

# Handling the heterogeneity of genomic and metabolic networks data within flexible workflows with the PADMet Toolbox

---

Marie Chevallier<sup>1</sup>, Meziane Aite<sup>1</sup>, Jeanne Got<sup>1</sup>, Guillaume Collet<sup>1</sup>, Nicolas Loira<sup>2</sup>, Maria Paz Cortes<sup>2</sup>,  
Clémence Frioux<sup>1</sup>, Julie Laniau<sup>1</sup>, Camille Trottier<sup>1</sup>, Alejandro Maas<sup>2</sup>, and Anne Siegel<sup>1</sup>

<sup>1</sup> DYLISS (INRIA - IRISA) – INRIA, Université de Rennes 1, CNRS : UMR6074 – Campus de Beaulieu 35042 Rennes cedex,  
France

<sup>2</sup> Centro de Modelamiento Matematico (CMM) – Chili

# Context : high scale system biology

2

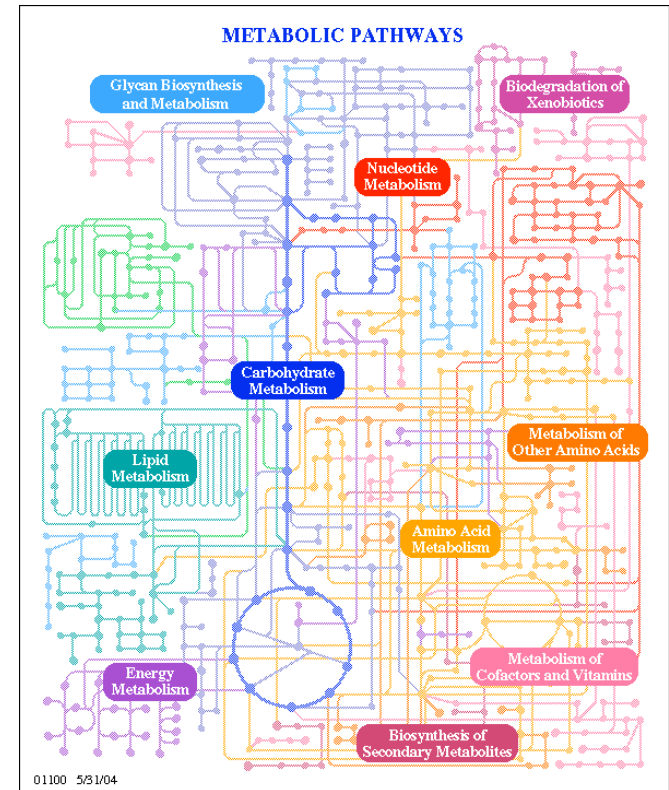
- Emergence of new technologies
  - ▣ Massive genome sequencing
  - ▣ High-throughput production of transcriptomic and metabolomic data
- Very high quantity of raw data
- How to manage these high scale data ?

Comprehend these omic data while reconstructing a metabolic network

# Genome scale metabolic network (MN)

3

- Describes all chemical and physical processes in a cell
  - ▣ Chemical reactions taking part into the cell metabolism
  - ▣ Transport reactions of compounds



Kegg

How to master such a complex and high connected network ?

# Metabolic network (MN) : modeling

4

## □ Bipartite directed graph

### ▣ Nodes :

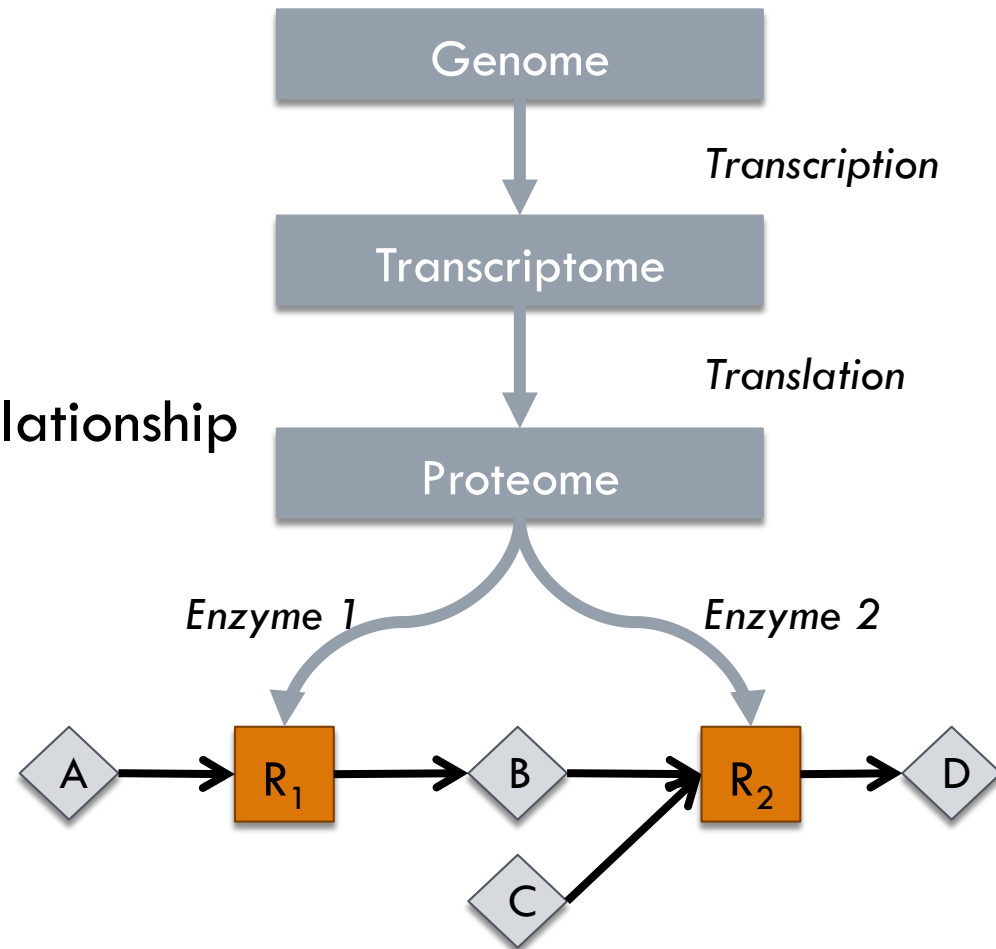
Reactions and Compounds

### ▣ Edges :

Consumption/Production relationship

◇ Compound

■ Reaction



# Metabolic network (MN) : modeling

5

## □ Bipartite directed graph

### ▣ Nodes :

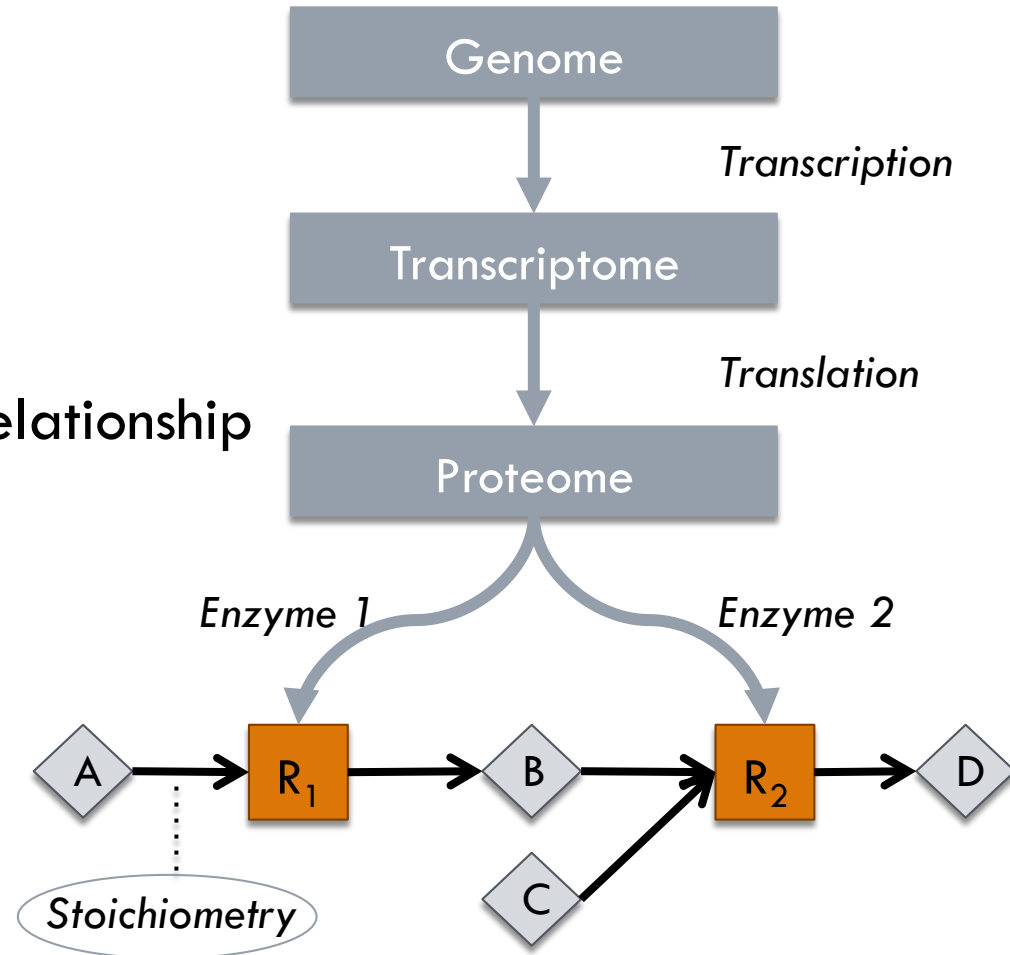
Reactions and Compounds

### ▣ Edges :

Consumption/Production relationship  
qualitative or quantitative

 Compound

 Reaction



# Metabolic network (MN) : modeling

6

- Bipartite directed graph
  - Nodes :  
Reactions and Compounds
  - Edges :  
Consumption/Production relationship
  
