Towards the structural screening of microbioal ecosystems : selected projects and MetaFoldScan presentation

Gwenaëlle André-Leroux

To cite this version:

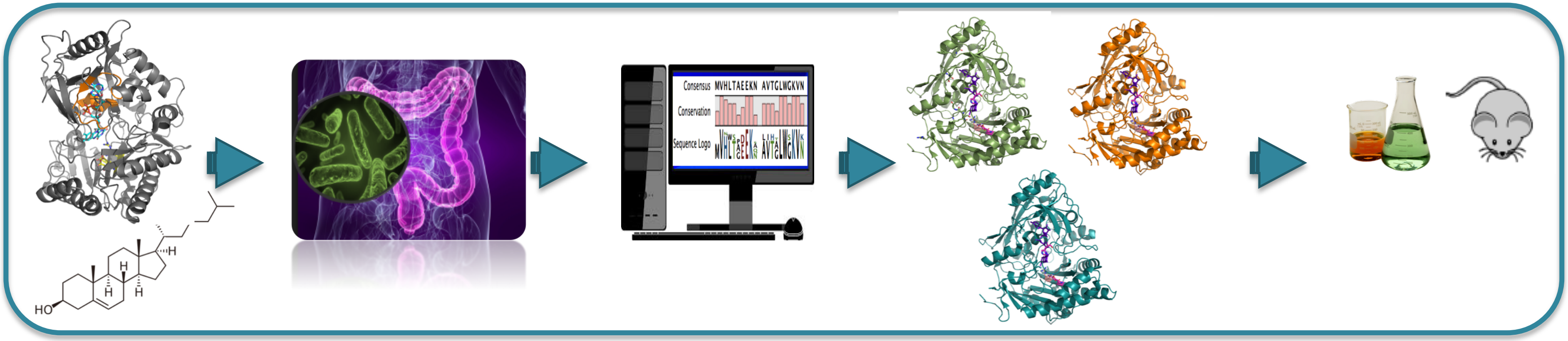
Gwenaëlle André-Leroux. Towards the structural screening of microbioal ecosystems : selected projects and MetaFoldScan presentation. Séminaire, Dec 2018, Montevideo, Uruguay. hal-02785774

HAL Id: hal-02785774
https://hal.inrae.fr/hal-02785774

Submitted on 4 Jun 2020

Towards the Structural Screening of Microbial Ecosystems

Selected projets and MetaFoldScan presentation



Gwenaëlle André-Leroux

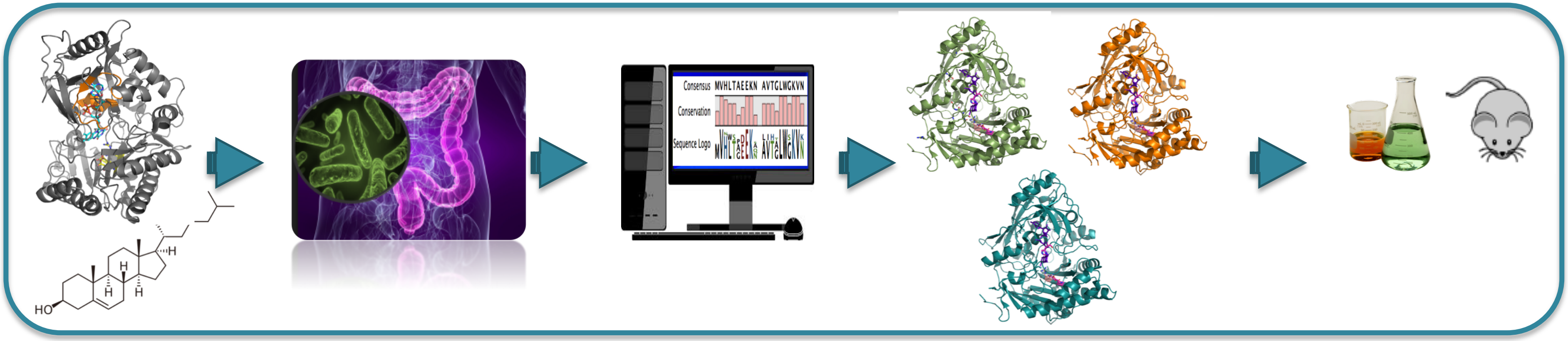
Montevideo University of Sciences December 2018

