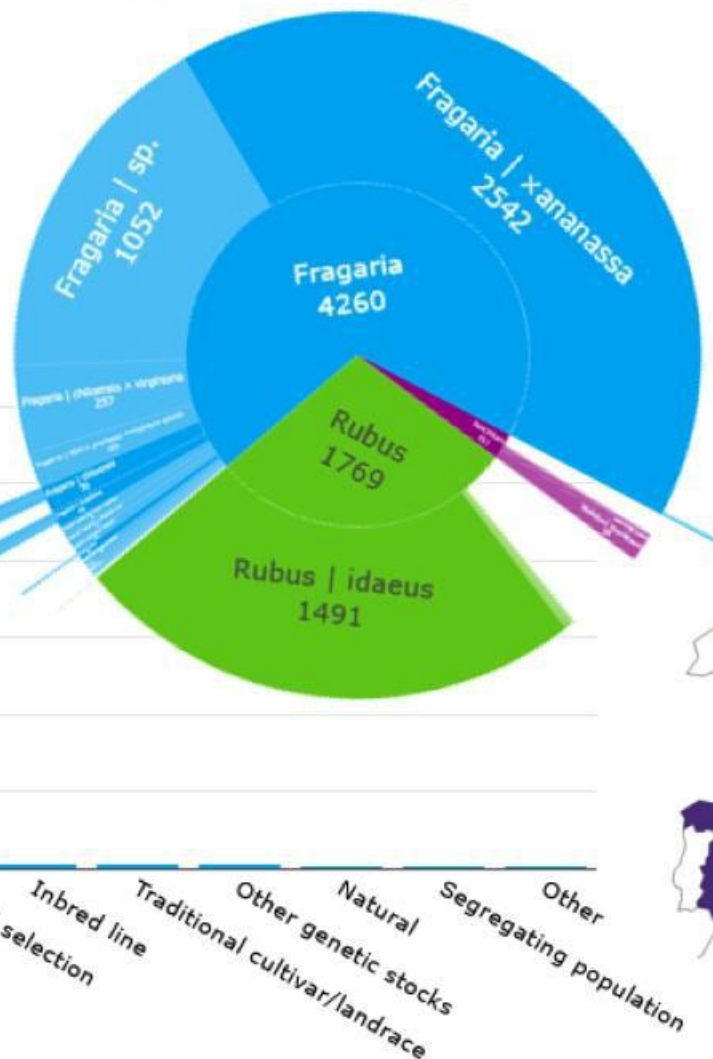
BreedingValue Berry Data @ Germinate

Germplasm passport data

- ~ 6K berry accessions



Germplasm grouped by biological status



BreedingValue Berry Data @ Germinate

Genotypic data

Maps

The following maps were found in this version of Germinate. To view markers contained on a map choose a map from the table below and click on its name.




[Use the table filtering to search for specific items.](#)
Showing 1 to 2 of 2 records
 100

Id	Map name	Map description	Marker count
1	2023 JHI Blueberry SNP Marker Data	2023 JHI Blueberry SNP Marker Data	4.011
2	2023 JHI Raspberry SNP Marker Data	2023 JHI Raspberry SNP Marker Data	4.989

Markers

The table below shows all markers within Germinate. Click on the marker to see more information about it.




[Use the table filtering to search for specific items.](#)
Showing 1 to 10 of 54.035 records
 10

Id	Marker name	Type	Synonyms
1	AX-123356912	50k array	
2	AX-123356914	50k array	

BreedingValue Berry Data @ Germinate

Phenotypic data (incl. metabolomics and sensory data)

Trials

Use the table filtering to search for specific items.