- SBML (Systems Biology Markup Language)
  - Standard format (based on XML)
  - Low flexibility (supplementary data)

Modeling standards : bipartite graph, SBML format

# SBML format drawbacks

7

- Not suited for supplementary data
- "note" section data : difficult to parse

```
</reaction>
<reaction id="RXN_45_13524" name="RXN-13524" reversible="false">
  <notes>
    <body xmlns="http://www.w3.org/1999/xhtml">
      <p>BIOCYC: RXN-13524</p>
      <p>EC Number: 1.14.99.-</p>
      <p>SUBSYSTEM: PWY-5748</p>
      <p>Confidence level: 0</p>
    </body>
  </notes>
  <listOfReactants>
    <speciesReference species="CPD_45_9424_c" stoichiometry="1"/>
    <speciesReference species="Donor_45_H2_c" stoichiometry="1"/>
    <speciesReference species="OXYGEN_45_MOLECULE_c" stoichiometry="1"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="CPD_45_14516_c" stoichiometry="1"/>
    <speciesReference species="Acceptor_c" stoichiometry="1"/>
    <speciesReference species="WATER_c" stoichiometry="1"/>
    <speciesReference species="PROTON_c" stoichiometry="1"/>
  </listOfProducts>
</reaction>
```

SBML format : inadequate way to manage additional data

# Reconstruction of metabolic networks

8

## ① Building a draft

- ▣ Genomic information : sequence annotation
- ▣ Orthology information : taxonomically related species

## ② Gap-filling the draft

- ▣ Target compounds producible from seeds compounds
- ▣ Topological or stoichiometric

➤ *Thiele & Palsson 2010*

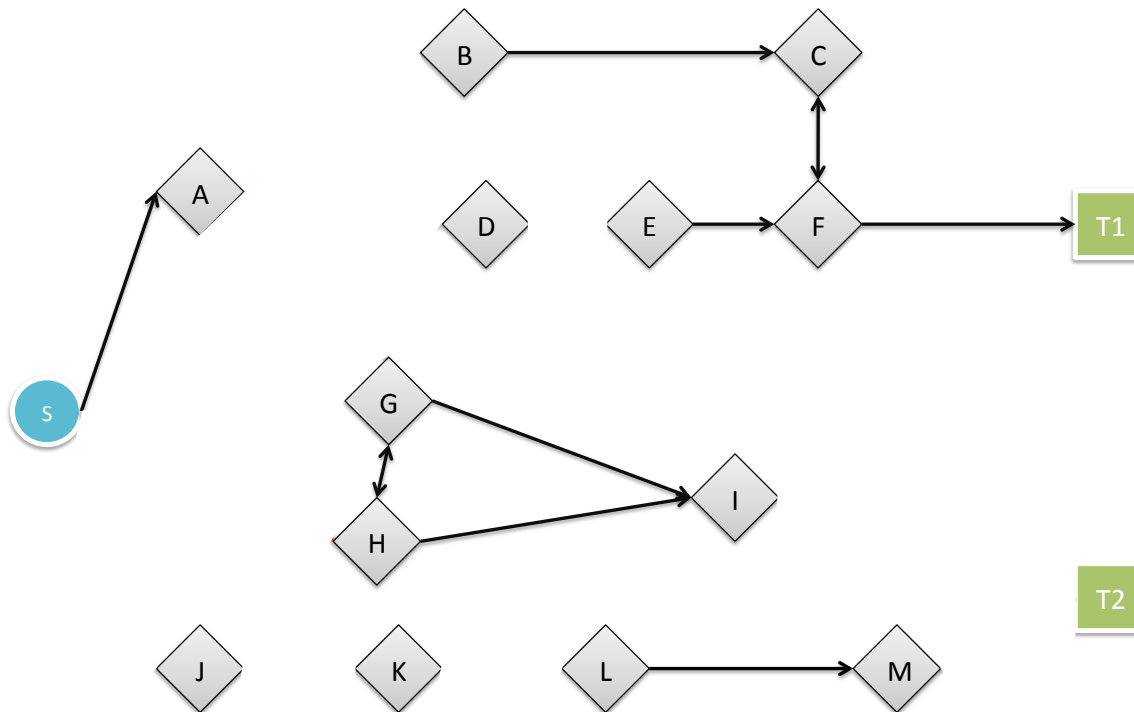
Studying non-model species require to consider every source of data



# Reconstruction of metabolic networks

9

## ① Building a draft



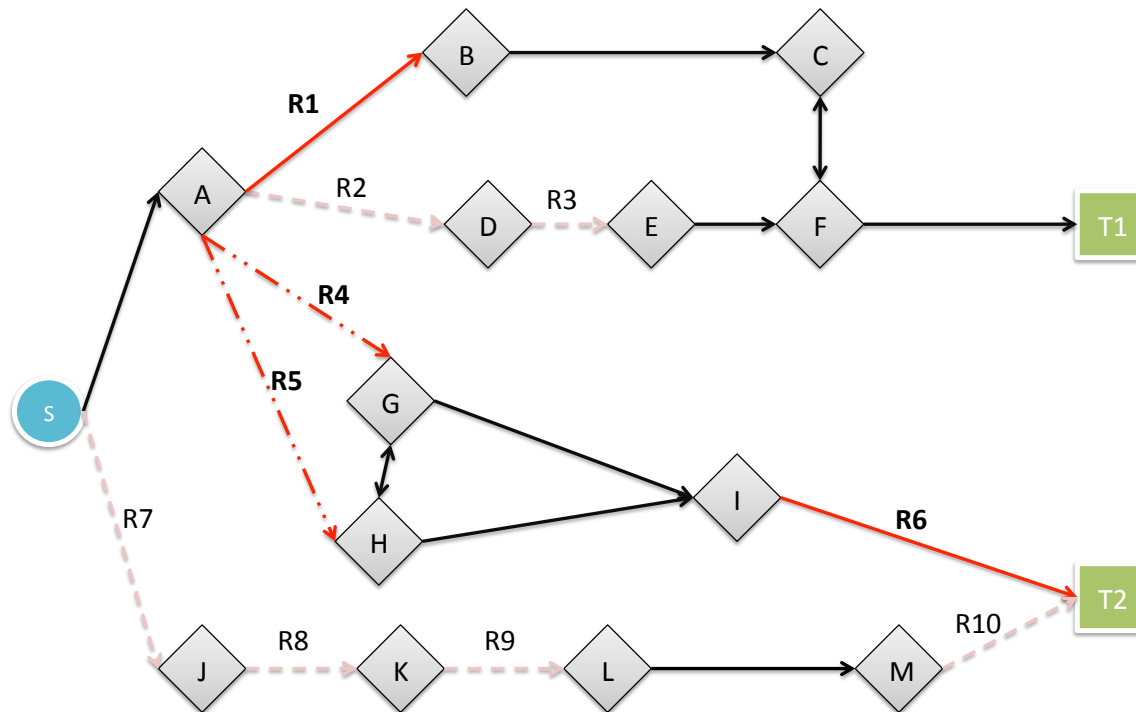
Draft MN for non-model species are filled with wholes

# Reconstruction of metabolic networks

10

## ② Gap-filling the draft

- Select solutions



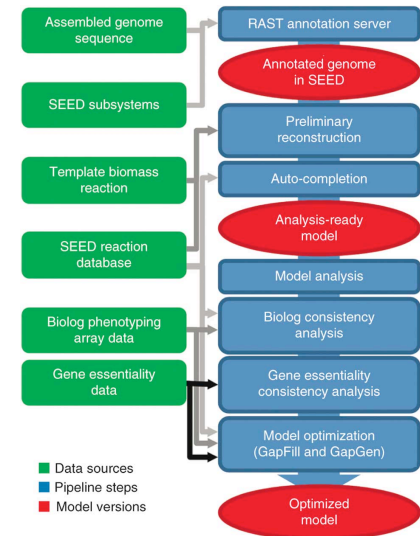
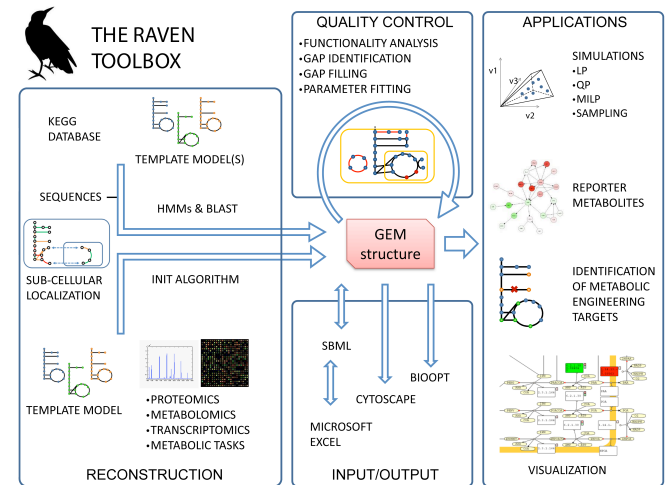
Several ways to select the best solution(s)