INRA
SCIENCE & IMPACT

MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES
MaIAGE
DU GÉNOME À L'ENVIRONNEMENT

Towards the Structural Screening of Microbial Ecosystems

Selected projets and MetaFoldScan presentation

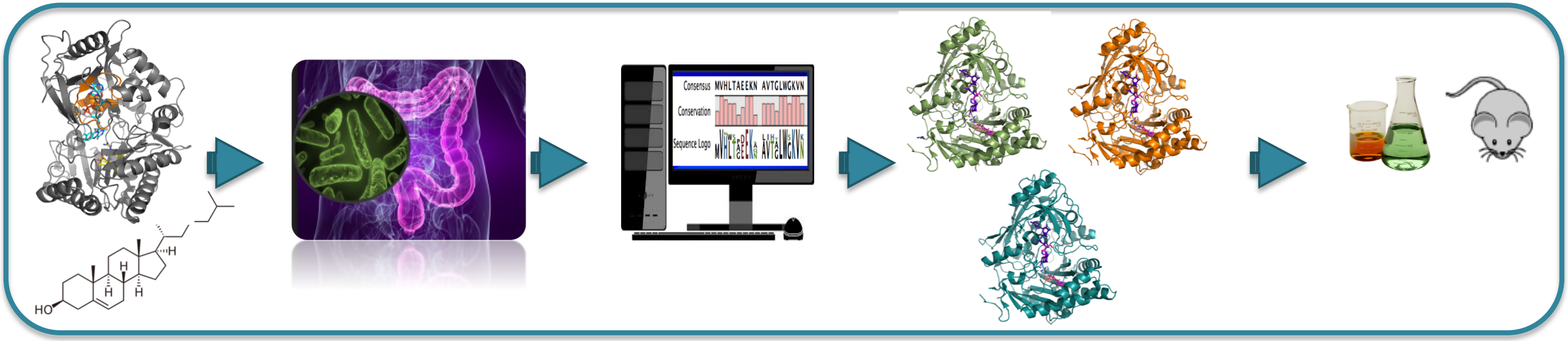


Outline

- Cholesterol conversion in the gut microbiota: the mystery enzyme(s) ?
- MetaFoldScan project : 3D screening of the gut microbiota
- Conclusions and perspectives

Towards the Structural Screening of Microbial Ecosystems

Selected projets and MetaFoldScan presentation



Outline

- Cholesterol conversion in the gut microbiota: the mystery enzyme(s) ?

- MetaFoldScan project : 3D screening of the gut microbiota

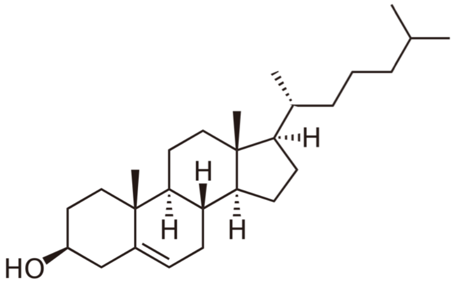
- Conclusions and perspectives

Metabolism of cholesterol by the gut microbiome

✪ Cholesterol :

- Is a sterol that plays a central role as cell-membrane agent.
- Is precursor of steroïd hormone (progesterone, testosterone, cortisone) and bile salt.
- Originates 30% from diet vs 70 % from bile and desquamed sterol of the gut epithelium.

✪ Cholesterol level and gut microbiota :



- Can be metabolized by colonic bacteria.
- The gut microbiota reduces cholesterol to coprostanol.
- Neomycin impacts serum cholesterol and fecal sterol in hypercholesterolemic patients.

✪ Cholesterol is absorbed by human intestin, not coprostanol :

- Cholesterol conversion relates to abundance of bacteria with cholesterol reducing activities.

  Gérard et al, 2004, Veiga et al, 2005
- Cholesterol conversion to coprostanol could result in lower cholesterolemia.

Two cholesterol-reducing bacterial strains have been isolated

✪ One strain from hog-sewage lagoon:

Gram+ coccobacillus *Eubacterium coprostanoligenes* ATCC 51222T Freier *et al*, 1993

Oral administration of *E. copro* → significant decrease of plasma cholesterol in dietary induced hypercholesterolemic rabbits Li *et al*, 1995

✪ One from « high converter » human:

Isolate from the *Bacteroides phylum, Bacteroides sp. Strain D8.* Gram- anaerobie. Gérard *et al*, 2007

Phylogenetic tree construction → this strain clusters in an independant clade with the two isolates of *Bacteroides dorei* species.

