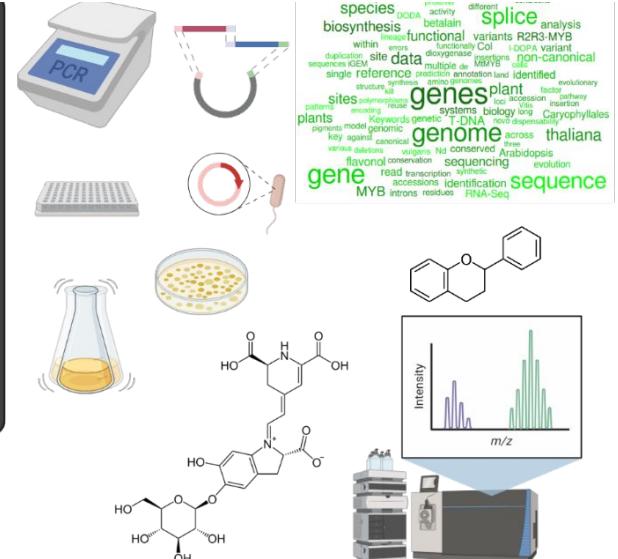
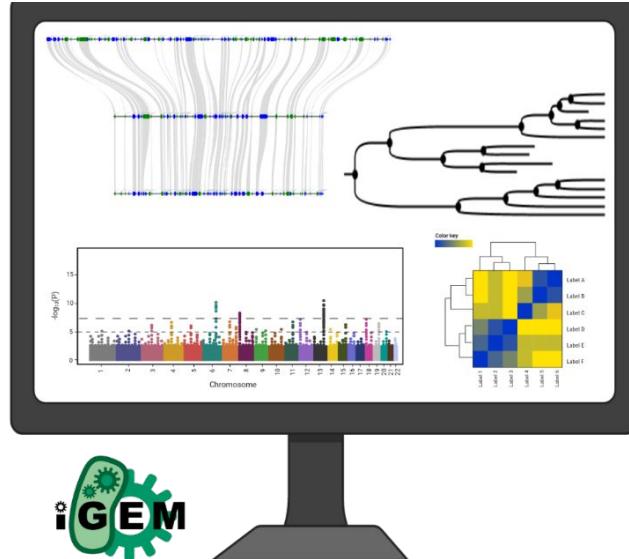
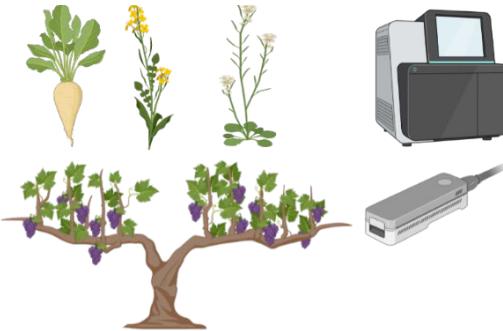




Technische
Universität
Braunschweig



GE32/MM12 - Databases and Services

Prof. Dr. Boas Pucker (Plant Biotechnology and Bioinformatics)

Availability of slides

- All materials are freely available (CC BY) - after the lectures:
 - StudIP: **GE32/MM12**
 - GitHub: <https://github.com/bpucker/teaching>
- Questions: Feel free to ask at any time
- Feedback, comments, or questions: [b.pucker\[a\]tu-braunschweig.de](mailto:b.pucker[a]tu-braunschweig.de)

My figures and content can be re-used in accordance with CC BY 4.0, but this might not apply to all images/logos. Some figure were constructed using bioRender.com.

How to find information in the literature?



How to find information in the literature?

- GoogleScholar

Google Scholar



- SciFinder



- PubMed



- WebOfScience/WebOfKnowledge

- PubPharm



Which sequence databases do you know?



Sequence databases

- Sequence Read Archive (SRA)
- European Nucleotide Archive (ENA)
- DNA Data Bank of Japan (DDBJ)
- GenBank (NCBI)
- Joint Genome Institute (JGI) / Phytozome
- PLAZA
- Ensembl genomes

The screenshot shows the SRA homepage with a blue header featuring the NIH logo and 'National Library of Medicine' text. Below the header, there's a search bar and a 'Log In' button. The main content area has a large image of a brain and the text 'SRA - Now available on the cloud'. It includes sections for 'Getting Started', 'Tools and Software', and 'Related Resources'. A sidebar on the left lists 'How to Submit', 'Data Search and Download', and 'How to use SRA in the cloud'. A footer at the bottom contains links for 'Submit', 'Search', 'Help', and 'Support'.

The screenshot shows the ENA homepage with a teal header featuring the ENA logo and 'European Nucleotide Archive'. Below the header, there's a search bar and a 'Log In' button. The main content area has a large image of a green plant and the text 'ENA - European Nucleotide Archive'. It includes sections for 'How to Submit', 'Search', 'Help', and 'Support'. A sidebar on the left lists 'Submit', 'Search', 'Help', and 'Support'. A footer at the bottom contains links for 'Submit', 'Search', 'Help', and 'Support'.

The screenshot shows the DDBJ Center homepage with a yellow logo and the text 'DDBJ' and 'Bioinformatics and DDBJ Center provides sharing and analysis services for data from life science researches and advances science.' Below the logo, there are six cards arranged in a 2x3 grid: 'Services' (Search, analysis, database services of DDBJ Center), 'Submission' (Navigation for how to submit your data), 'Super Computer' (NIH Supercomputer), 'Statistics' (Statistics of DDBJ Center services), 'Activities' (Training sessions and achievements of DDBJ Center), and 'About us' (About Bioinformatics and DDBJ Center).

The screenshot shows the JGI Phytozome 13 homepage with a green header featuring the JGI logo and 'Phytozome 13'. Below the header, there's a search bar and a 'Log In' button. The main content area has a large image of a green plant and the text 'Welcome to Phytozome'. It includes sections for 'Overview', 'Release Notes', and 'News'. A sidebar on the left lists 'Submit', 'Search', 'Help', and 'Support'. A footer at the bottom contains links for 'JGI Home' and 'JGI Data Portal'.

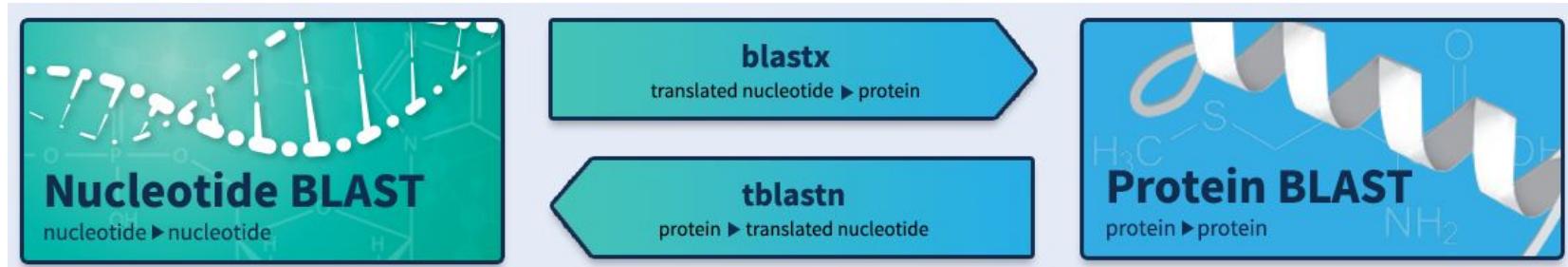
The screenshot shows the PLAZA platform homepage with a green header featuring the PLAZA logo. Below the header, there's a search bar and a 'Log In' button. The main content area has a large image of a green plant and the text 'What is PLAZA?'. It includes sections for 'Available PLAZA instances' (News, Dicots PLAZA 5.0, Monocots PLAZA 5.0, Dicotyledons PLAZA 1.0), 'Summary' (Dicots PLAZA 5.0, Monocots PLAZA 5.0, Dicotyledons PLAZA 1.0), and 'Species included' (Dicots PLAZA 5.0, Monocots PLAZA 5.0, Dicotyledons PLAZA 1.0). A footer at the bottom contains links for 'Home', 'Tools', and 'Quick Selection'.