# Workflows for MN reconstruction

11

- Involve a huge pool of tools
- Need specific steps of pre/post-processing of data
- Use a lot of different input formats
- Require to carefully manage data until the output  
Standardization, Heterogeneity, Traceability
- Often linked to a specific data base

Existing workflows require deep expertise from users



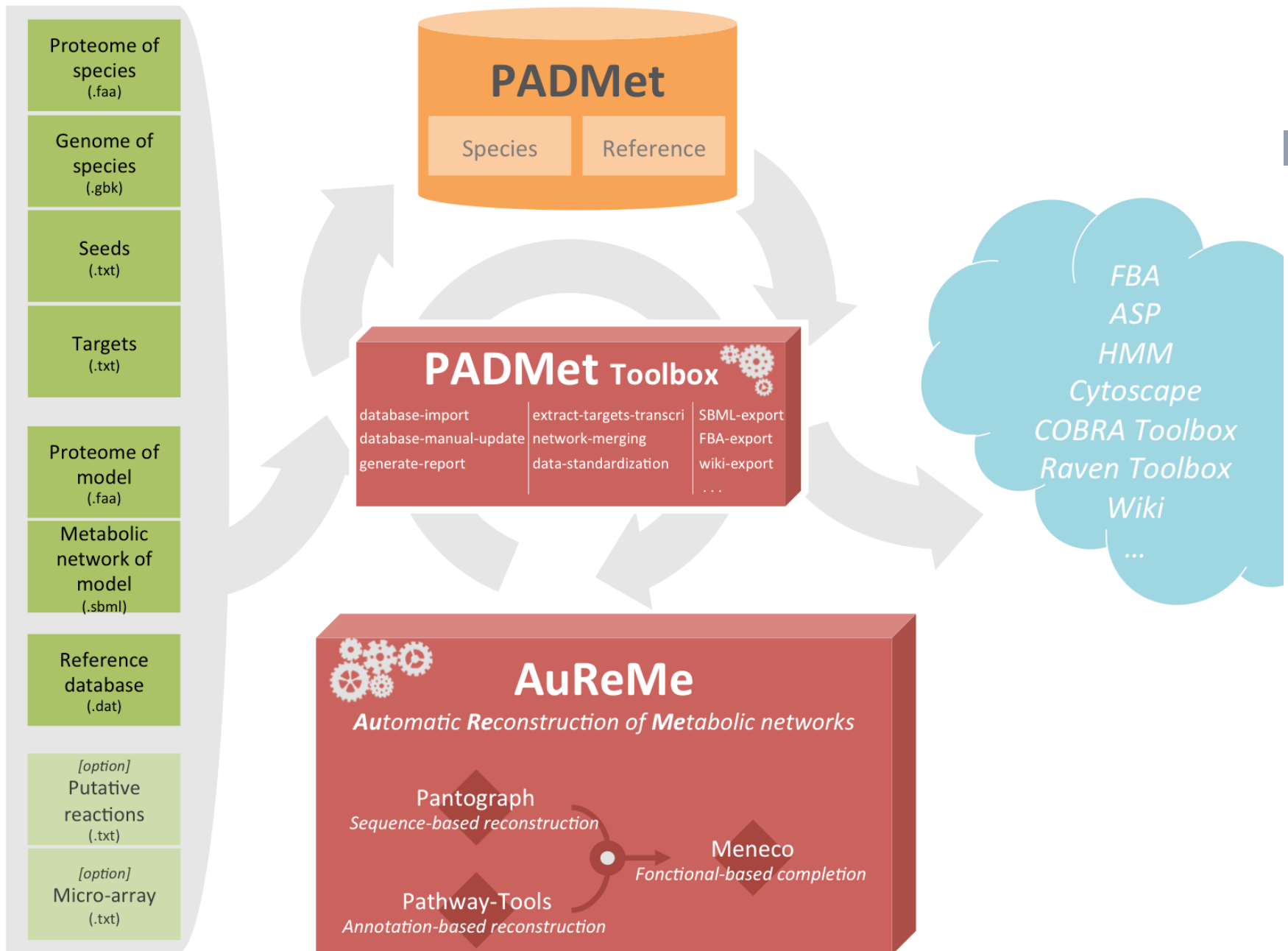
Model Seed

# Goals : centralizing and structuring data

12

- One *Toolbox* to *rule* them all
- One *Toolbox* to *find* them
- One *Format* to *bring* them all
- And in the *Workflow*, *bind* them

13



# The PADMmet format



14

## PADMmet : PortABle Database for METabolism

```

codes

:reaction RXN-13524 DIRECTION LEFT-TO-RIGHT COMMON-NAME pseudo-conhydrine hydroxylase EC-NUMBER EC-1.14.99
:pathway FWY-5748 COMMON-NAME 4gamma;-coniciene and conine biosynthesis
:compound CPD-14516 SMILES CCCC1(N=CC(O)CC1) COMMON-NAME 5-hydroxy-4gamma;-coniceine INCHI-KEY InChIKey=BVUNSAIHNEMOAG-UHFFFAOYSA-N MOLECULAR-WEIGHT
:compound CPD-9424 SMILES CCCC1(=[N+])CCCC1) COMMON-NAME 4gamma;-coniceine INCHI-KEY InChIKey=PQSOVRNZJIENNW-UHFFFAOYSA-O MOLECULAR-WEIGHT 126.221
:compound WATER SMILES [OH2] COMMON-NAME H2O INCHI-KEY InChIKey=XLYOFWQVPEJUNP-UHFFFAOYSA-N MOLECULAR-WEIGHT 18.015
:compound OXYGEN-MOLECULE SMILES O=O COMMON-NAME oxygen INCHI-KEY InChIKey=MYMOP12GZYHOMD-UHFFFAOYSA-N MOLECULAR-WEIGHT 31.999
:compound PROTON SMILES [H+] COMMON-NAME H+ INCHI-KEY InChIKey=GPRLSGONYQIRFK-UHFFFAOYSA-N MOLECULAR-WEIGHT 1.0078
[...]

relations

RXN-13524 is_a_class Small-Molecule-Reactions
RXN-13524 is_a_class Chemical-Reactions
RXN-13524 consumes CPD-9424 COMPARTMENT UNKNOWN STOICHIOMETRY 1
RXN-13524 consumes Donor-H2 COMPARTMENT UNKNOWN STOICHIOMETRY 1
RXN-13524 consumes OXYGEN-MOLECULE COMPARTMENT UNKNOWN STOICHIOMETRY 1
RXN-13524 produces CPD-14516 COMPARTMENT UNKNOWN STOICHIOMETRY 1
RXN-13524 produces Acceptor COMPARTMENT UNKNOWN STOICHIOMETRY 1
RXN-13524 produces WATER COMPARTMENT UNKNOWN STOICHIOMETRY 1
RXN-13524 produces PROTON COMPARTMENT UNKNOWN STOICHIOMETRY 1
RXN-13524 is_in_pathway FWY-5748
[...]

```

PADMmet format : for studied species and reference data base

# The PADMet format



15

PADMet : PortABle Database for METabolism

## Policy

```

reaction has_namename
reaction has_xref xref
reaction has_suppData suppData
reaction is_in_pathway pathway
reaction consumes compound      STOICHIOMETRY  X  COMPARTMENT  Y
reaction produces compound      STOICHIOMETRY  X  COMPARTMENT  Y
reaction is_linked_to  geneASSIGNMENT X
...
  
```

+ class, compound, gene, pathway and protein

```

...
  
```

Better management of supplementary data  
More information and all-in-one solution

# The PADMet format



16

PADMet : PortABle Database for METabolism

Nodes

## Nodes

```
reaction 1-ACYLGLYCEROL-3-P-ACYLTRANSFER-RXN DIRECTION LEFT-TO-RIGHT
COMMON-NAME 1-acylglycerol-3-phosphate O-acyltransferase EC-NUMBER
EC-2.3.1.51
```

```
compound 1-AMINO-PROPAN-2-ONE-3-PHOSPHATE SMILES C(C(=O)C[N+])OP(=O)
([O-])[O-] COMMON-NAME 3-amino-1-hydroxyacetone 1-phosphate MOLECULAR-
WEIGHT 168.066
```

...

```
RXN-13524 is_in_pathway PWY-5748
[...]
```

Better management of supplementary data  
More information and all-in-one solution



# The PADMet format

PADM

17

## PADMet : PortABle Database for METabolism

```

codes

```

```

:reaction RXN-13524 DIRECTION LEFT-TO-RIGHT COMMON-NAME pseudo-conhydrine hydroxylase EC-NUMBER EC-1.14.99
:pathway FWY-5748 COMMON-NAME 4gamma;-coniciene and coniine biosynthesis
:compound CPD-14516 SMILES CCCC1(N=CC(O)CC1) COMMON-NAME 5-hydroxy-4gamma;-coniceine INCHI-KEY InChIKey=BVUNSAIHNEMOAG-UHFFFAOYSA-N MOLECULAR-WEIGHT
:compound CPD-9424 SMILES CCCC1(=[N+])CCCC1) COMMON-NAME 4gamma;-coniceine INCHI-KEY InChIKey=PQSOVRNZJIENN-UHFFFAOYSA-O MOLECULAR-WEIGHT 126.221
:compound CPD-14516 SMILES CCCC1(N=CC(O)CC1) COMMON-NAME 5-hydroxy-4gamma;-coniceine INCHI-KEY InChIKey=BVUNSAIHNEMOAG-UHFFFAOYSA-N MOLECULAR-WEIGHT 146.145

```

## Relations

1.1.1.262-RXN	consumes	NAD	COMPARTMENT	CCO	STOICHIOMETRY	1
1.1.1.262-RXN	produces	PROTON	COMPARTMENT	CCO	STOICHIOMETRY	1
...						

```

RXN-13524 produces CPD-14516 COMPARTMENT UNKNOWN STOICHIOMETRY 1
RXN-13524 produces Acceptor COMPARTMENT UNKNOWN STOICHIOMETRY 1
RXN-13524 produces WATER COMPARTMENT UNKNOWN STOICHIOMETRY 1
RXN-13524 produces PROTON COMPARTMENT UNKNOWN STOICHIOMETRY 1
RXN-13524 is_in_pathway FWY-5748
[...]