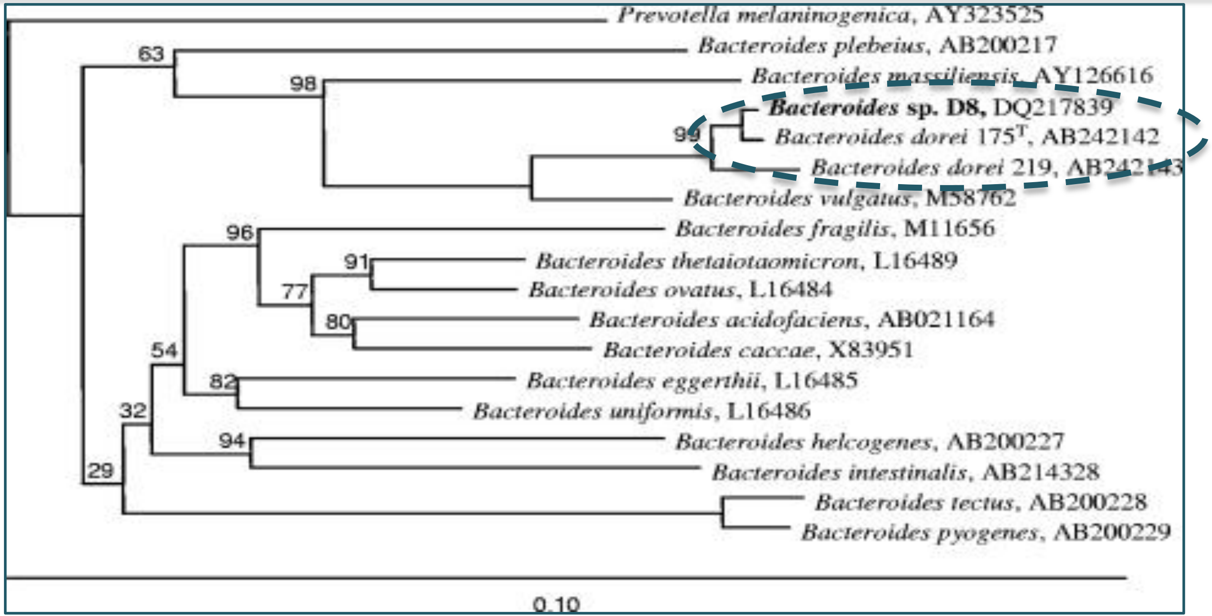
Intestinal microbiota

>1,000 genera of bacteria.

2 phyla

Bacteroidetes

Firmicutes



Two distinct cholesterol-reducing bacterial strains have been isolated

Phylogenetically close

Phylogenetically remote

Strain \ Medium	SBM	BCM
B. Dorei 175^T	-	-
B. Dorei D8	+	-
E. coprostanoligenes	-	+

SBM : Standard Brain Medium
BCM: Basal Cholesterol Medium

Coll. C. Juste and P. Gérard, Micalis

Metabolic pathway for cholesterol to coprostanol conversion in *B. dorei D8* ?

Which are the enzymes of the catabolic pathway that degrade cholesterol into coprostanol?

Two cholesterol-reducing bacterial strains have been isolated

✪ One strain from hog-sewage lagoon:

Gram+ coccobacillus *Eubacterium coprostanoligenes* ATCC 51222T Freier *et al*, 1993

Oral administration of *E. copro* → significant decrease of plasma cholesterol in dietary induced hypercholesterolemic rabbits Li *et al*, 1995

✪ One from « high converter » human:

Isolate from the *Bacteroides phylum, Bacteroides sp. Strain D8.* Gram- anaerobie. Gérard *et al*, 2007

Phylogenetic tree construction → this strain clusters in an independant clade with the two isolates of *Bacteroides dorei* species.

Microbial genes or enzymes involved in cholesterol metabolism in the gut are unknown.