How to search for sequences?



Basic Local Alignment Search Tool (BLAST)

- BLASTn = nucleotide query against a nucleotide database
- BLASTp = protein query against a protein database
- BLASTx = (translated) nucleotide query against a protein database
- tBLASTn = protein query against (translated) nucleotide database



DIAMOND

- BLAST equivalent with respect to result quality and format
- About 50x faster sequence search
- RAM requirements are substantially higher than for BLAST



Introduction

DIAMOND is a sequence aligner for protein and translated DNA searches, designed for high performance analysis of big sequence data. The key features are:

- Pairwise alignment of proteins and translated DNA at 100x-10,000x speed of BLAST.
- Frame shift alignments for long read analysis.
- Low resource requirements and suitable for running on standard desktops or laptops.
- Various output formats, including BLAST pairwise, tabular and XML, as well as taxonomic classification.

Build passing build passing downloads 284k Anaconda.org 2.0.15 downloads 248k total Citations 5693

Documentation

The online documentation is located at the [GitHub Wiki](#).

Support

Diamond is actively supported and developed software. Please use the [issue tracker](#) for malfunctions and the [GitHub discussions](#) for questions, comments, feature requests, etc.

About

DIAMOND is currently developed by Benjamin Buchfink at the Drost lab, Max Planck Institute for Biology, Tübingen, Germany (since 2019). Its development was supported for one year by the German Federal Ministry for Economic Affairs and Energy through an EXIST grant in 2018-2019. It was developed independently by Benjamin Buchfink from 2016-2018. Its initial version was developed in 2013-2015 by Benjamin Buchfink at the Huson lab, University of Tübingen, Germany.

<https://github.com/bbuchfink/diamond>
Buchfink et al., 2021: 10.1038/s41592-021-01101-x

Do you know websites with bioinformatic tools?



EBI services

- EBI = European Bioinformatics Institute
- Broad range of different services:
 - DNA&RNA
 - Gene expression
 - Literature
 - Structures

Featured data resources



AlphaFold DB

Database for protein structure predictions for numerous species

CC-BY



BioModels

A repository of peer-reviewed, published, computational models.

[Web API](#) | [CCO](#)



ChEMBL

An open data resource of binding, functional and ADMET bioactivity data.

[Web API](#) | [CC-BY](#)

Featured tools



Clustal Omega

Multiple sequence alignment of DNA or protein sequences. Clustal Omega replaces the older ClustalW alignment tools.

[Web API](#)



HMMER

Fast sensitive protein homology searches using profile hidden Markov models (HMMs) for querying against both sequence and HMM target databases.

[Web API](#)



Annotation Platform

Consolidating text-mined and curated annotations

[Web API](#)

<https://www.ebi.ac.uk/services/data-resources-and-tools>



EBI services (gene expression)

- Expression Atlas = Visualization of gene expression
- ArrayExpress = collection of various gene expression analysis data sets
- PRIDE = database of protein abundances (MS-based)

Search

Refine by

Type
 Data resources
 Tools

Category
 Chemical biology
 Cross domain
 DNA & RNA
 Gene expression
 Literature
 Ontologies
 Proteins
 Structures
 Systems

Sort by: Featured

Expression Atlas

An added-value database that shows which genes/proteins are expressed under which conditions, and how expression differs between conditions.

DATA RESOURCE CC-BY

PRIDE

An archive of protein expression data determined by mass spectrometry.

DATA RESOURCE Web API | CCO

ArrayExpress

A database of functional genomics experiments, including microarray and RNAseq expression data typically related to publications.

DATA RESOURCE Web API

EGA

A service for permanent archiving and sharing of all types of potentially identifiable genetic, molecular and phenotypic data resulting from biomedical research projects.

DATA RESOURCE

Ensembl VEP

Analyse your variants and predict the functional consequences of known and unknown variants

TOOL Web API

EVA

A repository and browser for all types of genetic variation data

DATA RESOURCE EMBL-EBI Terms of use

RNA-Sequer

RNASeq-er API - a gateway to systematically updated analysis of public RNA-Seq data

DATA RESOURCE Web API

<https://www.ebi.ac.uk/services/data-resources-and-tools>

EBI services (DNA&RNA)

- Ensembl = genome databases
- Clustal Omega = sequence alignment tool
- ENA = European Nucleotide Archive (sequence read submission)
- Gramene = comparative plant biology
- MAFFT = sequence alignment tool
- Rfam = database of RNA structures
- T-coffee = sequence alignment tool
- TreeFam = database of phylogenetic trees

EBI services - Gramene

- Genome browsers allow inspection of plant annotations
- Pangenomic data sets = enables comparative genomics
- Bulk download options = enables users to conduct more customized analysis
- BLAST = sequence comparison
- Expression Atlas (dedicated to plants)

Welcome to Gramene release 65

Gramene is a curated, open-source, integrated data resource for comparative functional genomics in crops and model plant species.

 Genome Browser Genome annotations, variation and comparative tools	 Plant Reactome Browse and analyze metabolic and regulatory pathways
 Plant Pan Genomes Gramene-powered sites focused on specific crops	 BLAST Query our genomes with a DNA or protein sequence
 CLIMtools Environment x Genome x Phenotype Associations in A. thaliana	 Plant Expression Atlas Browse plant expression results at EBI
 bioMart An advanced genomic query interface powered by BioMart	 Ensembl Tools Tools for fetching and searching genomic data
 Bulk Downloads FTP download of our data	 Archive Legacy tools and data (markers, Cyc pathways, etc)

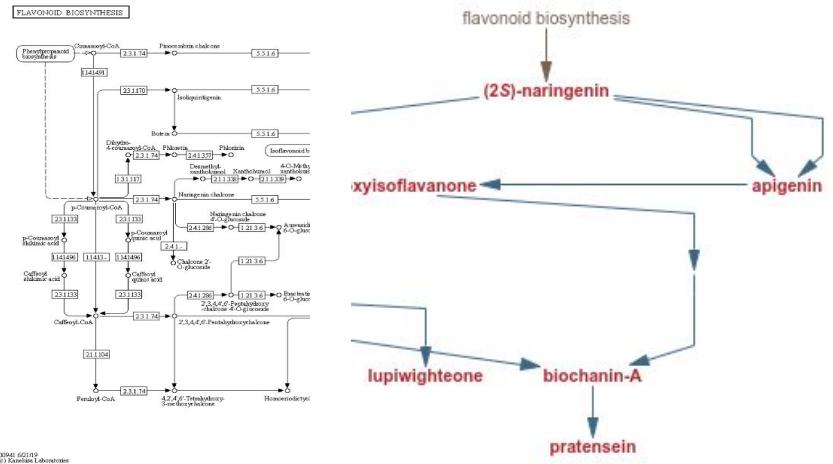
EMBOSS

- EMBOSS = European Molecular Biology Open Software Suite
- Collection of (small) bioinformatic tools on a website
- Link: <https://www.bioinformatics.nl/emboss-explorer/>
- Homepage: <http://emboss.sourceforge.net/>

Where would you look for pathways?