Showing 1 to 7 of 7 records

	Id	Dataset name	Dataset description	Experiment Id	Experiment name	Dataset type	Data type	License	Data objects	Data points	Linked data resources	Ext.
☐	2	2020 JHI Blueberry LAB field mapping population Tunnel J&A	A blueberry cross used for genetic mapping	2	Generic Experiment ⓘ	Trials		CC BY-SA	301	3.58K	Show	
☐	3	2021 JHI Blueberry LAB field mapping population Tunnel J&A	A blueberry cross used for genetic mapping	2	Generic Experiment ⓘ	Trials		CC BY-SA	299	3.45K	Show	
☐	4	2022 JHI Blueberry LAB field mapping population Tunnel J&A	A blueberry cross used for genetic mapping	2	Generic Experiment ⓘ	Trials		CC BY-SA	298	3.85K	Show	
☐	17	2022 JHI Blueberry Bullionfield cultivars	Replicated commercial blueberry cultivars	2	Generic Experiment ⓘ	Trials		CC BY-SA	29	992	Show	
☐	18	2022 JHI Blueberry replicated mapping population	Mapping progeny blueberry cultivars	2	Generic Experiment ⓘ	Trials		CC BY-SA	272	3.41K	Show	
☐	19	2023 JHI Blueberry Bullionfield cultivars	Replicated commercial blueberry cultivars	2	Generic Experiment ⓘ	Trials		CC BY-SA	29	937	Show	
☐	21	2023 JHI Blueberry LAB field mapping population Tunnel J&A	A blueberry cross used for genetic mapping	2	Generic Experiment ⓘ	Trials		CC BY-SA	293	3.78K	Show	

Use the checkboxes to select one or more rows.



