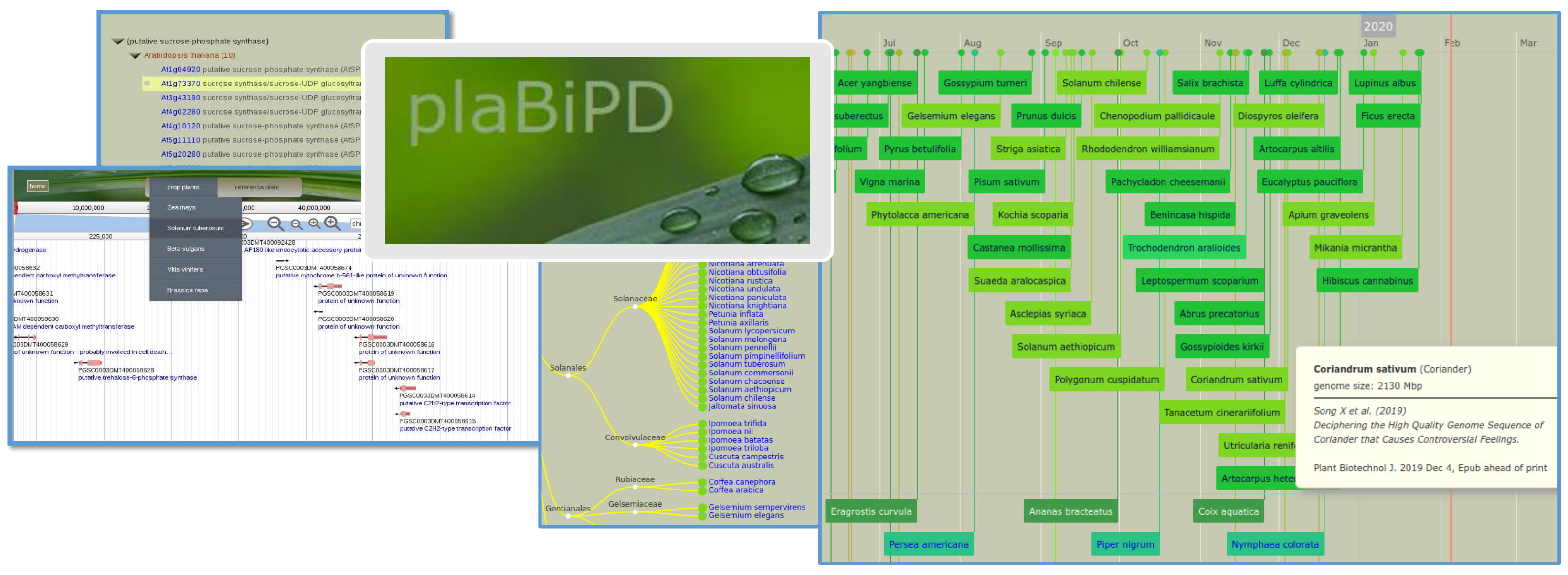
GCBN German Crop Bio Greenformatics Network @ FZJ

