

9. Introduction to relational annotation

Guillaume GAUTREAU, 29/09/2022

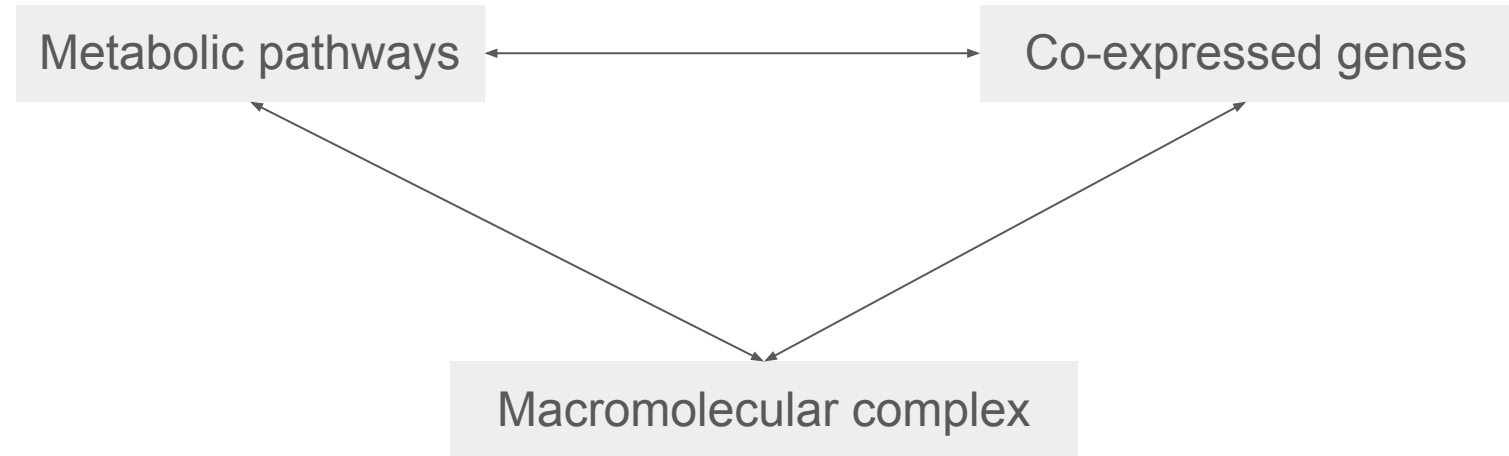
How the previously identified proteins work together ?

Metabolic pathways

Co-expressed genes

Macromolecular complex

How the previously identified proteins work together ?



How the previously identified proteins work together ?

Metabolic pathways

Co-expressed genes

Macromolecular complex

Metabolic pathways

Metabolic pathways are composed of:

- **Compounds** transformed by....
- **Reactions** which are catalyzed by...
- **Enzymes** detected through the functional annotation

Database of Compounds : ChEBI

Database of Reactions : RHEA

Database of Enzymes:

- Expasy Enzyme
- BRENDA (can be identified via PRIAM)

Database of Pathways :

- KEGG
- BioCyc

ChEBI (Chemical Entities of Biological Interest)

- Dictionary of molecular entities focused on 'small' chemical compounds :
 - atom
 - molecule
 - ion
 - ion pair
 - radical
 - radical ion
 - complex
 - conformer.
- Could be natural or synthetic compounds but linked to biological reactions
- Example: D-glucopyranose 6-phosphate²⁻

<https://www.ebi.ac.uk/chebi/searchId.do?chebId=CHEBI:61548>

REHA (Swiss Institute of Bioinformatics)

- A database of chemical and transport reactions of biological interest
- Based on ChEBI compounds ID
- Reaction are curated from peer-review literature
- Example: glucose-6-phosphate dehydrogenase
 - <https://www.rhea-db.org/rhea/15841>

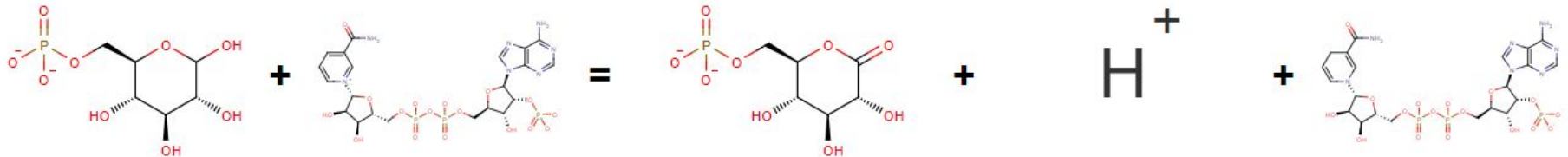
D-glucose 6-phosphate

NADP⁺

6-phospho-D-glucono-1,5-lactone

H⁺

NADPH



zoom

zoom

zoom

zoom

zoom

Expasy Enzyme (Swiss Institute of Bioinformatics)