```

Better management of supplementary data  
More information and all-in-one solution

# The PADMet Toolbox



18

- Features
  - Import
  - Export + Connection to other tools
  - Data management
  - Analysis
  - Curation
  - Visualization (wiki, Cytoscape and pathways)
  - Menecheck : topological analysis of MN
  
- Adaptability

# The PADMmet Toolbox



19

## □ Portability (Python lib)

### Classe Tinygraphdb

---

#### Attributs

- policy (class)
- info (dict)
- dicOfNode (dict)
- dicOfRelationIn (dict)
- dicOfRelationOut (dict)

---

#### Fonctions

- 31 fonctions différentes
- 1609 lignes

### Modules:

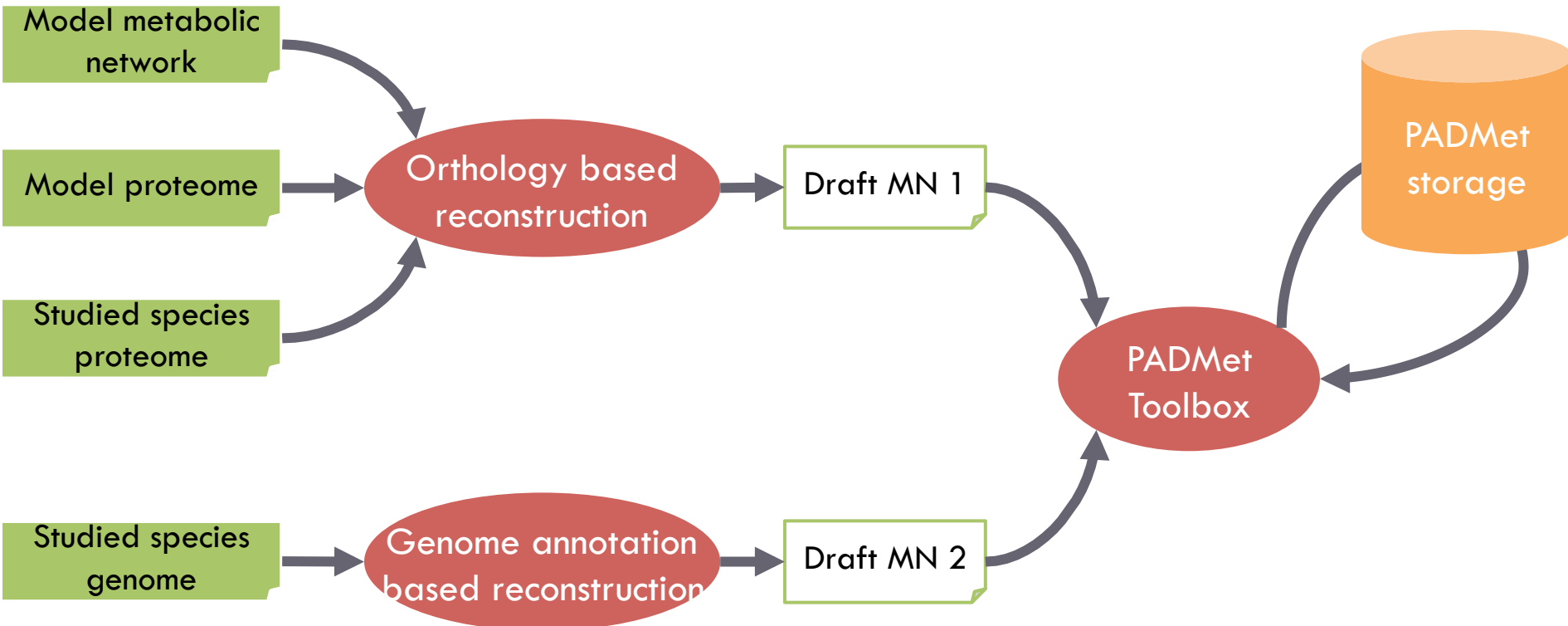
- **SBMLplugin**
  - 7 fcts / 177 lignes
- **SBMLgenerator**
  - 8 fcts / 565 lignes
- **ASPgenerator**
  - 3 fcts / 240 lignes
- **WIKIgenerator**
  - 18 fcts / 816 lignes
- **FBA\_workflow**
  - 7 fcts
- **metacycToPADMet**
  - 6 fcts / 572 lignes

# AuReMe workflow

(Automatic Reconstruction of Metabolic network)

20

## ① Building a draft

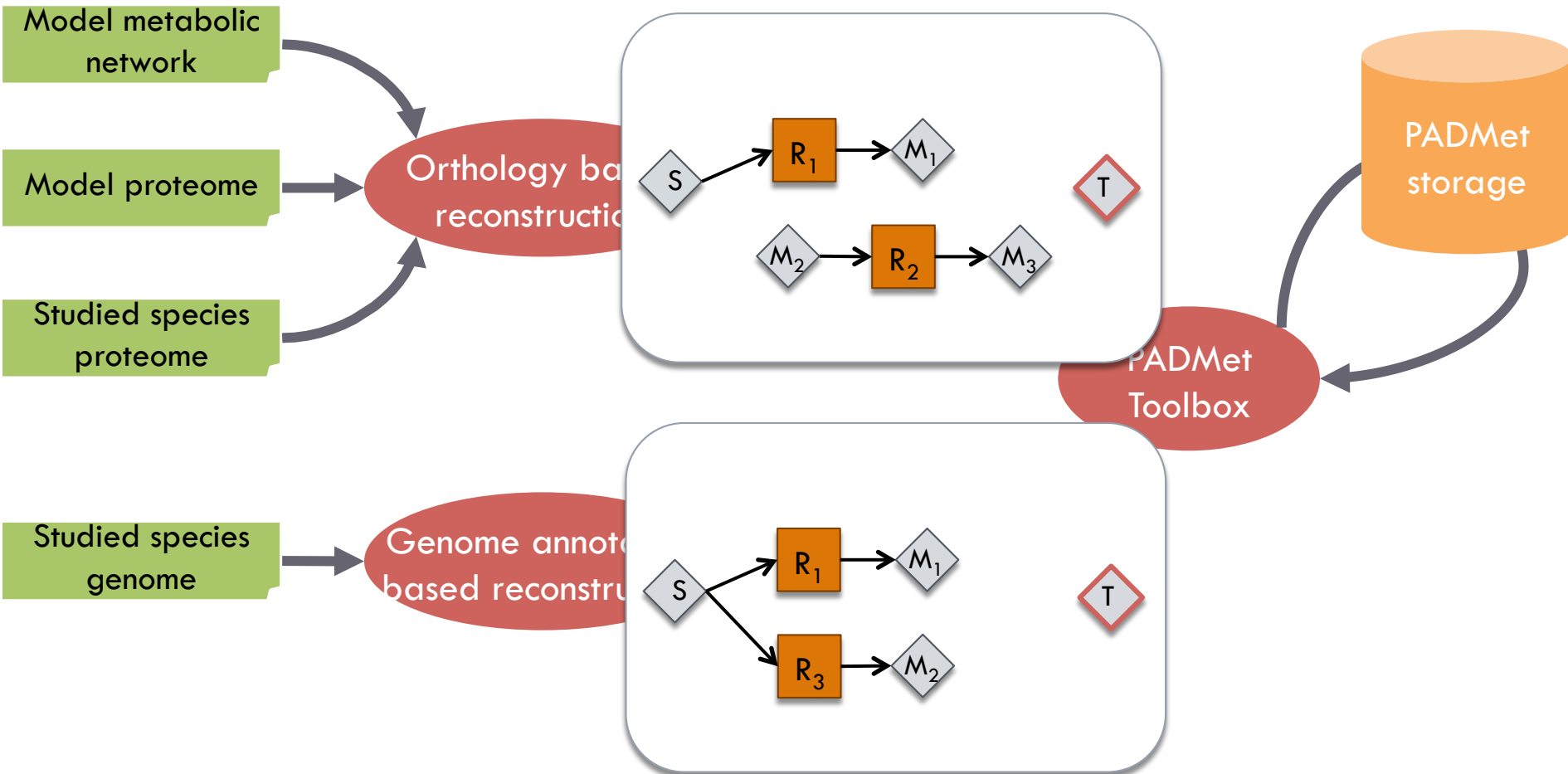


# AuReMe workflow

(Automatic Reconstruction of Metabolic network)