Marie Bolger, Rainer Schwacke, Björn Usadel
Forschungszentrum Jülich, Germany

Fk031A536C

Short description of the project

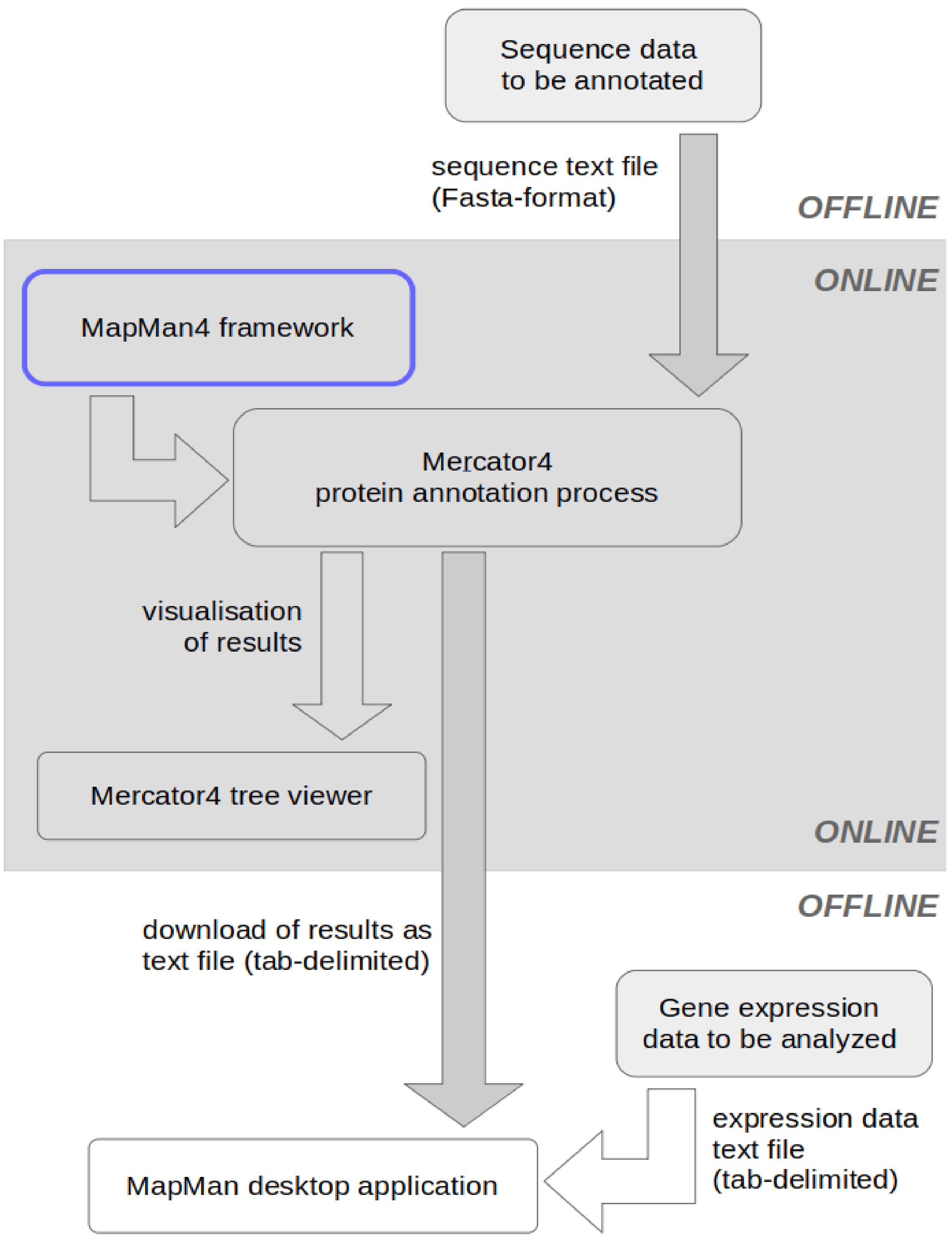
The Forschungszentrum Jülich (FZJ) node of GCBN provides tools and resources tailored to the plant research community. With a particular focus on crop plants, the web resource *www.plabipd.de* serves as the access point. FZJ tracks advances in published sequenced plant genomes and develops algorithms for functional annotation of plant proteins.



Screenshots of some resources available at *www.plabipd.de*

de.NBI services

The **Mercator4** online annotation service allows users to submit plant protein or DNA sequences for classification into **MapMan4** categories. The results from Mercator4 can be compared online to reference plant species or other submitted user jobs.
They can also be downloaded and directly used to explore RNASeq data in the MapMan desktop application on pathway diagrams.


Sequence data to be annotated

sequence text file (Fasta-format)

OFFLINE

ONLINE

MapMan4 framework

Mercator4 protein annotation process

visualisation of results

Mercator4 tree viewer

ONLINE

OFFLINE

download of results as text file (tab-delimited)

Gene expression data to be analyzed

expression data text file (tab-delimited)