- Example glucose-6-phosphate dehydrogenase is catalyzed by :
 - An enzyme classified 1.1.1.49
 - <https://enzyme.expasy.org/EC/1.1.1.49>
- Reminder: enzyme hierarchical classification from the Enzyme Commission:
 - The 1st digit defines the general type of reaction
 - The 2nd digit indicates the subclass.
 - The 3rd digit gives the sub-subclass.
 - The 4th digit is the serial number of the enzyme in its sub-subclass.

EC class	Reaction type
1	Oxidoreductases
2	Transferases
3	Hydrolases
4	Lyases
5	Isomerases
6	Ligases

BRENDA

- Similar to Expasy Enzyme but with a bit more information
- Same example :
 - <https://www.brenda-enzymes.org/enzyme.php?ecno=1.1.1.49>
- BRENDA activities can be inferred using PRIAM : <http://priam.prabi.fr/>



BRENDA



The Comprehensive Enzyme Information System

KEGG : Kyoto Encyclopedia of Genes and Genomes

- A massive database linking all the known information around pathways
- Main functionalities :
 - KEGG COMPOUND: similar to ChEBI
 - KEGG REACTION: similar to REHA
 - KEGG ENZYME: similar to ExPASy Enzyme and BRENDA
 - KEGG PATHWAY : manual drawn most important pathways
 - example : <https://www.genome.jp/pathway/map00010+C00668>
 - KEGG ORTHOLOGY : family of ortholog proteins
 - KEGG MODULE : manually defined functional units of orthologs sets.
 - Example: glycolysis <https://www.genome.jp/module/M00001+C00668>
 - and many more

MetaCyc / BioCyc

- Similar functionalities to KEGG but some differences :
 - No equivalent of KO orthologs
 - Associated to a local software: Pathway tools allowing many complicated thing as for example:
 - to predict pathways from a set of EC number with gap filling
 - Commercial use is restricted
- Example glycolysis:
 - <https://biocyc.org/ECOLI/NEW-IMAGE?type=PATHWAY&object=GLYCOLYSIS>

How the previously identified proteins work together ?

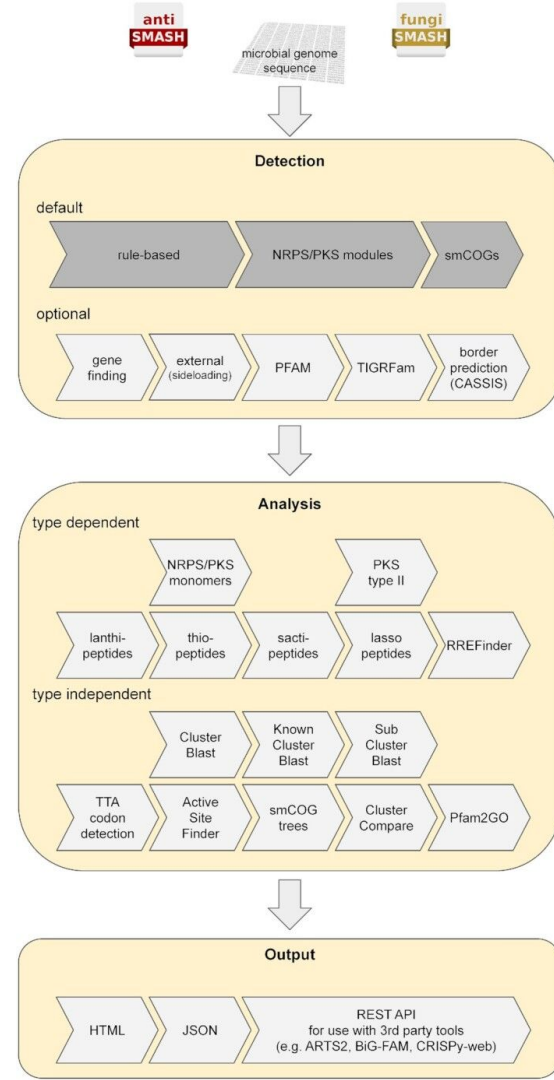
Metabolic pathways

Co-expressed genes

Macromolecular complex

AntiSMASH

- Allows the rapid genome-wide identification of secondary metabolite biosynthesis gene clusters in bacterial and fungal genomes.
- Example of biological process detected:
 - Polyketides, non-ribosomal peptides, terpenes, aminoglycosides, aminocoumarins, indolocarbazoles, lantibiotics, bacteriocins, nucleosides, beta-lactams, butyrolactones, siderophores, melanins and others
- <https://antismash.secondarymetabolites.org/#!/start>



How the previously identified proteins work together ?