21

## ① Building a draft



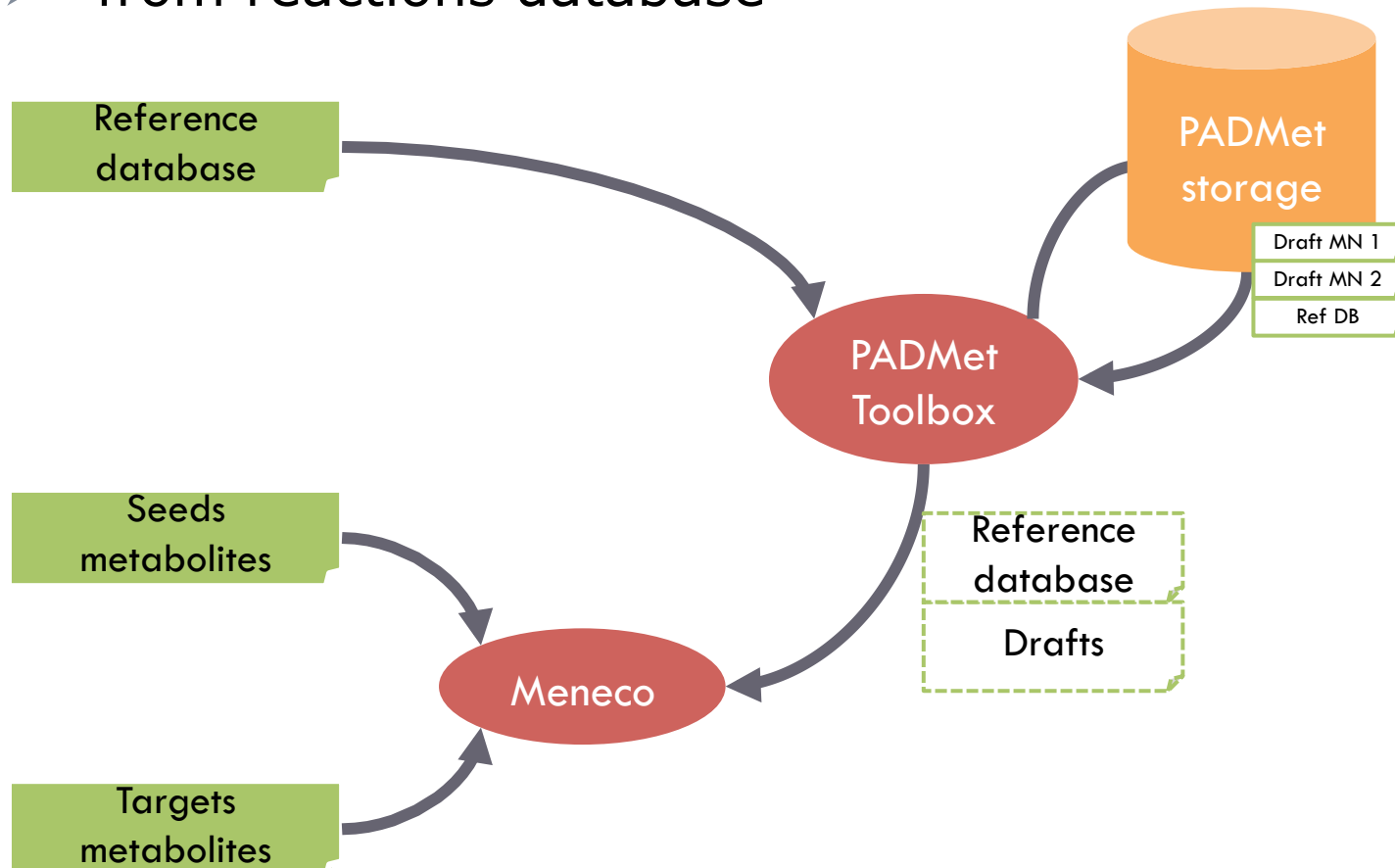
# AuReMe workflow

(Automatic Reconstruction of Metabolic network)

22

## ② Gap-filling the draft

➤ from reactions database



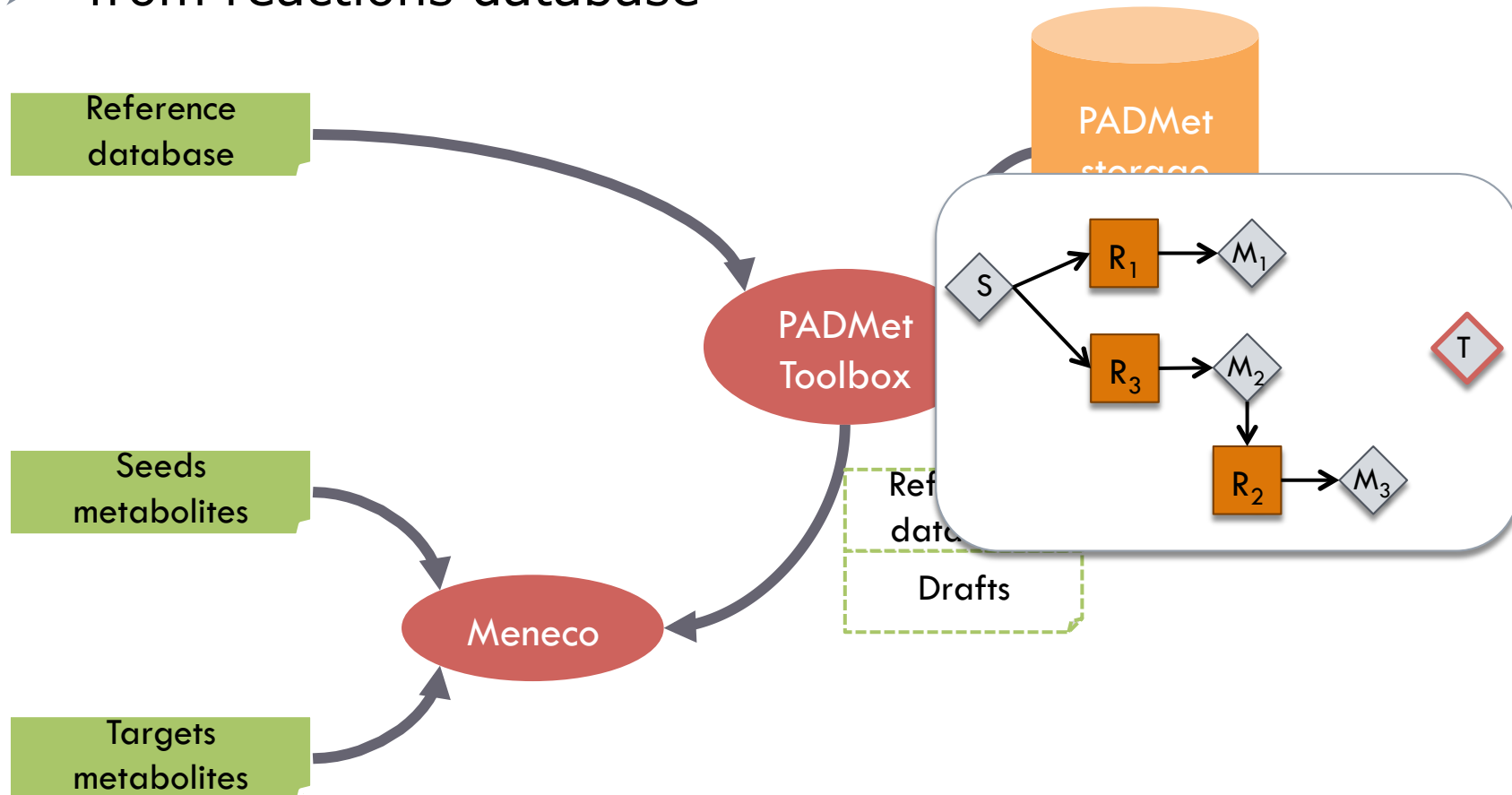
# AuReMe workflow

(Automatic Reconstruction of Metabolic network)

