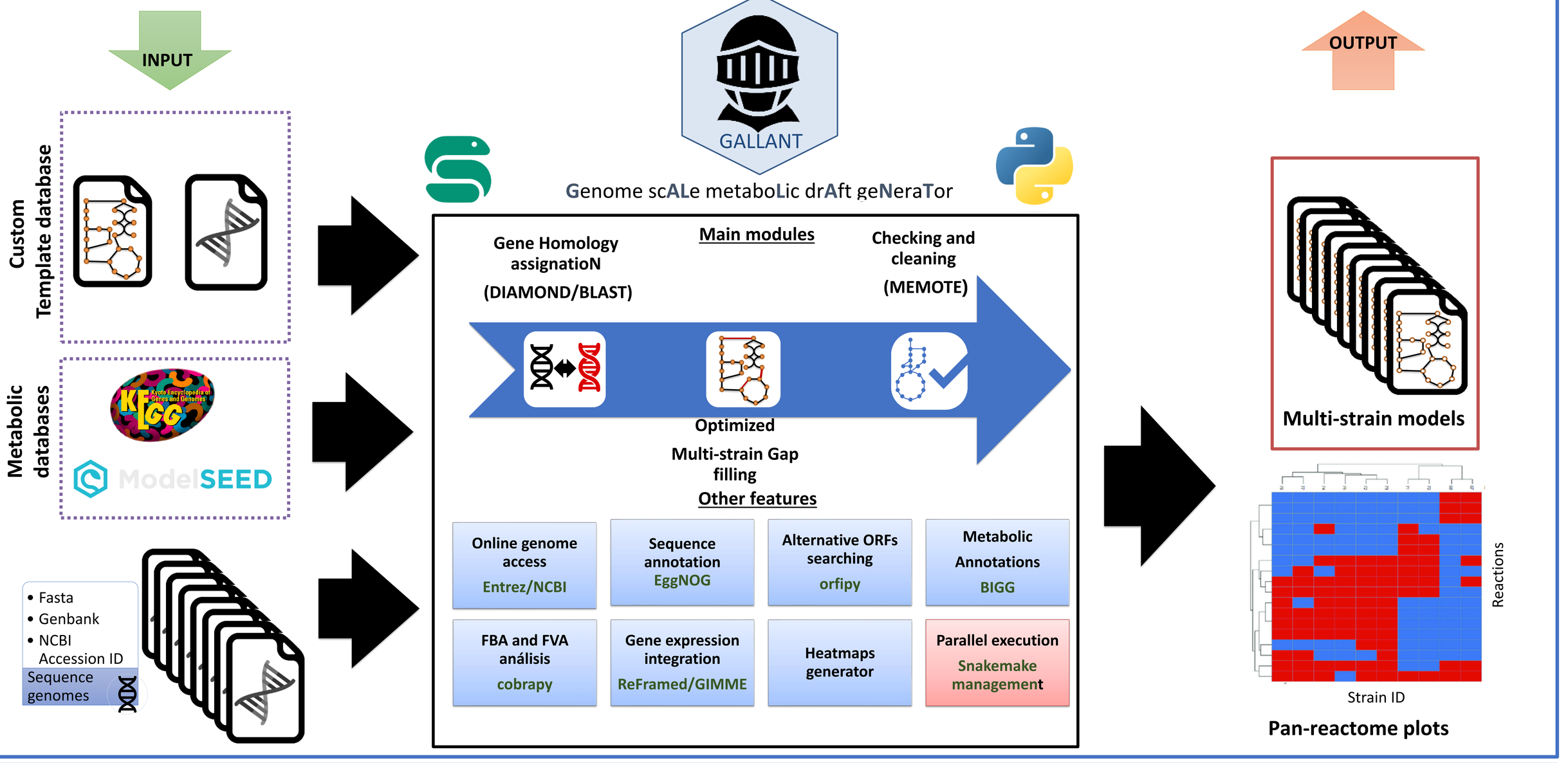
GALLANT: A standardized workflow for multi-strain Genome-scale metabolic modeling.