Pathway databases

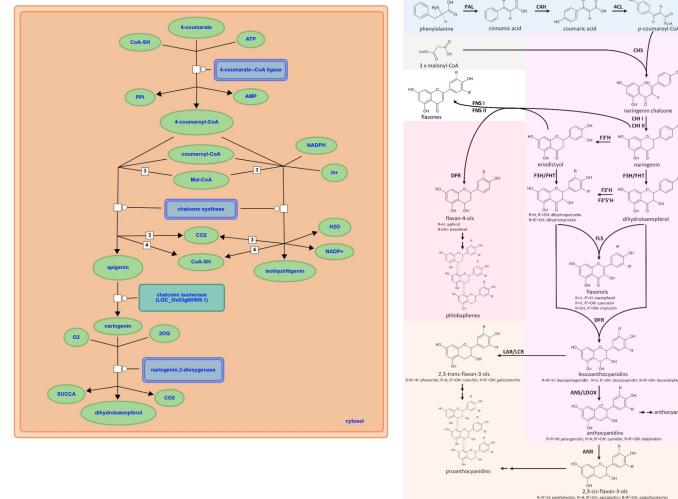
- KEGG: <https://www.genome.jp>



- MetaCyc: <https://metacyc.org>

- Plant Reactome: <https://plantreactome.gramene.org/>

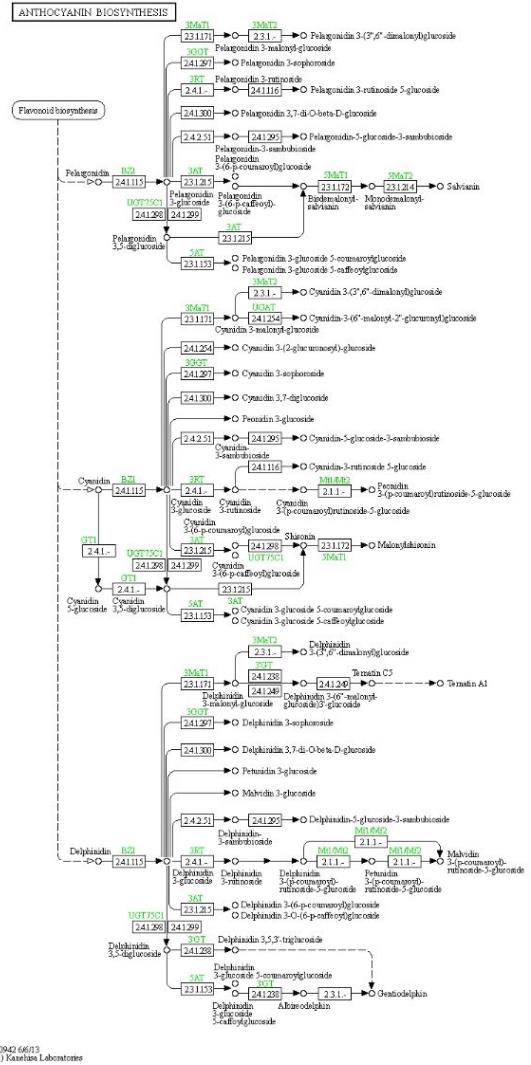
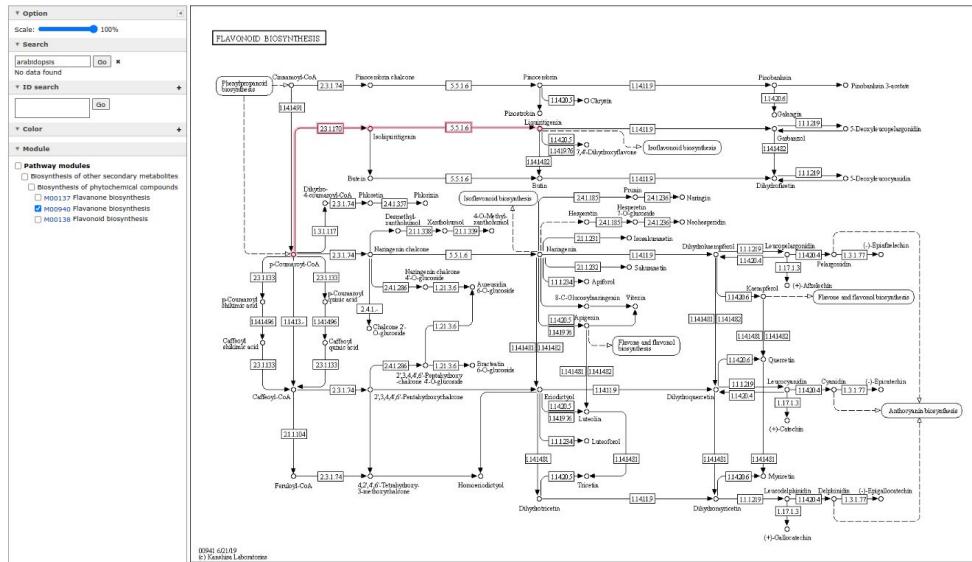
- Publications



Kanehisa et al., 2015: 10.1093/nar/gkv1070
Pucker et al., 2020: 10.3390/plants9091103

KEGG - Flavonoid biosynthesis

- Selection of one particular species
- Selection of individual branches in pathways (flavone biosynthesis)
- Gene names are given in addition to EC numbers



00940:6613
(c) Kansaike Laboratories

https://www.genome.jp/kegg-bin/show_pathway?map00940

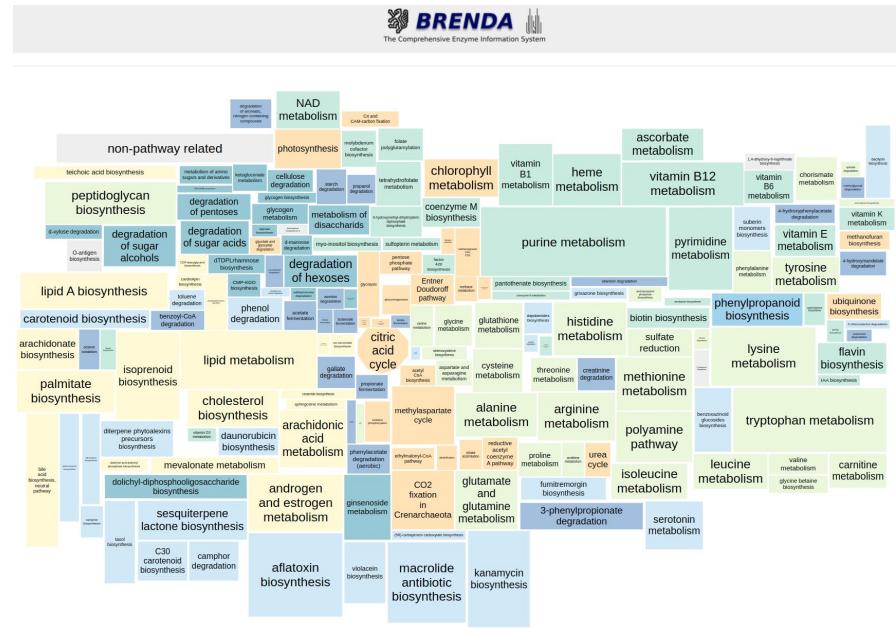


Details about enzymes?