MapMan desktop application

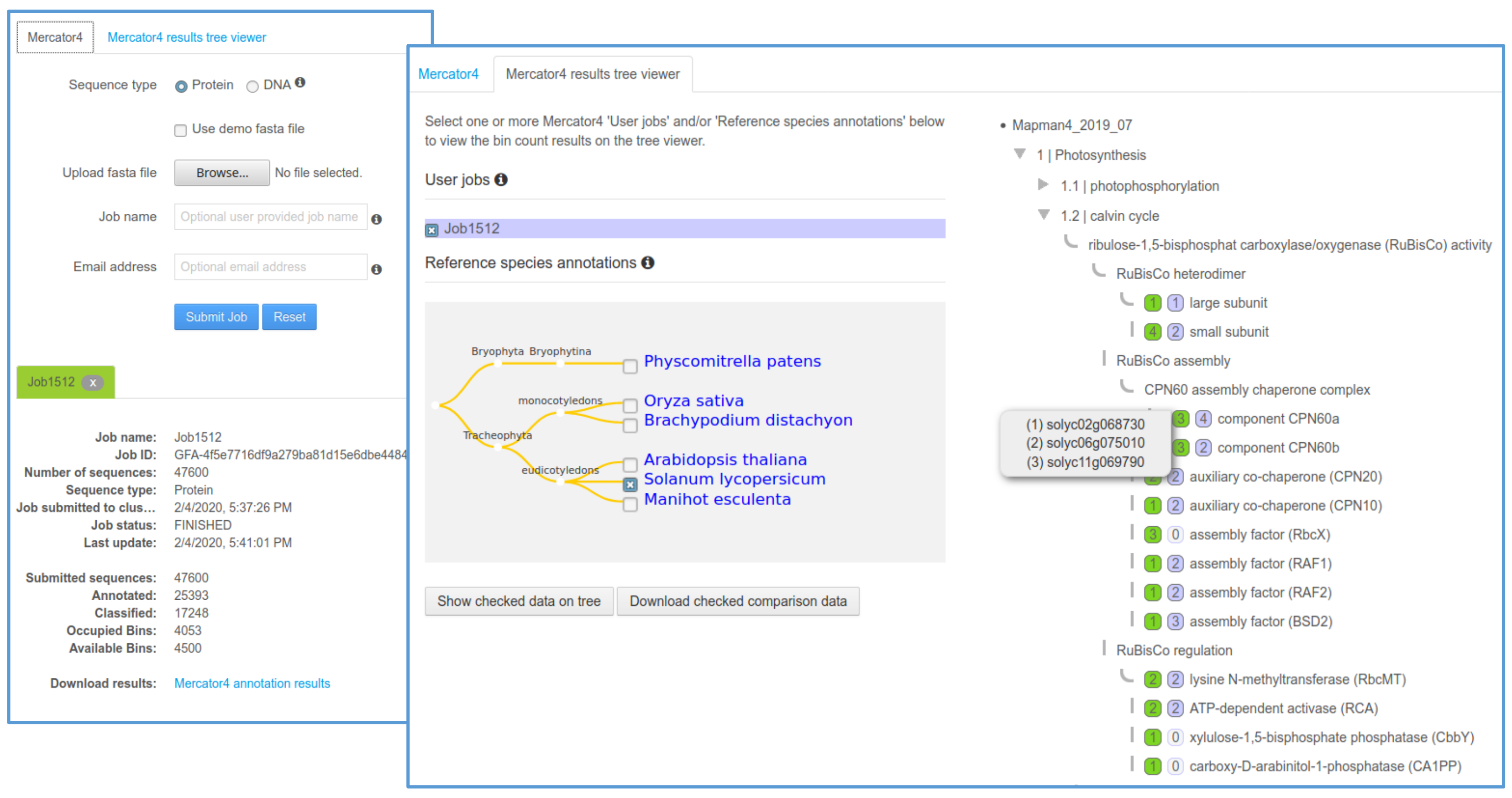
Overview of the Mercator4 plant protein annotation process.

General information on the project

Core group Björn Usadel (PI), Rainer Schwacke (0.5x de.NBI-position, Mapman4, *www.plabipd.de*) and Marie E. Bolger (Mercator4, www.*plabipd.de*). The project is supported by Anthony M. Bolger (HPC cluster and programming).

Progress report

The FZJ node of GCBN has further developed the **MapMan4** annotation framework for protein function predictions as the foundation of the **Mercator4** service. This is driven by clustering and analysis of proteins from publicly available plant proteomes and has led to the generation of more than 5900 protein categories (MapMan4 v.2, July 2019). In order to increase annotation speed, the back-end high performance computing (HPC) cluster has been further upgraded and extended. The high quality annotations together with the cluster upgrades enable protein function prediction for whole plant proteomes and transcriptomes within minutes.



Sequence type ● Protein ○ DNA
☐ Use demo fasta file
Upload fasta file: Browse... No file selected.
Job name: Optional user provided job name
Email address: Optional email address

Job name: Job1512
Job ID: GFA-4f5e7716df9a279ba81d15e6dbe4484
Number of sequences: 47600
Sequence type: Protein
Job submitted to clus...: 2/4/2020, 5:37:26 PM
Job status: FINISHED
Last update: 2/4/2020, 5:41:01 PM

Submitted sequences: 47600
Annotated: 25393
Classified: 17248
Occupied Bins: 4053
Available Bins: 4500

Download results: Mercator4 annotation results

Select one or more Mercator4 'User jobs' and/or 'Reference species annotations' below to view the bin count results on the tree viewer.