23

## ② Gap-filling the draft

➤ from reactions database



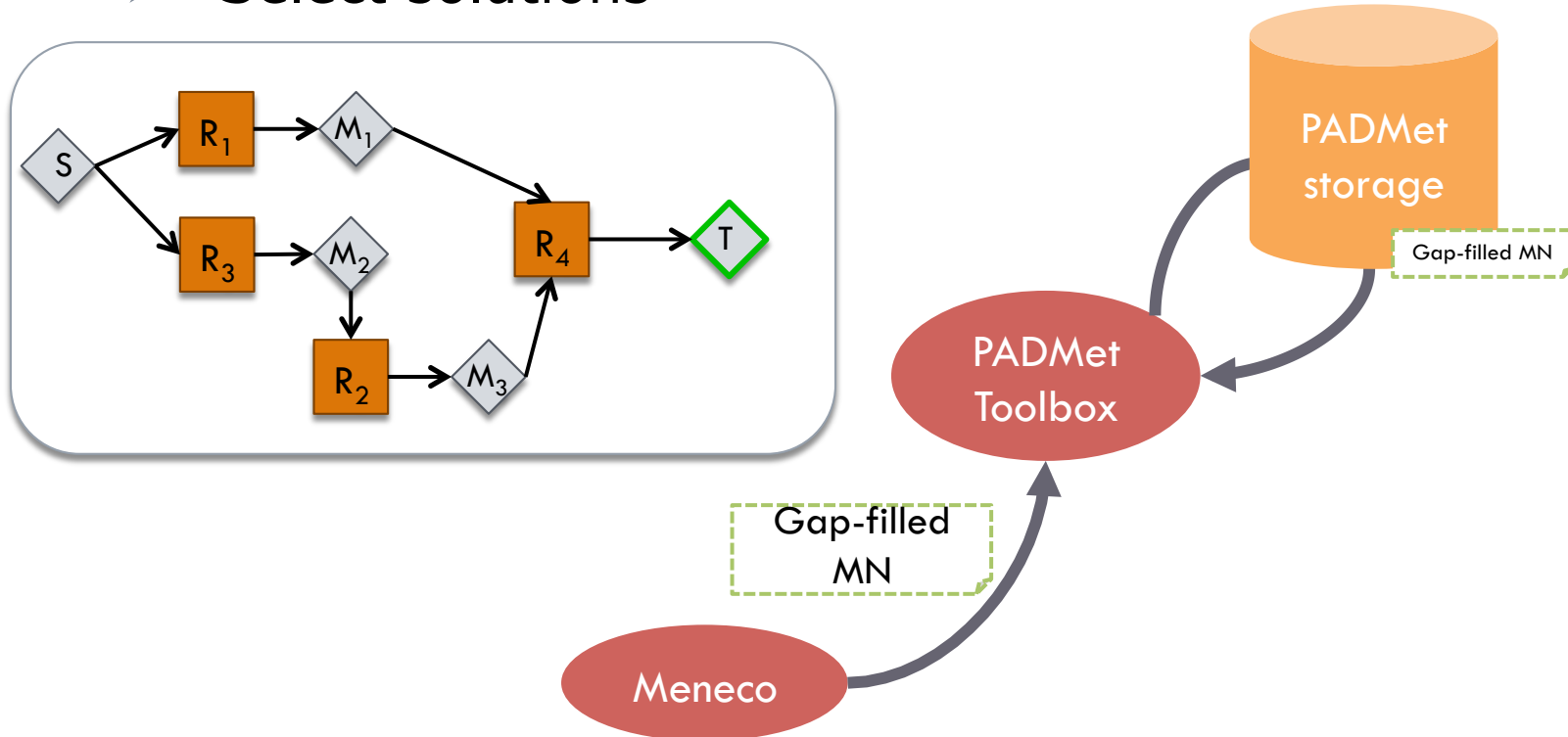
# AuReMe workflow

(Automatic Reconstruction of Metabolic network)

24

## ② Gap-filling the draft

➤ Select solutions



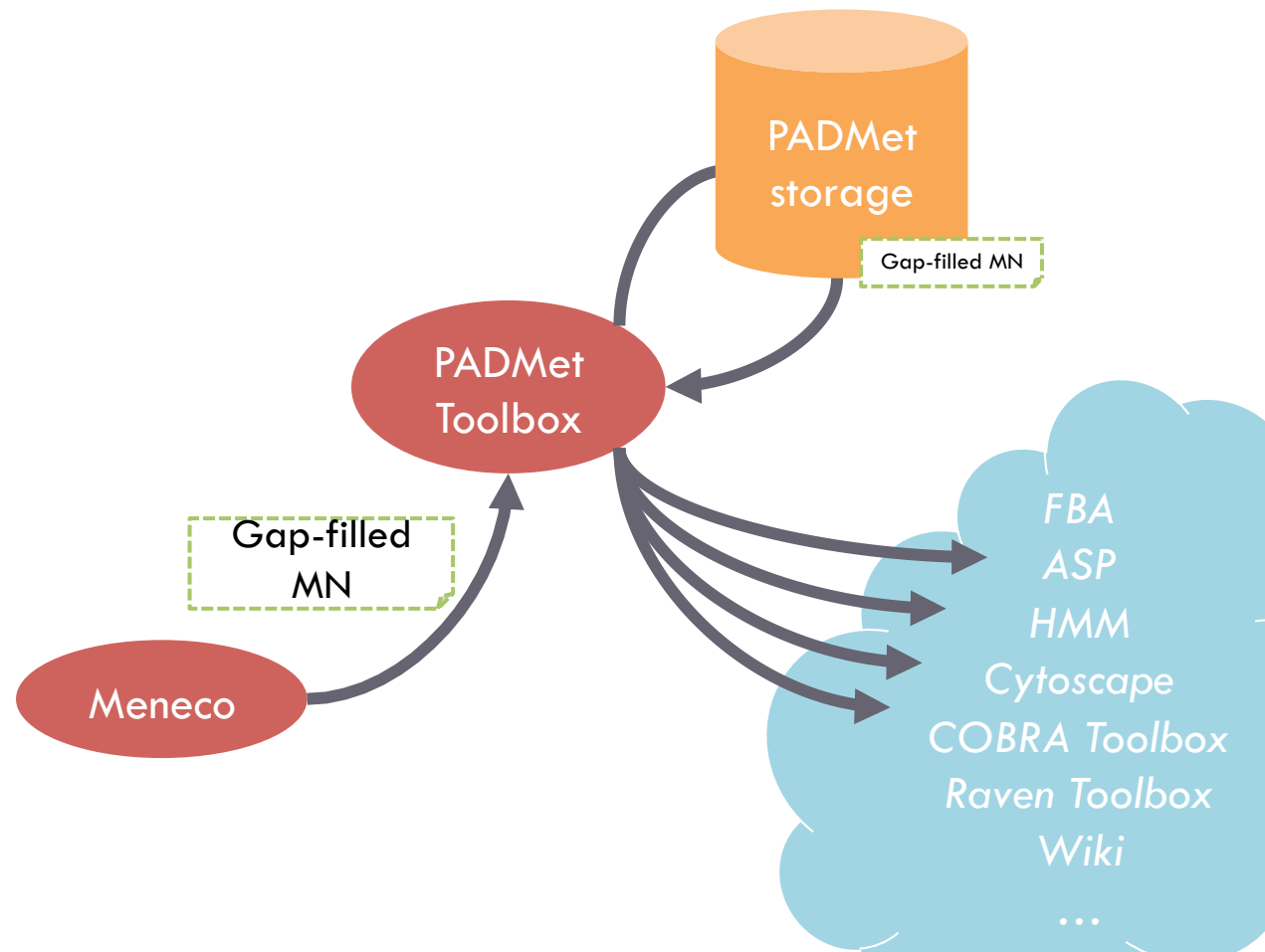


# AuReMe workflow

(Automatic Reconstruction of Metabolic network)