BRENDA: BRaunschweig ENzyme DAtabase

- Enzyme database hosted at TU Braunschweig (BRICS)
- Text and structure-based queries
- Visualization of pathways
- Manual curation of datasets
- Many details about enzyme properties (substrates, kinetics, mutants, ...)



www.brenda-enzymes.org
Chang et al., 2021: 10.1093/nar/gkaa1025

BRENDA - example (1)

- DFR (dihydroflavonol 4-reductase) is a central enzyme in the anthocyanin biosynthesis
- BRENDA provides information about reaction mechanism and species-specific parameters
- Enzymatic properties of multiple species allow quick comparison

Information on EC 1.1.1.219 - dihydroflavonol 4-reductase and Organism(s) *Arabidopsis thaliana*

for references in articles please use BRENDA:EC1.1.1.219

EC Tree

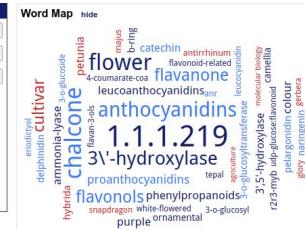
L-1 Oxidoreductases
L-1.1 Acting on the CH-OH group of donors
L-1.1.1 With NAD⁺ or NADP⁺ as acceptor
L-1.1.1.219 dihydroflavonol 4-reductase

IUBMB Comments

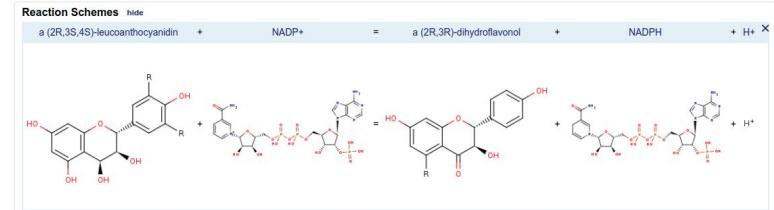
This plant enzyme, involved in the biosynthesis of anthocyanidins, is known to act on (+)-dihydrokaempferol, (+)-taxifolin, and (+)-dihydromyricetin, although some enzymes may act only on a subset of these compounds. Each dihydroflavonol is reduced to the corresponding cis-flavan-3-diol. NAD⁺ can act instead of NADP⁺, but more slowly.

Specify your search results
Mark a special word or phrase in this record:
Search Reference ID:
Search UniProt Accession:
Select one or more organisms in this record:
All organisms
Allium cepa
Anthurium andraeanum
Arabidopsis thaliana
Brassica rapa

This record set is specific for:
Arabidopsis thaliana



The taxonomic range for the selected organisms is: *Arabidopsis thaliana*
The expected taxonomic range for this enzyme is: Bacteria, Eukaryota, Archaea



Synonyms
[Boag058630](#), [BrDFR1](#), [BrDFR10](#), [BrDFR11](#), [BrDFR12](#), [BrDFR2](#), [BrDFR3](#), [BrDFR4](#), [BrDFR5](#), [BrDFR6](#), [BrDFR7](#)

BRENDA - example (2)

- Specific substrates and products are listed (reaction details)
- K_M and V_{max} values are included
- References point to publications and other databases; synonyms
- EC number
- Biochemical details: pH optimum, pI value, molecular weight

Substrate ▾ | Product ▾ | Reaction Diagram | Organism ▾ | Uniprot | Commentary (Substrate) ▾ | Literature (Substrate) ▾ | Commentary (Product) ▾ | Literature (Product) ▾ | Reversibility (reversible/in-reversible/?) and specified ▾

dihydrokempferol + NADPH + H ⁺	leucopelargonidin + NADP+	?	Arabidopsis thaliana	P51102	-	702000	-	-	?
dihydromyricetin + NADPH + H ⁺	leucodelphinidin + NADP+	?	Arabidopsis thaliana	P51102	-	702000	-	-	?
dihydroquercetin + NADPH + H ⁺	leucocyanidin + NADP+	?	Arabidopsis thaliana	P51102	-	702000	-	-	?
additional information	?	-	Arabidopsis thaliana	P51102	does not catalyze naringenin	702000	-	-	?

Cofactor ▾ | Organism ▾ | Uniprot ▾ | Commentary ▾ | Literature ▾ | Image
NADPH | Arabidopsis thaliana | P51102 | - | 702000 | ?

Specific Activity [μmol/min/mg] | Organism ▾ | Uniprot ▾ | Commentary ▾ | Literature ▾

0.00002	Arabidopsis thaliana	P51102	with dihydrokempferol as substrate	702000
0.0006	Arabidopsis thaliana	P51102	with eriodictyol as substrate	702000
0.001	Arabidopsis thaliana	P51102	with dihydromyricetin as substrate	702000
0.0013	Arabidopsis thaliana	P51102	with dihydroquercitin as substrate	702000

Organism ▾ | Commentary ▾ | Literature ▾ | Uniprot ▾ | Sequence DB ▾ | Source ▾
Arabidopsis thaliana | - | 702000 | P51102 | UniProt | BRENDA

General Information ▾ | Organism ▾ | Uniprot | Commentary ▾ | Literature ▾

physiological function	Arabidopsis thaliana	P51102	DFR plays a key role in determining intensity and pigment coloration because its specificity and activities dictate the type and amount of the colorless leucoanthocyanidins	702000
------------------------	----------------------	--------	--	--------

UniProt ▾ | Entry Name ▾ | Organism ▾ | No. of AA ▾ | No. of Transm. Helices ▾ | Molecular Weight [Da] ▾ | Source ▾ | Sequence ▾ | Localization Prediction ▾

P51102	DFRA_ARATH	Arabidopsis thaliana	382	0	42775	Swiss-Prot	Show Sequence	other Location (Reliability: 5)
--------	------------	----------------------	-----	---	-------	------------	---------------	---------------------------------

Cloned (Commentary) ▾ | Organism ▾ | Uniprot ▾ | Literature ▾

into pTrcHis2-TOPO and heterologously expressed in Escherichia coli TOP10F' strain. DFR cDNA cloned into pRSF-FHT and inserted into Escherichia coli BL21Star to create E-coli strain	Arabidopsis thaliana	P51102	702000
---	----------------------	--------	--------

Ref. ▾ | Authors ▾ | Title ▾ | Journal ▾ | Vol. ▾ | Pages ▾ | Year ▾ | Organism (Uniprot) ▾ | Pubmed ID ▾ | Source ▾

702000	Leonard, E.; Yan, Y.; Chemler, J.; Matern, U.; Martens, S.; Koffas, M.	Characterization of dihydroflavonol 4-reductases for recombinant plant pigment biosynthesis applications	BioCat., Biotransform.	26	243-251	2008	Fragaria x ananassa (Q22617), Ipomoea nil (Q24607), Arabidopsis thaliana (P51102), Rosa hybrid cultivar (Q41158), Lilium sp. (Q6UAQ7), Anthurium andraeanum (Q84L22)	-	BRENDA
--------	--	--	------------------------	----	---------	------	--	---	--------

Select items on the left to see more content.