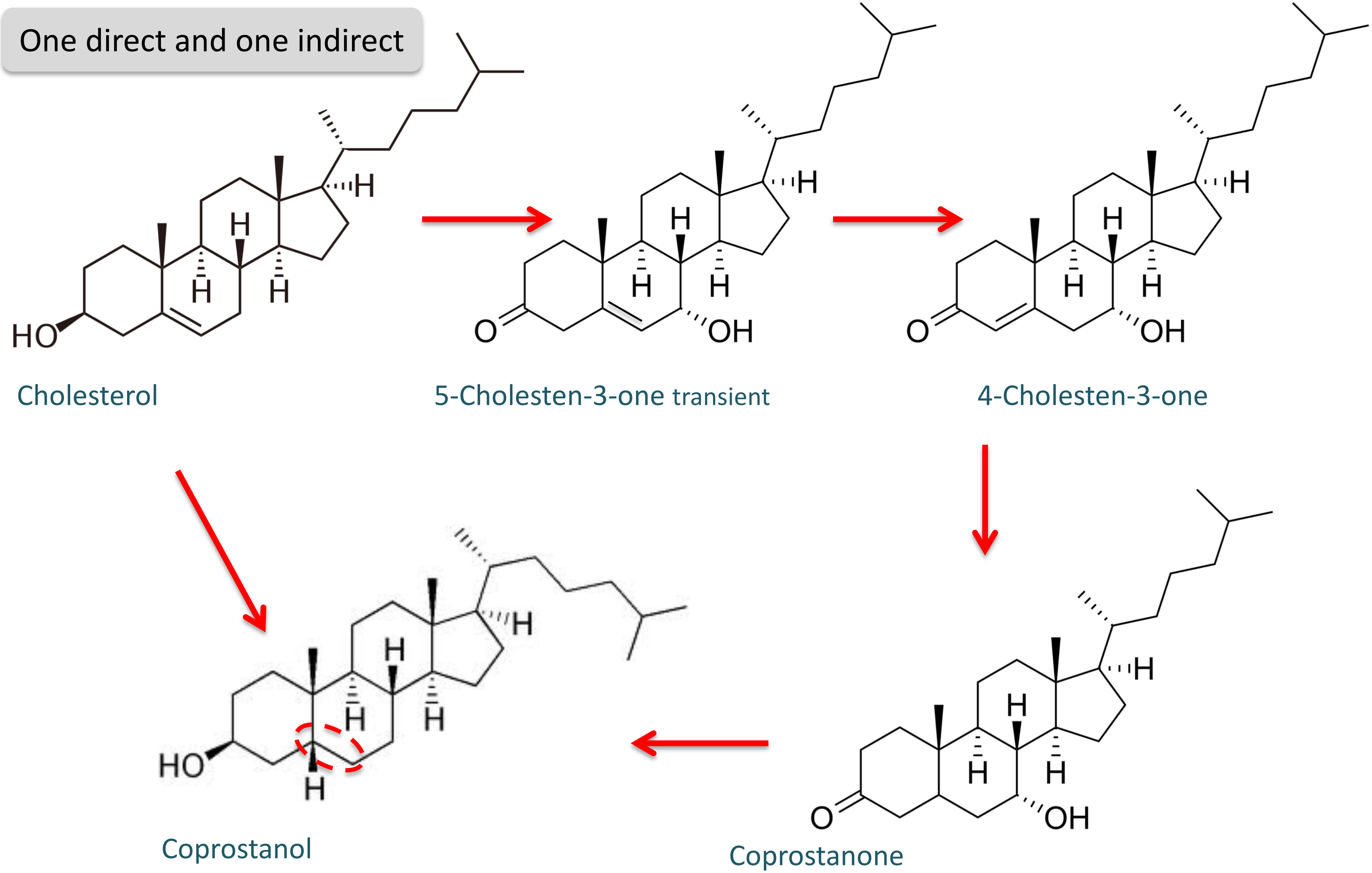
However …

Genomes of *B. Dorei* D8 & *B. Dorei* 175^T sequenced and assembled at MaIAGE

Valentin Loux

Cholesterol metabolic pathway in aerobie bacteria

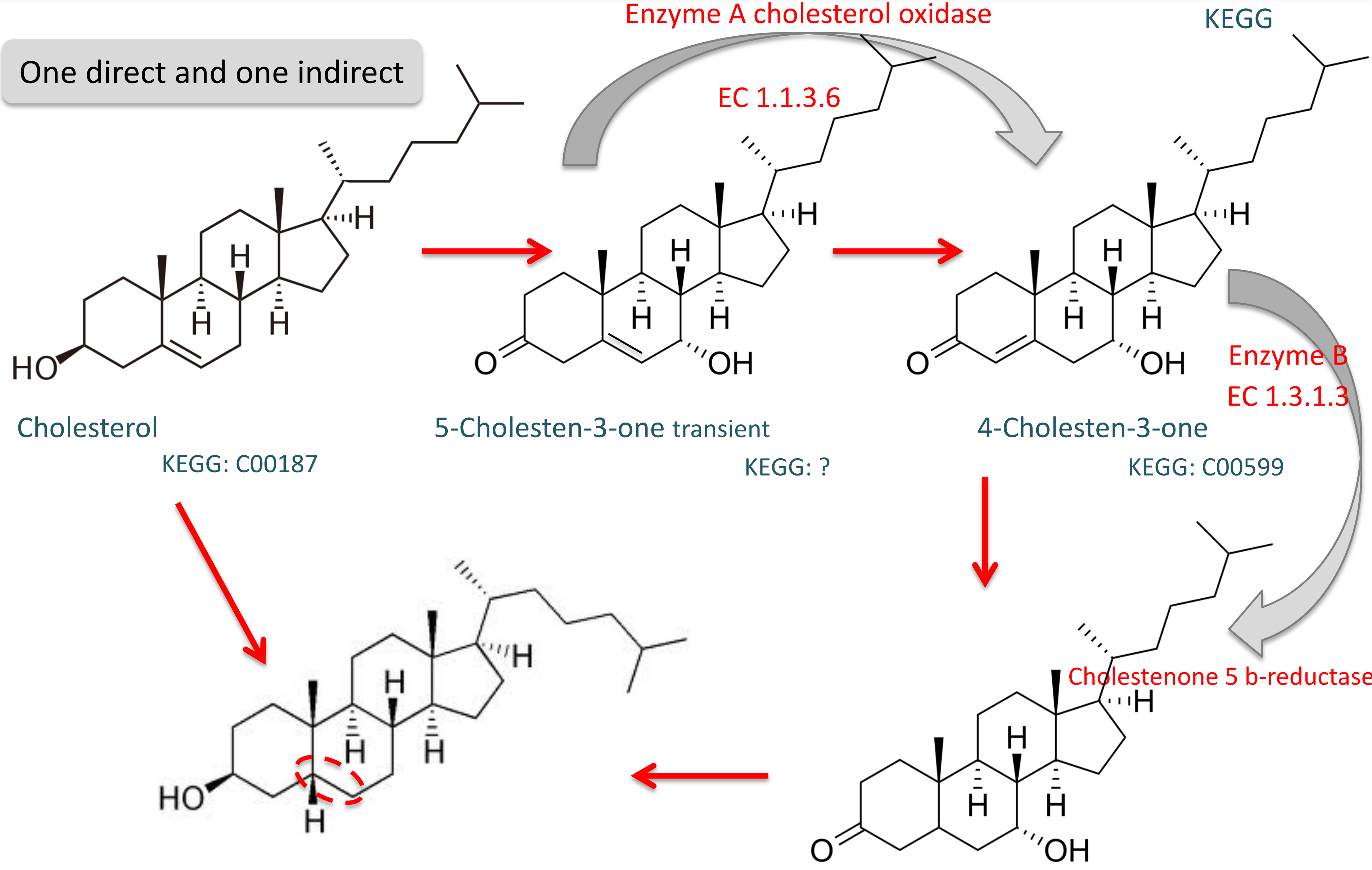
One direct and one indirect



KEGG

Cholesterol metabolic pathway in aerobie bacteria

One direct and one indirect



KEGG

Enzyme A cholesterol oxidase

EC 1.1.3.6

Cholesterol

KEGG: C00187

5-Cholesten-3-one transient

KEGG: ?