25

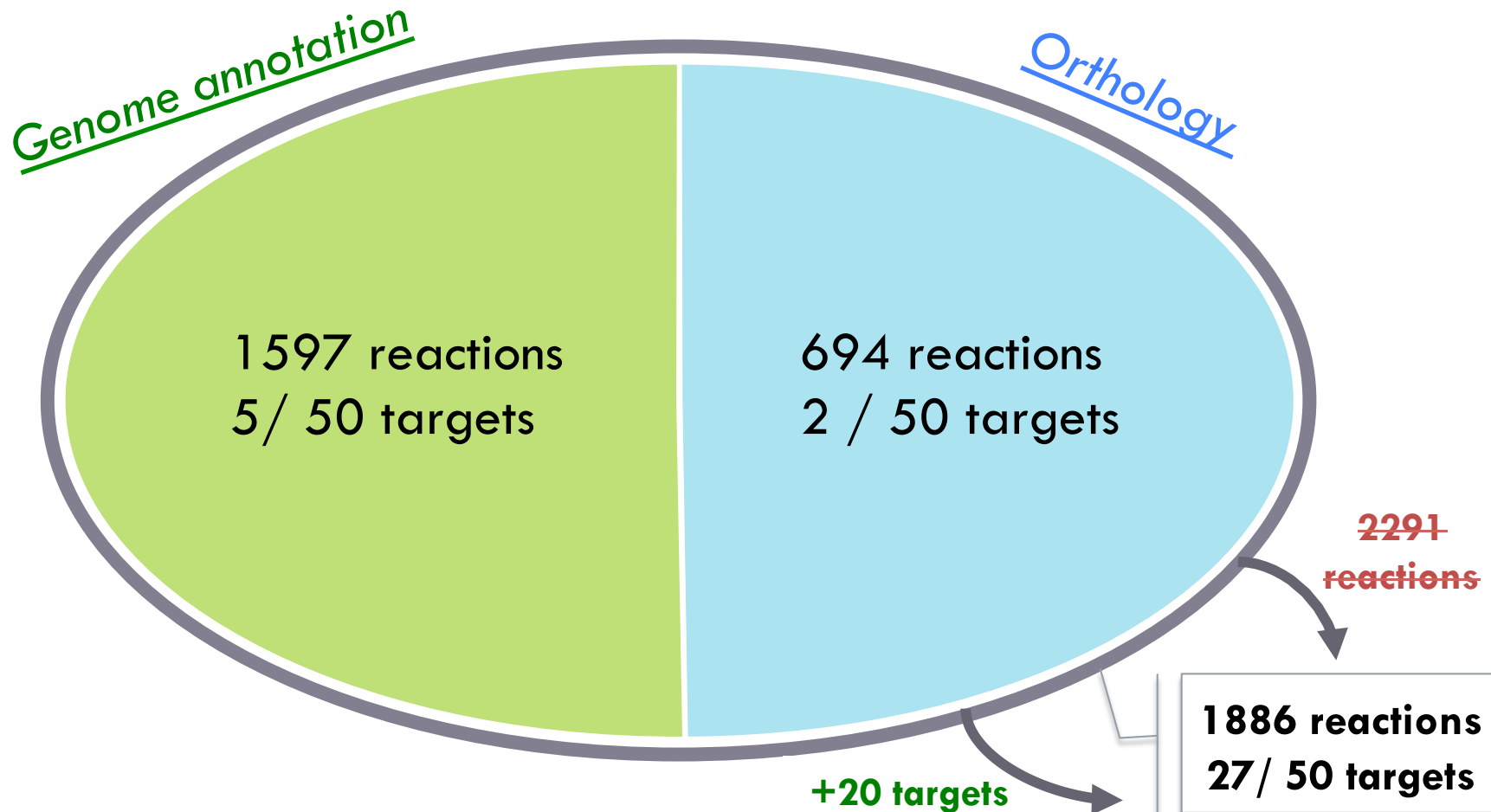
## □ Export and/or analyse results



# Complementary reconstructions

26

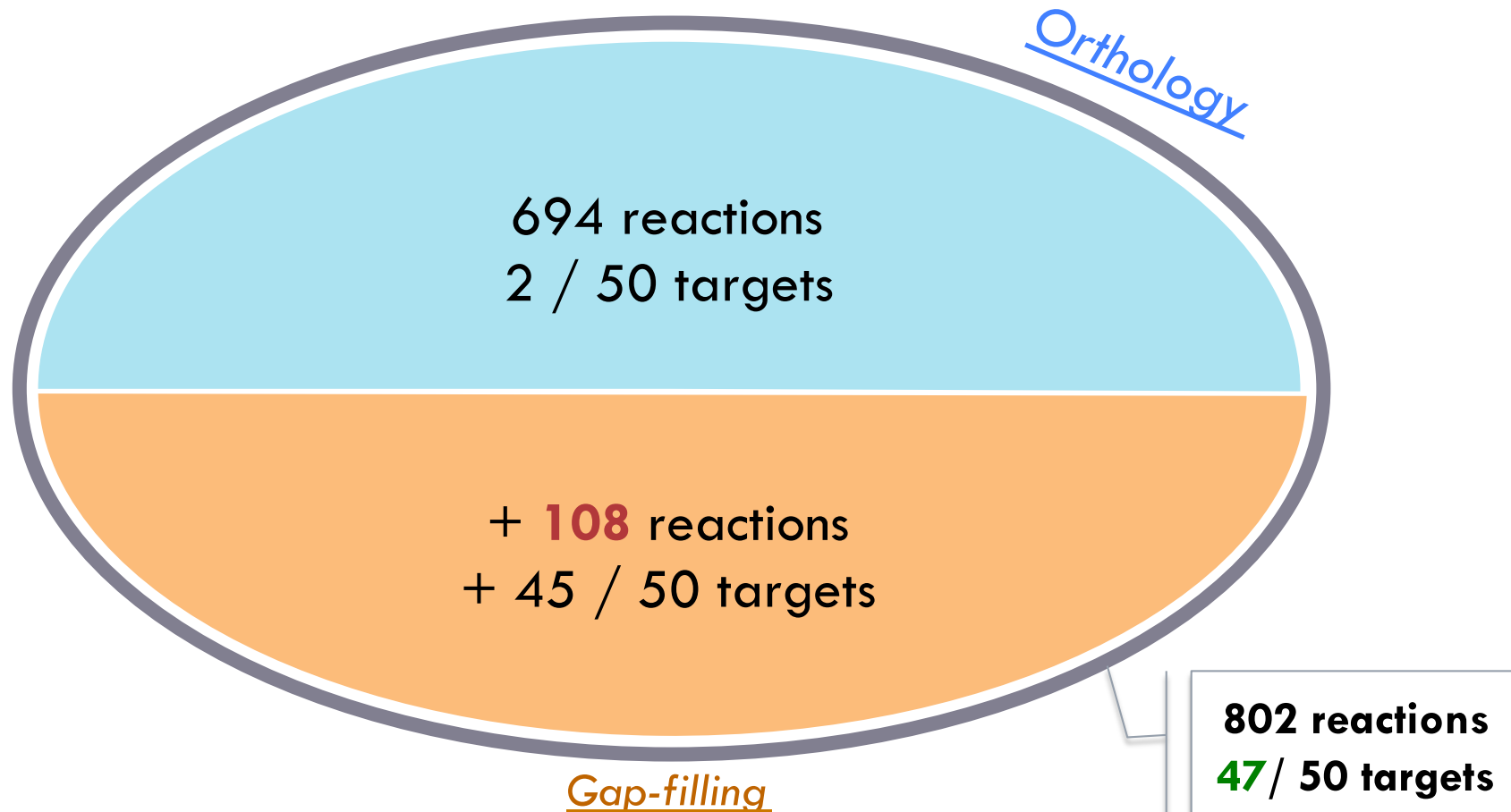
- Reconstruction on *Ectocarpus siliculosus* (brown algae)



# Complementary reconstructions

27

- Reconstruction on *Ectocarpus siliculosus* (brown algae)



# Complementary reconstructions

28

- Reconstruction on *Ectocarpus siliculosus* (brown algae)