External Links (specific for EC-Number 1.1.1.219)

- ExplorEnz (official portal for IUBMB Enzyme Nomenclature)
- Exasy Enzyme Nomenclature Database
- KEGG
- MetaCyc
- SABIO-RK
- NCBI: PubMed, Protein, Nucleotide, Structure, Gene, OMIM
- Enzyme Nomenclature (alternative site)
- UniProt
- PDB
- PROSITE Database of protein families and domains
- InterPro (database of protein families, domains and functional sites)

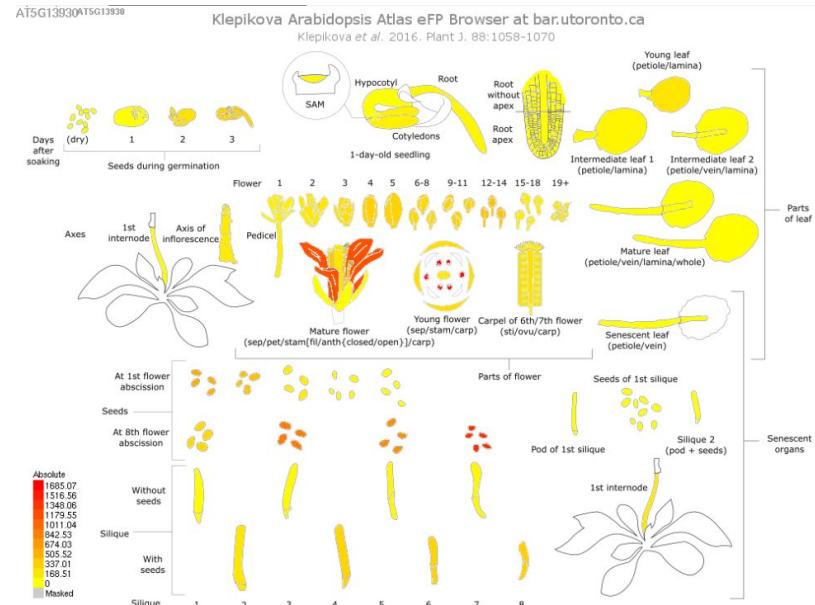
What if you need more details about a particular gene?



What if you need more details about a particular gene?

- Publications:
 - PubMed
 - GoogleScholar
- Organism specific databases:
 - TAIR (*Arabidopsis thaliana*)
 - BananaGenomeHub (*Musa acuminata*)
 - TomatoGenome (*Solanum lycopersicum*)
- Genome browsers

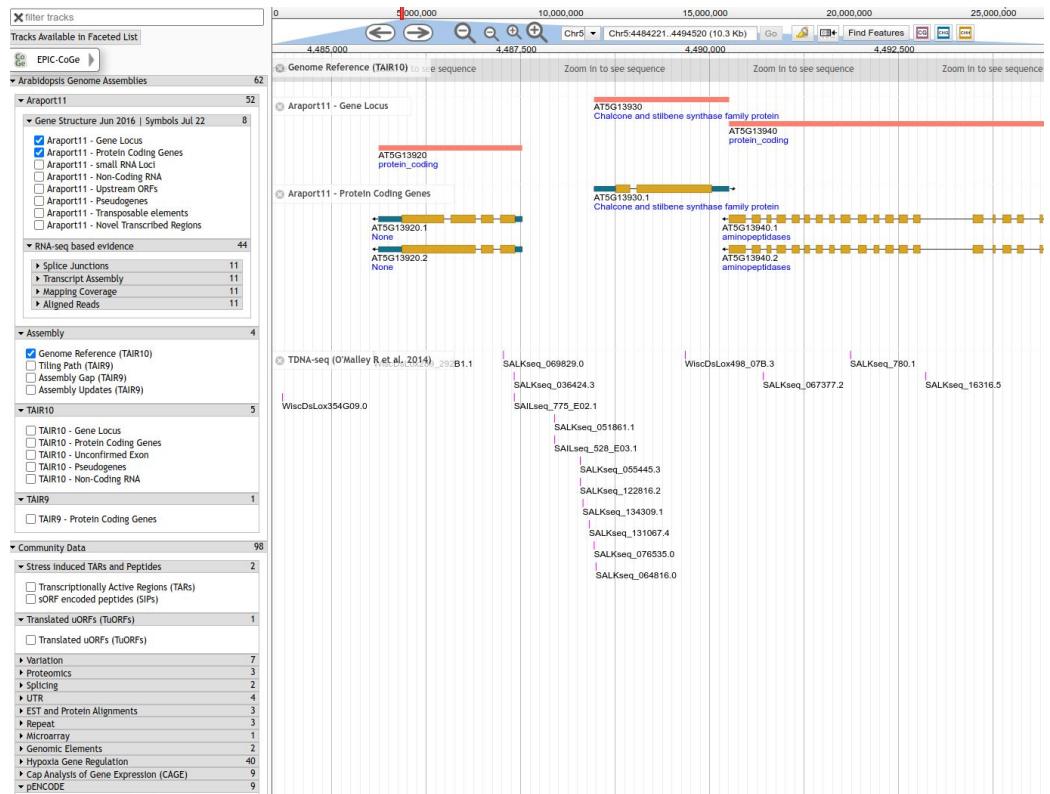
- Information about the gene function (annotation text, GO terms, ...)
- Information about gene structure including splice variants
- Information about gene expression (BAR eFP browser)
- Knock-out lines of a gene of interest
- External links to various databases
- References to corresponding publications



<https://www.arabidopsis.org/servlets/TairObject?type=locus&name=At5g13930>

Genome browser (jbrowse)

- Genome browsers enable interactive inspection of gene structures
- Feature tracks can be selected/deselected
- Functional annotation and related information can be displayed



Where would you look for gene expression data?



Where would you look for gene expression data?

- Electronic Fluorescent Pictograph (eFP)
- Gene Expression Omnibus (GEO)
- ArrayExpress