User jobs
☒ Job1512

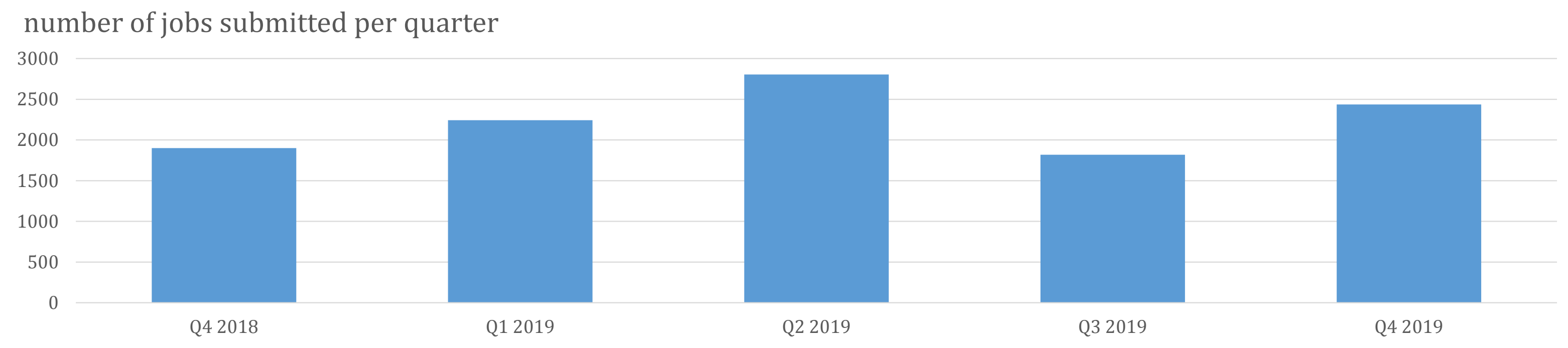
Reference species annotations

☐ Physcomitrella patens
☐ Oryza sativa
☐ Brachypodium distachyon
☐ Arabidopsis thaliana
☒ Solanum lycopersicum
☐ Manihot esculenta

- Mapman4_2019_07
  - 1 | Photosynthesis
    - 1.1 | photophosphorylation
    - 1.2 | calvin cycle
      - ribulose-1,5-bisphosphat carboxylase/oxygenase (RuBisCo) activity
        - RuBisCo heterodimer
          - 1 1 large subunit
          - 4 2 small subunit
        - RuBisCo assembly
          - CPN60 assembly chaperone complex
            - 3 4 component CPN60a
            - 3 2 component CPN60b
          - 2 auxiliary co-chaperone (CPN20)
          - 1 2 auxiliary co-chaperone (CPN10)
          - 3 0 assembly factor (RbcX)
          - 1 2 assembly factor (RAF1)
          - 1 2 assembly factor (RAF2)
          - 1 3 assembly factor (BSD2)
        - RuBisCo regulation
          - 2 2 lysine N-methyltransferase (RbcMT)
          - 2 2 ATP-dependent activase (RCA)
          - 1 0 xylulose-1,5-bisphosphate phosphatase (CbbY)
          - 1 0 carboxy-D-arabinitol-1-phosphatase (CA1PP)

(1) solyc02g068730
(2) solyc06g075010
(3) solyc11g069790

Screenshots of the Mercator4 online annotation service.



Usage of the Mercator4 online annotation service.

Additionally, FZJ maintains a selection of bioinformatics tools (**MapMan desktop application**, **RobiNA** and **Trimmomatic**) for the analysis and interpretation of sequencing data. The command line tool Trimmomatic is used for trimming Illumina data sets and has been incorporated into automated cluster pipelines and Galaxy pipelines. RobiNA allows the visual analysis of large microarray or small RNAseq data sets and provides the input for the MapMan desktop application, a tool to analyse plant functional pathways and visualize transcriptomics, metabolomics and proteomics quantitative data.

de.NBI Training and education

de.NBI – CeBiTec Nanopore Best Practice Workshop 2019
Bielefeld, 25. - 27. Sep. 2019

Publications

Schwacke R, Ponce-Soto GY, Krause K, Bolger AM, Arsova B, Hallab A, Gruden K, Stitt M, **Bolger ME**, **Usadel B** (2019)
Mapman4: a refined protein classification and annotation framework applicable to multi-omics data analysis.
Mol Plant. 2019 Jun 3, 12(6): 879-892

Bolger AM, Poorter H, Dumschott K, **Bolger ME**, Arend D, Osorio S, Gundlach H, Mayer KFX, Lange M, Scholz U, **Usadel B** (2019)
Computational aspects underlying genome to phenome analysis in plants.
Plant J. 2019 Jan, 97(1): 182-198

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
Projektträger Jülich Forschungszentrum Jülich