Genome annotation

1597 reactions  
5 / 50 targets

+ 66 reactions  
+ 42 / 50 targets

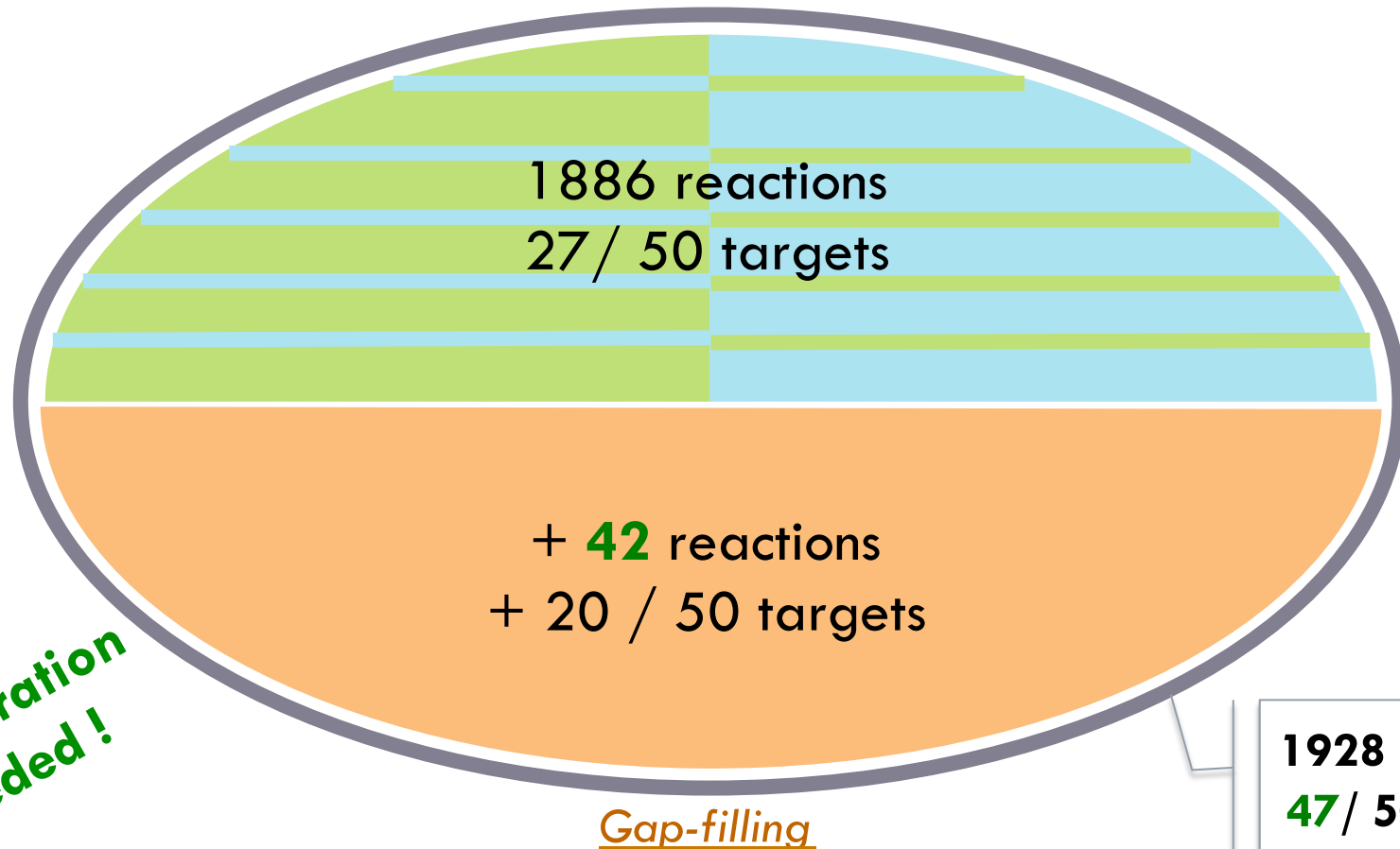
Gap-filling

**1663 reactions**  
**47 / 50 targets**

# Complementary reconstructions

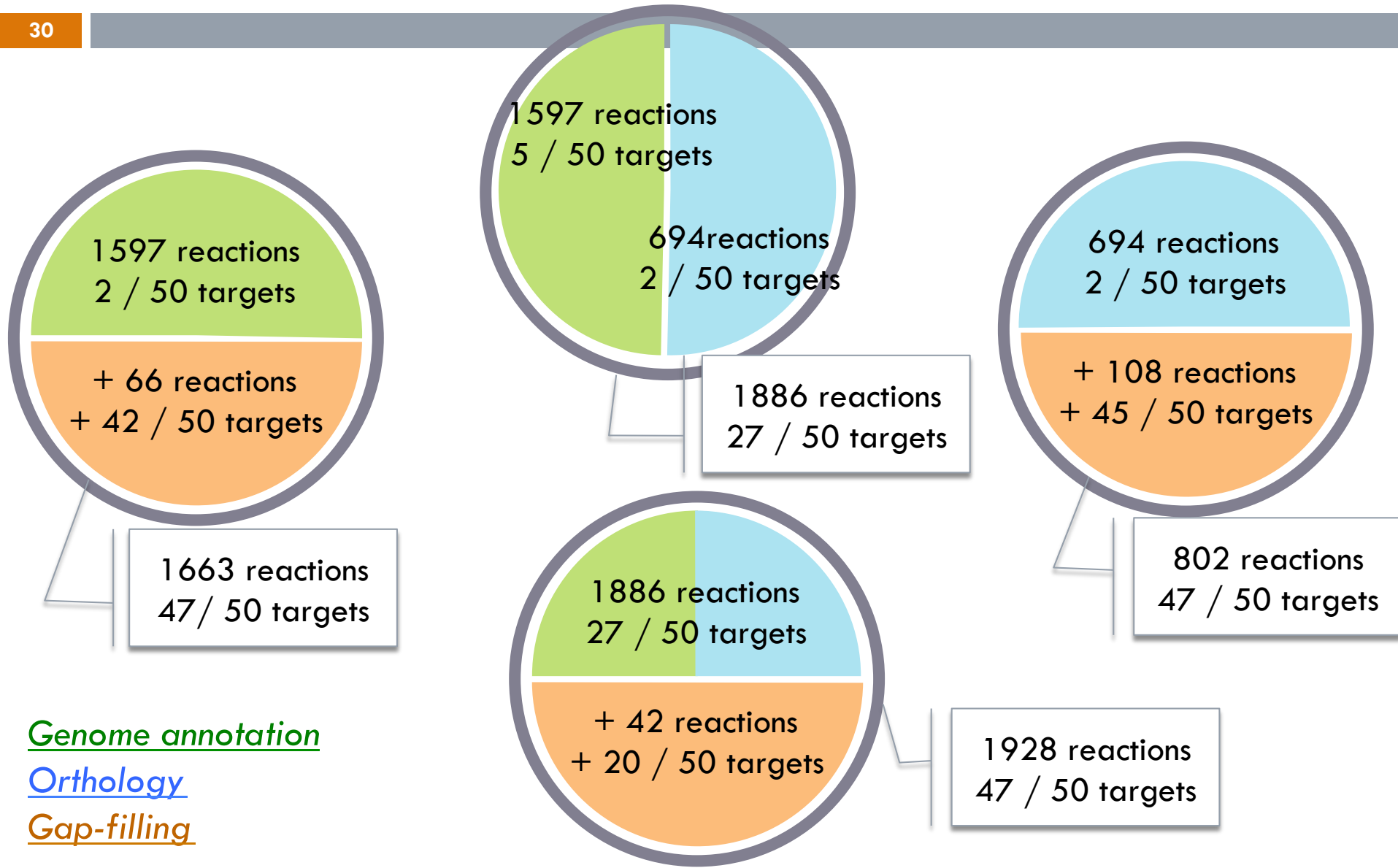
29

- Reconstruction on *Ectocarpus siliculosus* (brown algae)  
Genome annotation + Orthology



# Complementary reconstructions

30



# Easy to use solution

31

- Easy to share with Docker
- Easy to play with Galaxy
  - ▣ Coming soon...
  
- PADMet
  - ▣ Standalone
  - ▣ Toolbox : Python library
  - ▣ Format : extendable
  - ▣ Customizable



# Easily explorable results

32

- Connection to several other tools :



- Reports for analysis and curation



# Conclusion : AuReMe & PADMet

33

- Total reconstruction of metabolic networks
- High cross-information of different data sources
  - ▣ Very usefull for exotic species
- Highly manipulable
- Very easy to use and analyse
- Portable in docker containers

DyLISS

*Inria*  
INVENTEURS DU MONDE NUMÉRIQUE

UMR IRISA



**CMM**  
Center for  
Mathematical  
Modeling  
Advanced Math for Business



Thank you for your attention

Clémence Frioux, Meziane Aite, Jeanne Got, Anne Siegel, Julie Laniau, Guillaume Collet  
and all the Dyliss team

Camille Trottier, Jérémie Bourdon and Damien Eveillard

Nicolas Loira, Maria-Paz Cortes, Philippe Bordron, Alejandro Maas  
and all the Chilean team from the CMM

# Metabolic network : 2 models

35

- Topological



- Producibility =  
**paths** in the graph

- Stoichiometric

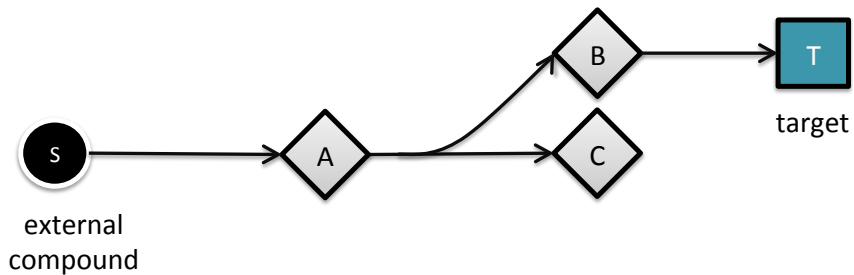


- Producibility =  
**flux** in pipes

# Metabolic network : 2 models

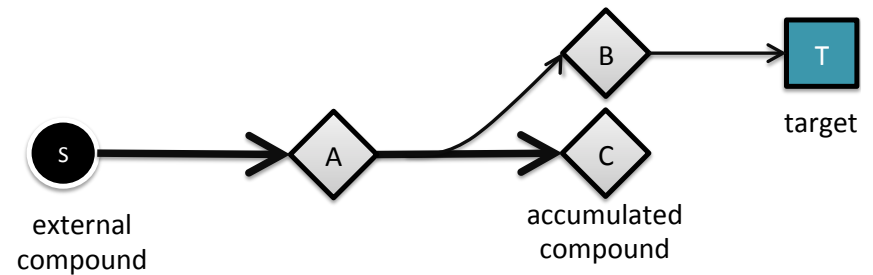
36

## □ Topological



✓ Target is produced

## □ Stoichiometric



✗ Target is NOT produced

# PADMet : reports

37

## □ On metabolites :

dbRef_id	Common name	Produced (p), Consumed (c), Both (cp)
CPD-13043	7-carboxy-7-deazaguanine	CP

**a**

## □ On reactions :

dbRef_id	Common name	formula (with id)	formula (with common name)	in pathways	associated genes
ACETOOHBUTSYN-RXN	2-aceto-2-hydroxy-butyrate synthase	1 PROTON + 1 2-OXOBUTANOATE + 1 PYRUVATE => 1 CARBON-DIOXIDE + 1 2-ACETO-2-HYDROXY-BUTYRATE	1 H+ + (...)	ILEUSYN-PWY; PWY-5104;...	brl001f_2818; brl001f_2819

**b**

## □ On pathways :

dbRef_id	Common name	Number of reaction found	Total number of reaction	Ratio (Reaction found / Total)
P185-PWY	formaldehyde assimilation III (dihydroxyacetone cycle)	10	12	0.83

**c**

# PADMmet : report about

38

(Metabolites / Reactions / Pathways)

- ID
- Common name
- If produced / Consumed only or both (based on reaction reversibility)



dbRef_id	Common name	Produced (p), Consumed (c), Both (cp)
CPD-13043	7-carboxy-7-deazaguanine	CP

# PADMet : report about

39



(Metabolites / Reactions / Pathways)

- ID
- Common name
- Formula (with compounds ids or common name)
- Pathways involved in
- Associated genes

dbRef_id	Common name	formula (with id)	formula (with common name)	in pathways	associated genes
ACETOOHBUTSY N-RXN	2-aceto-2-hydroxy-butyrate synthase	1 PROTON + 1 2-OXOBUTANOATE + 1 PYRUVATE => 1 CARBON-DIOXIDE + 1 2-ACETO-2- HYDROXY-BUTYRATE	1 H+ + (...)	ILEUSYN- PWY; PWY-5104; ...	brl001f_281 8; brl001f_281 9

# PADMet : report about

40

(Metabolites / Reactions / Pathways)



- ID
- Common name
- Number of reaction present in the network (associated to this pathway)
- Total number of reaction (associated to this pathway)
- Completion ratio

dbRef_id	Common name	Number of reaction found	Total number of reaction	Ratio (Reaction found / Total)
P185-PWY	formaldehyde assimilation III (dihydroxyacetone cycle)	10	12	0.83



# PADMet : wiki

41

Set \$wgLogo to the URL path to your own logo image.

Page [Discussion](#)

Read

[View source](#)

[View history](#)

Search

## Main Page

**MediaWiki has been successfully installed.**

Consult the [User's Guide](#) for information on using the wiki software.

## Getting started

- [Configuration settings list](#)
- [MediaWiki FAQ](#)
- [MediaWiki release mailing list](#)
- [Localise MediaWiki for your language](#)

This page was last modified on 16 November 2015, at 15:33.

Content is available under [Creative Commons Attribution](#) unless otherwise noted.

[Privacy policy](#) [About synechococcusgem.a15-62](#) [Disclaimers](#)

[Main page](#)  
[Semantic search](#)

Categories

[Reactions](#)

[Genes](#)

[Metabolites](#)

[Pathways](#)

Origin of reactions

[EC-number](#)

[Gene Ontology term](#)

[Functional](#)

[Description](#)

[Gap-Filling \(Pathway Tools\)](#)

[Gap-Filling \(Functional\)](#)

[Manual](#)

Tools

[What links here](#)

[Related changes](#)

[Special pages](#)

[Printable version](#)

[Permanent link](#)

[Page information](#)

[Browse properties](#)