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

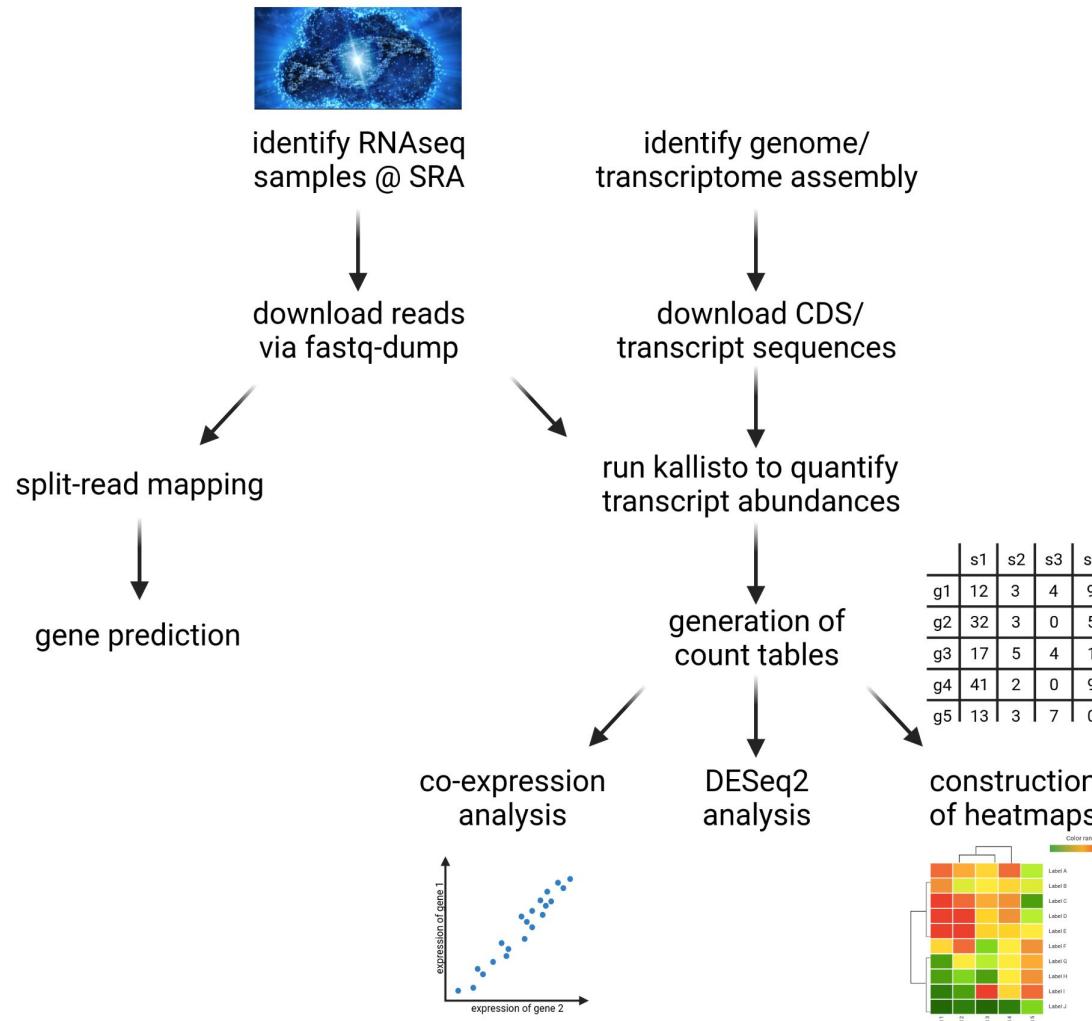
Getting Started	Tools	Browse Content
Overview	Search for Studies at GEO DataSets	Repository Browser
FAQ	Search for Gene Expression at GEO Profiles	DataSets: 4348
About GEO DataSets	Search GEO Documentation	Series: 179161
About GEO Profiles	Analyze a Study with GEO2R	Platforms: 24146
About GEO2R Analysis	Studies with Genome Data Viewer Tracks	Samples: 5167932
How to Construct a Query	Programmatic Access	
How to Download Data	FTP Site	
	ENCODE Data Listings and Tracks	

Information for Submitters		
Login to Submit	Submission Guidelines	MIAME Standards
	Update Guidelines	Citing and Linking to GEO
		Guidelines for Reviewers
		GEO Publications



<http://bar.utoronto.ca/>
<https://www.ncbi.nlm.nih.gov/geo/>
<https://www.ebi.ac.uk/arrayexpress/>

Large scale gene expression analysis

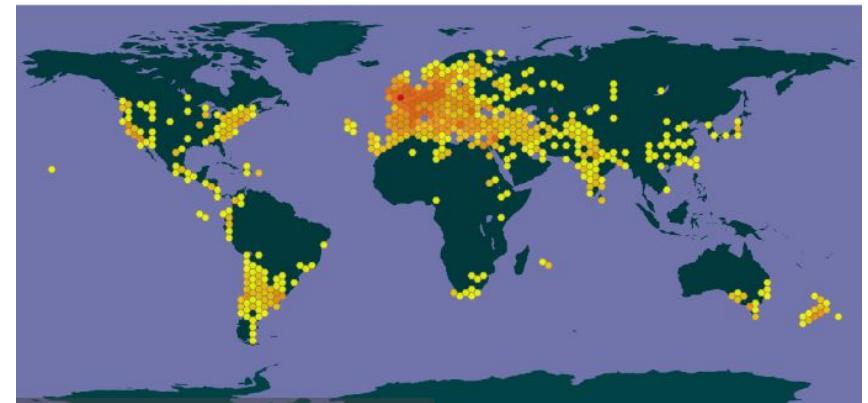


How to find the habitat of a plant species?



Global Biodiversity Information Facility (GBIF)

- Databases with all observations of a species
- Download of large data tables
- DOIs for referencing of data sets (via Datacite)
- Crawling ensures that GBIF is continuously updated



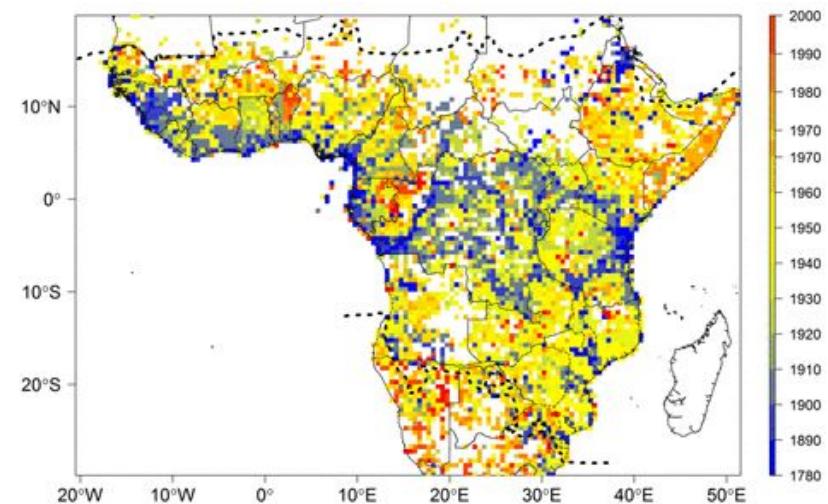
Observations of *Beta vulgaris*

Darwin Core / Darwin Core Archive

- DwC = Darwin Core
- DwC-A = Darwin Core Archive
- Biodiversity informatics data standard
- self-contained data set for species occurrences
- Reduced file size for sparse data sets
- Two components:
 - CSV file contains the data
 - meta.xml file contains information about the data structure

RAINBIO

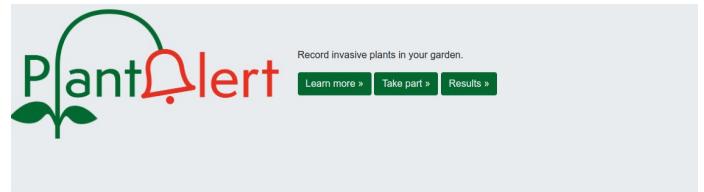
- High quality georeferenced occurrences of vascular plants in Africa
- Conservation with R (ConR) package:
<https://cran.r-project.org/web/packages/ConR/index.html>



Dauby et al., 2016: 10.3897/phytokeys.74.9723

PlantAlert

- Collecting records of invasive garden plants
- Identify invasiveness of plant species
- Challenge: identification of future problems



Plant Alert is a citizen science project for gardeners. Help other gardeners and protect the countryside by warning about invasive plants before they become a problem.