David San León Granado (dsanleon@cnb.csic.es), Juan Nogales Enrique. Systems Biology, National Center for Biotechnology (CNB-CSIC), Madrid, Spain

SUMMARY

Genome-scale metabolic models (GEMs) of bacteria are one of the most important drivers of recent advances in metabolic engineering and systems biology. However, single strain GEMs define the metabolic capabilities of the strain in question, thus limiting the ability to perform metabolic studies at the species level. To overcome this, here we present GALLANT, a standardized and reproducible workflow that returns multi-strain GEMs to analyze the unique capabilities of individual strains, but with the added value that it is also able to group GEMs according to their pan-genome in order to perform studies of metabolic potential at the species level. The main goal of GALLANT is to deliver output GEMs to the highest standards. To this end, GALLANT uses a high-quality, customized GEM database as a template and it combines it with other major databases like BIGG and KEGG to return final GEM drafts. To reduce the need for subsequent manual curation, it features an optimized automatic gap-filling module for multi-strain modeling, while built-in GEM quality assurance tools such as MEMOTE guide the user in the process of refining final results. GALLANT's robust modular design supports parallel delivery of single strain GEMs and it is fully geared up to be run on high performance servers in order to scale up the process and reduce execution times considerably.

WORKFLOW DESCRIPTION



RESULTS

Input

Template

Bacillus Subtilis sp. 168

A curated *Bacillus Subtilis* sp.168 GEM based on iBsu1103 GEM was used as template.

Targets

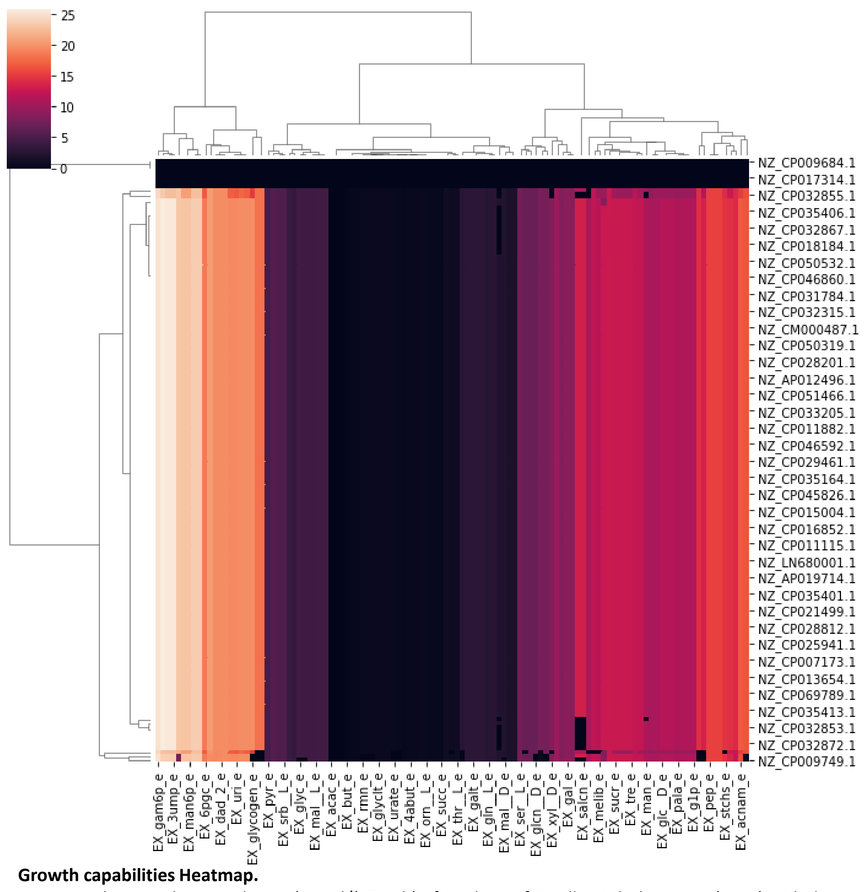
All *Bacillus subtilis* with complete genome downloaded from NCBI

488 strains

Bacillus Subtilis homology analysis

Homology Heatmap. The color represents if there is homology between a gene (row) from the *Bacillus subtilis* 168 template and the correspondence strain (columns). Black represents ausence and light color represents presence. The binary data was hierarchical clustered and the selected distance was Hamming distance. The strain and gene labels are hidden to increase the visibility. Only homologies over 30% of identity (protein level), 25 % of coverage and an e-value below 0.01 are considered

Carbon source - Growth capabilities



Growth capabilities Heatmap.
Heatmap showing the growth rate (mmol/(gDW h) of a subset of *Bacillus Subtilis* strains (rows) with diverse carbon sources (columns). The growth rate was calculated with Flux Balance Analysis (FBA) implemented in cobrapy and a minimal media. The column labels represents the Exchange pseudo-reactions of multiple carbon sources (BIGG ids) and the row labels contains the NCBI *Bacillus subtilis* IDs.i