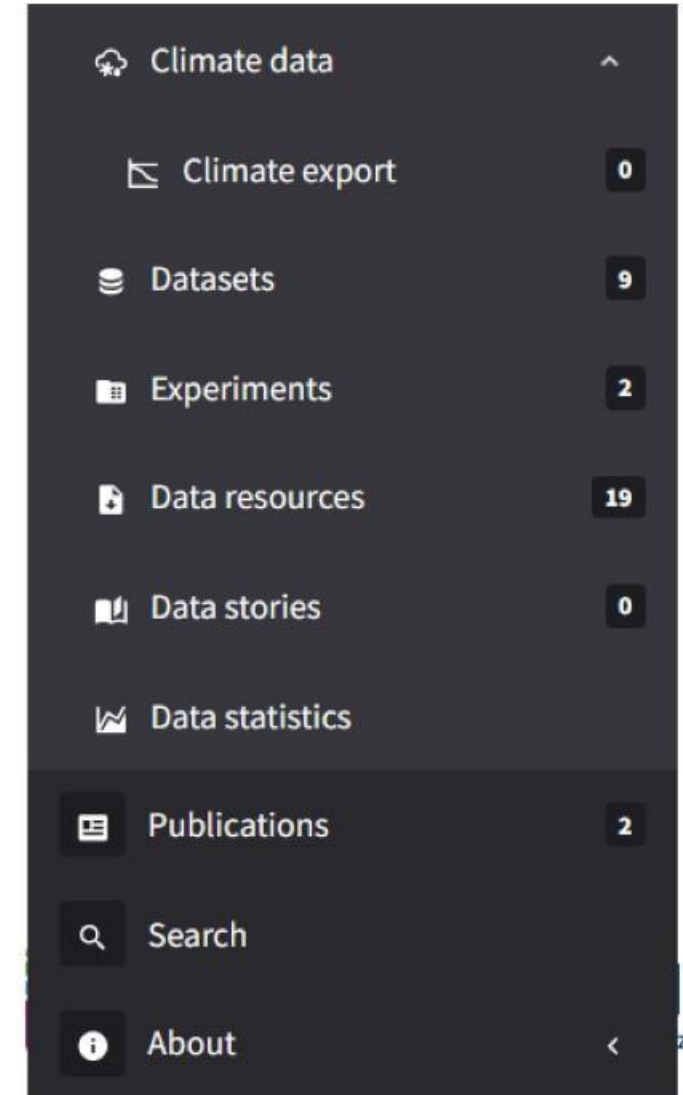
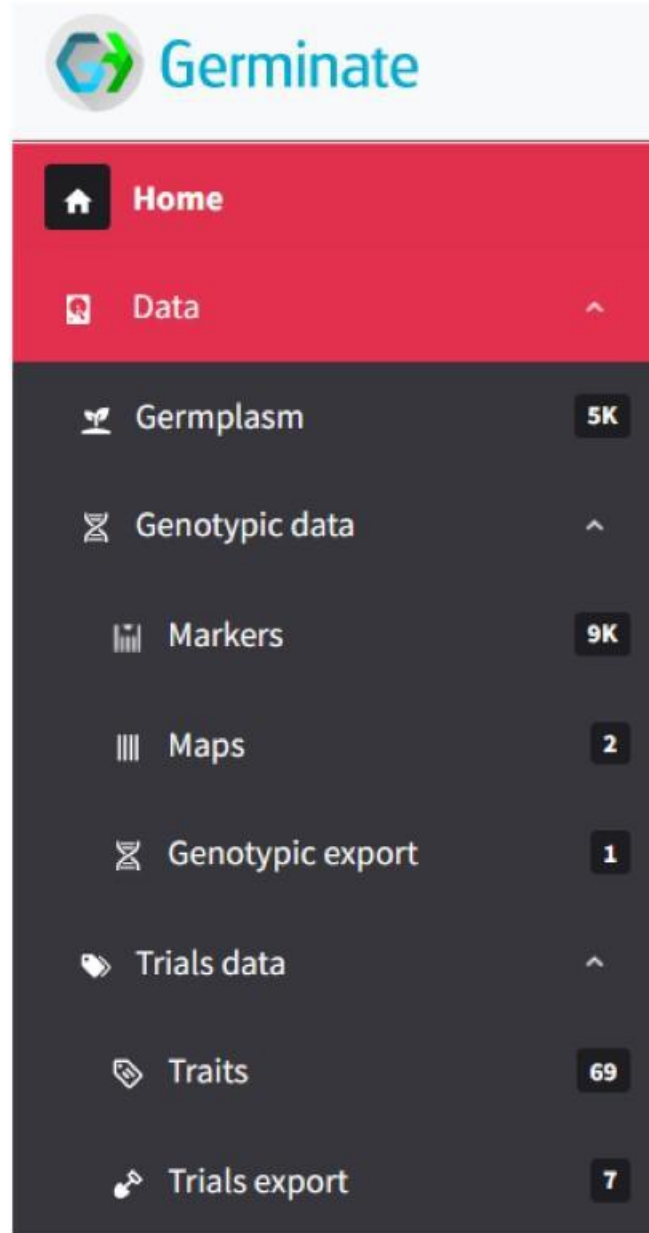
Next



BreedingValue Berry Data @ Germinate

Further data

- Data, data, data...
- Photographs of plants and fruits
- Metabolic profiles
- Location
- Climate



Take-Home

- Data integration becomes increasingly important.
- Data integration thoughts to be included in project planning and data management
- Sound planning, structuring and preparations needed
- Suitable data repository indispensable for good data management and easy access, also for users outside the respective project or study



Thanks to

- **YOU ! ...FOR YOUR ATTENTION**
- **JAMES-HUTTON INSTITUTE**, INVERGOWRIE, SCOTLAND, UK ...FOR PROVIDING THE PLATFORM AND SOME SLIDES (PAUL SHAW, SEBASTIAN RAUBACH, DEPARTMENT OF INFORMATION AND COMPUTATIONAL SCIENCES, SUSAN MCCALLUM, DEPARTMENT OF CELL AND MOLECULAR SCIENCES)
- **BREEDINGVALUE PARTNERS** ...FOR COLLABORATION AND PROVIDING DATA
- **EU** ...FOR FUNDING
- **IBG-4 AT FZJ** ...FOR GENOMICS TOOLS (RAINER SCHWACKE, MARIE BOLGER) AND TRAINING



Member of the Helmholtz Association