Keeping ornamental plants in the garden

While maintaining the benefits and contributions of non-native plants to our gardens, we need to know at an early stage which of those plants might go on to endanger habitats outside gardens. Gardeners across Britain and Ireland can play a crucial role here: they are likely to be the first to notice any ornamental plants showing signs of invasiveness.

Data collected from Plant Alert will be used in risk assessments of species as well as to provide gardeners and nurseries with advice on which plants could also become difficult to manage in gardens.



Himalayan balsam (*Impatiens glandulifera*) escaped from garden cultivation.
photo credit: Maja Dumat

Warn us about invasive plants in your garden before they escape >

Which plants should be reported?

Many ornamental plants in your garden will spread and this is a sign that they are growing well. We only want to know about those that are spreading to an extent that you have to control them to prevent them overgrowing other plants or parts of your garden where you do not want them. We are not really interested in garden weeds, i.e. plants that are not used as ornamental plants and may be growing in your garden (e.g. couch grass, dandelion). However, if you are not sure if a plant you have to manage in your garden is a weed or an ornamental, just report it.

Why report invasive garden plants?

Invasive non-native plants are causing major problems for native biodiversity, ecosystems, infrastructure, the built environment and human health. The majority of invasive plants have been initially introduced as ornamental garden plants and then spread from gardens into the wider environment. To prevent more species becoming invasive, gardeners can contribute by reporting early signs of invasiveness of ornamental plants in gardens.

Are all non-native plants a problem?

Only a minority of escaped ornamental species are causing problems. Well known examples include Japanese Knotweed (*Fallopia japonica*), Rhododendron (*Rhododendron ponticum*), Himalayan Balsam (*Impatiens glandulifera*) and several aquatic plants. The challenge is to identify the potential future problematic plants out of about 70,000 ornamental plants available to gardeners in Britain.

Tweets by @Plant_Alert

Plant Alert Retweeted
Invasive Species Week
@invasivesp
• • Look out for Winterprimrose this summer!
Download a free ID sheet here to help you:
nonnativeSpecies.org/assets/uploads...

If you think you've spotted it, report your sightings here: record.org.uk/enter-non-native...

Remember to #CheckCleanDry to help:
#StopTheSpread!



Jul 19, 2022

Plant Alert Retweeted
Invasive Species Week
@invasivesp
Skunk cabbage is also in our top list of species reported from gardens. But now banned from sale!
<https://twitter.com/BSBIScience/status/1542761411503349761>

Jul 4, 2022

Plant Alert Retweeted
Invasive Species Week
@invasivesp
Want to help the wildlife in your garden?
<https://www.invasivespeciesweek.org/>

Embed

View on Twitter



admin login

<https://www.youtube.com/watch?v=jn4pZyCTMls&t=2s>
<https://plantalert.org/>

Where to look for metabolite data?



MetaboLights

- Comprehensive collection of metabolite data sets
- Search by study keywords (example: flavonoids)

The image displays the MetaboLights website. At the top, there is a dark blue header bar with the MetaboLights logo (a stylized bar chart icon) and the word "MetaboLights" in white. Below the logo, there is a search bar containing the placeholder text "Examples: Alanine, Homo s...". A navigation menu bar below the search bar includes links for "Home", "Browse Studies", "Browse Compounds", "Browse Species", "Download", "Help", "Give us feedback", and "About". The main content area features a large title "MetaboLights" in blue, followed by a brief description: "MetaboLights is a database for Metabolomics experiments and derived information. The database is cross-species, cross-technique and covers metabolite structures and their reference spectra as well as their biological roles, locations and concentrations, and experimental data from metabolic experiments. MetaboLights is the recommended Metabolomics repository for a number of leading journals." Below the description are two links: "More about us" and "Quick tour >". Further down, there are three main navigation boxes: "Study" (with "BROWSE" and "ORCID SEARCH" options), "Compound Library" (with "COMPOUNDS" and "SPECIES" options), and "Training" (with "METABOLOMICS TRAIN ONLINE" and "METABOLIGHTS QUICK TOUR" options).

Example dataset

- Descriptors (keywords, title)
- Protocols: details about the methods
- Samples: information about tissue/conditions/replicates
- Assays: details about extraction and analysis in HPLC/GC
- Metabolites: list of identified metabolites with retention time, mass, chemical formula
- Files: text files and json files

Status Public Release Date 2021-12-09

MTBLS1804: Elucidations on the structures of some putative flavonoids identified in postharvest withered grapes (*Vitis vinifera L.*) by quadrupole time-of-flight mass spectrometry.

Mirko De Rosso, Amarita Panighel, Antonio Dalla Vedova, Riccardo Flaminii

Grape dehydration is an oenological process used for production of high-quality reinforced and sweet wines. High-resolution mass spectrometry (ultrahigh-performance liquid chromatography/quadrupole time-of-flight [UHPLC/QTOF]) was used to deepen the characterization of some flavonoids previously proposed in Corvina and Raboso Piave withered grapes. By performing a automated data analysis workflow and orthogonal identification approach (tandem mass spectrometry [MS/MS] , in silico fragmentation, and calculating putative retention time), elucidation on the structures of six compounds previously proposed was achieved (taxifolin-pentoside, two tetrahydroxyflavone-O-hexoside derivatives, a tetrahydroxy-dimethoxyflavone-O-hexoside derivative, a pentahydroxyflavone, and peonidin-O-pentoside); and the structures of five putative new grape flavonoids were characterized (dihydromyricetin-O-hexoside, taxifolin-di-O-hexoside, isorhamnetin, and a pinoquercetin isomer). Findings enlarge the panorama of flavonoids in grape and of their possible biosynthetic pathways.

Vitis vinifera

PUBLICATIONS

Elucidations on the structures of some putative flavonoids identified in postharvest withered grapes (*Vitis vinifera L.*) by quadrupole time-of-flight mass spectrometry

De Rosso M, Panighel A, Dalla Vedova A, Flaminii R.

Descriptors Protocols Samples Assays Metabolites Files

FTP Download Aspera Download Help

ISA METADATA search meta data

a_MTBL1804_LC_MS_alternating_reverse-phase_metabolite_profiling.txt December 09 2021 16:37:19

i_Investigation.txt December 09 2021 16:37:05

m_MTBL1804_LC_MS_alternating_reverse-phase_metabolite_profiling_v2_maf.tsv December 09 2021 16:28:48

s_MTBL1804.txt December 09 2021 16:28:48

RAW / DERIVED FILES Rawdatafile february 02 2021 08:50:38

AUDIT FILES audit December 09 2021 16:37:05

DERIVED DATA search derived files

files-all.json April 21 2022 00:19:53

metexplore_mapping.json June 17 2020 00:08:07

validation_files.json July 22 2022 18:46:07

validation_report.json July 22 2022 18:46:10

<https://www.ebi.ac.uk/metabolights/MTBLS1804/descriptors>

Where would you deposit large datasets?