Metabolic pathways

Co-expressed genes

Macromolecular complex

Synteny and Operon

- Synteny :
 - Physical co-localization of gene in a same or similar genetic loci (synton)
 - Can be identified by comparative genomic
 - Often corresponds to operons
 - Often a correspondence between gene proximity and metabolic pathways

- Operon : genes expressed under the control of a common operator
 - Can be detected via decided methods :
 - OperonHunter: <https://github.com/ridassaf/OperonHunter>
 - Octopus: <https://github.com/FredHutch/octopus>
 - OperonFinder: <https://www.mefyi.com/operon>
 - Some Known and Predicted Operons: <http://www.microbesonline.org/>

How the previously identified proteins work together ?

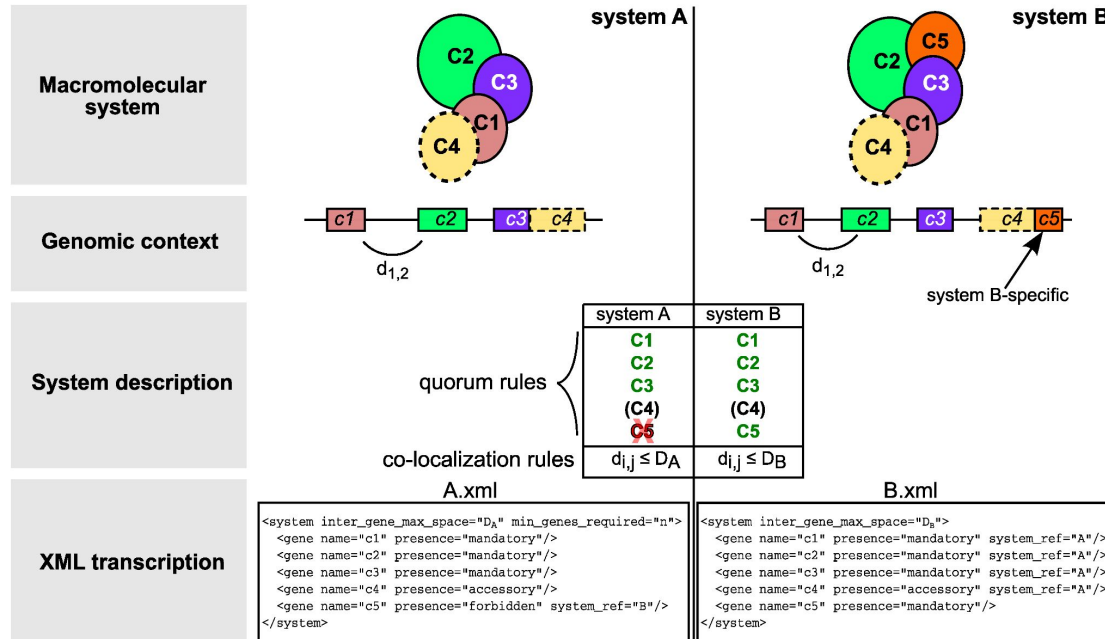
Metabolic pathways

Co-expressed genes

Macromolecular complex

MacSyFinder

- Genes of macromolecular systems are often colocalized
- Using dedicated rules, MacSyFinder is able to predict them



STRING database

- Database of known and predicted protein-protein interactions
- The interactions include direct (physical) and indirect (functional) associations :
 - from computational prediction
 - from knowledge transfer between organisms
 - from interactions aggregated from other (primary) databases.

- <https://string-db.org/>

- Example: LacZ
 - <https://string-db.org/cgi/network?taskId=bjoDRPJ5Clrz&sessionId=brjeSWzHBa78>

Conclusion

- Many databases for relational annotation too
- A lot of them are not described here
- Many new tools are released each month (try to keep up to date)
- Don't forget ncRNA in macromolecular complexes