Dryad

- Any file format / data type is accepted
- Support for best practice by curators
- Compliance with funder or publisher requirements
- Nonprofit membership organization
- Dryad's Data Publishing Charges (DPCs): \$120 per submission; additional \$50/10GB exceeding 50GB
- Content published under CC0 to facilitate reuse

The image shows a screenshot of the Dryad website. At the top, there are four colored boxes: blue for 'Login', orange for 'Submit', green for 'Review', and red for 'Cite'. Below these are two sections: 'How it works' and 'Why use Dryad?'.
How it works:

- Login:** Use your ORCID. If your institution is a [Dryad member](#), connect to your existing credentials.
- Submit:** Whether or not your data are related to an article, [upload](#) your data files and receive a citable DOI.
- Review:** Our [curators](#) will check through your submission to ensure the data are usable. They may contact you with advice or questions.
- Cite:** [Cite](#) and promote your data publication!

Why use Dryad?

- Any field. Any format.** Submit data in any file format from any field of research. Share all of the data from a project in one place.
- Quality control and assistance.** Our curators will check your files before they are released, and help you follow best practices.
- Straightforward compliance.** Submit your data to satisfy publisher and funder requirements for preservation and availability with a minimum of effort. We work directly with many publishers -- including Wiley, The Royal Society, and PLOS -- to streamline the process.
- Community-led.** Dryad is a nonprofit membership organization that is committed to making data available for research and educational reuse now and into the future. Modest Data Publishing Charges help ensure our sustainability.

e!DAL (IPK database)

- Hosted at the IPK
- Data sets are released under a CC0 license to enable reuse
- No restrictions with respect to file formats
- Support for review process

The screenshot shows a detailed view of a dataset entry on the e!DAL platform. At the top, the e!DAL logo and the text "e!DAL - Plant Genomics & Phenomics Research Data Repository" are visible, along with the IPK Leibniz Institute logo.

Citation: B. Pucker (2019-03-06): Chromosome-level Assembly Reveals the Niederzenz (Nd-1) Genome Structure and Gene Set. DOI:10.5447/ipk/2019/4

Abstract: SMRT sequencing read assemblies of *Arabidopsis thaliana* ecotype Nd-1 and files of the gene predictions

License: CC0 1.0 Universal (Creative Commons Public Domain Dedication)

DOI: 10.5447/ipk/2019/4

Content: 5 Directories 38 Files (2.2 GB)

Files:

//dholtgra@cebitec.uni-bielefeld.de/Chromosome-level Assembly Reveals the Niederzenz (Nd-1) Genome Structure and Gene Set [5 Directories 1 Files]
↳ Up to parent directory
↳ falcon 377.2 MB
↳ flye 370.2 MB
↳ pangenomic 153.4 MB
↳ canu 1000.7 MB
↳ miniasm 378.7 MB
↳ README.txt 446 B

Metadata

CONTRIBUTOR: Daniela Holtgräwe, Kai Bernd Stadermann, Bernd Weisshaar [Show full information]	COVERAGE: none
CREATOR: Boas Pucker [Show full information]	DATE: Event: event
PUBLISHER: Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland OT Gatersleben, Correnstraße 3, 06466, Germany	CREATED: TimePoint: Wed Mar 06 12:15:16 CET 2019
SIZE: 2.2 GB	UPDATED: TimePoint: Wed Mar 06 12:19:23 CET 2019
SUBJECT: <i>Arabidopsis thaliana</i> ecotype Niederzenz Nd-1, pseudochromosomes, chondrome, PacBio assembly, plastome sequence, gene prediction, Canu, Falcon, Flye, Miniasm	LANGUAGE: de_DE
	RELATION: none
	SOURCE: none

Revision: 2 - CreationDate: Wed Mar 06 12:15:16 CET 2019 - RevisionDate: Wed Mar 06 12:19:23 CET 2019

GigaDB

- Publication of data sets belonging to GigaScience articles
- CC0 to enable future studies
- Data sets are citeable by DOIs

Repository details

GigaDB

[General](#) [Institutions](#) [Terms](#) [Standards](#)

Name of repository **GigaDB**
Additional name(s) GigaScience Database
Repository URL <http://gigadb.org/>
Subject(s) Basic Biological and Medical Research | Plant Genetics | Animal Genetics, Cell and Developmental Biology | Metabolism, Biochemistry and Genetics of Microorganisms | Human Genetics | Biology | Life Sciences | Plant Sciences | Zoology | Microbiology, Virology and Immunology | Medicine | Medicare | Software Technology | Neurosciences | Computer Science | Computer Science, Electrical and System Engineering | Engineering Sciences

Description GigaDB primarily serves as a repository to host data and tools associated with articles published by GigaScience Press; GigaScience and GigaByte (both are online, open-access journals). GigaDB defines a dataset as a group of files (e.g., sequencing data, analyses, imaging files, software programs) that are related to and support a unit-of-work (article or study). GigaDB allows the integration of manuscript publication with supporting data and tools.

Contact database@gigasciencejournal.com

Content type(s) Source code | Images | Audiovisual data | Raw data | Plain text | Archived data | Images | Scientific and statistical data formats | other

Keyword(s) CT data | FAIR | MRI image data | biocuration | biodiversity | biology | biomedical sciences | biomedicine | data publishing | genomics | human genetics | microscopy data | optical imaging | transcriptome sequence

Persistent identifier(s) of the repository RRID:SCR_004002
RRID:nix_158413
FAIRsharing_doi:10.25504/FAIRsharing.rcbwsf

Repository size 2068 datasets

Repository type(s) disciplinary

Mission statement for designated community <http://gigadb.org/site/term>

Research data repository language(s) English

Data and/or service provider data provider

[Back to search](#) [Submit a change request](#) [Get a badge](#)



Cite this re3data.org record:

re3data.org: GigaDB; editing status 2021-07-19; re3data.org - Registry of Research Data Repositories.
[http://doi.org/10.17616/R3TG83](https://doi.org/10.17616/R3TG83) last accessed: 2022-07-23

Institute-specific data publications

- Data publications can be hosted at libraries
- Assignment of DOIs
- TUBS is planning data publication service

PUB - Publications at Bielefeld University

[Home](#) [Publications](#) [Authors](#) [About](#) ▾

Deutsch [Login](#)

[Open Access at
Bielefeld University](#)



[Research Data
Management](#)



[PUB Theses](#)



DINI-Certificate 2019

"PUB - Publications at Bielefeld University" is the institutional repository of Bielefeld University. It constitutes the central depot for publication metadata, publications and research data with the mission to reflect the scientific work of the university's researchers. Bielefeld scientists of all disciplines can use this service to create and maintain their personal publication lists. The records are freely accessible and partially linked with fulltexts and research data. [Learn more...](#)

- 80167 [Publication References](#)
Bibliographic data, partially enriched with fulltexts (PDFs etc.)
- 5499 [Individual Author Pages](#)
Documentation of Bielefeld researchers' publishing activities
- 393 [Data Publications](#)
Bibliographic data, partially enriched with research data
- 14459 [Open Access Publications](#)
Freely accessible documents
- 2501 [Theses](#)
Bielefeld dissertations, habilitation, and selected master and bachelor theses

Time for questions!



Questions

1. Where do you find scientific literature?
2. Which sequence databases do you know?
3. What are the different BLAST types?
4. What are the advantages/disadvantages of DIAMOND?
5. Where can you find visualizations of pathways?
6. What is a genome browser?
7. Which information can be displayed in a genome browser?
8. What are the important steps in processing SRA data sets for gene expression?
9. Where can you find information about the geographic distribution of plant species?
10. Where would you deposit the large datasets of your next publication?