4-Cholesten-3-one

KEGG: C00599

Enzyme B

EC 1.3.1.3

Cholestenone 5 b-reductase

Coprostanol

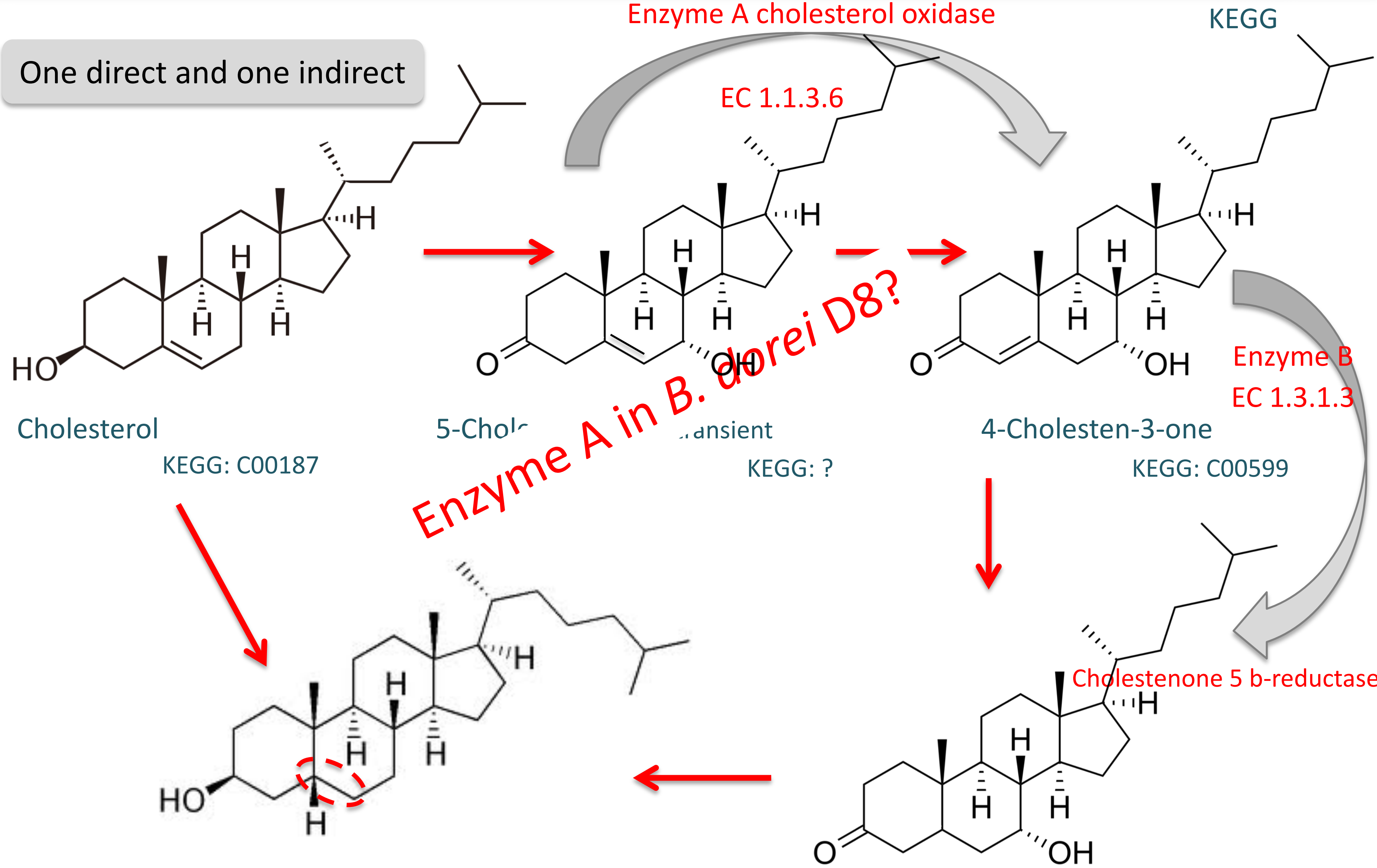
KEGG: ?

Coprostanone

KEGG:C03091

Cholesterol metabolic pathway in aerobie bacteria

One direct and one indirect



Enzyme A cholesterol oxidase

EC 1.1.3.6

KEGG

Cholesterol

KEGG: C00187

5-Chol ansient

KEGG: ?

4-Cholesten-3-one

KEGG: C00599

Enzyme B

EC 1.3.1.3

Enzyme A in B. dorei D8?

Cholestenone 5 b-reductase

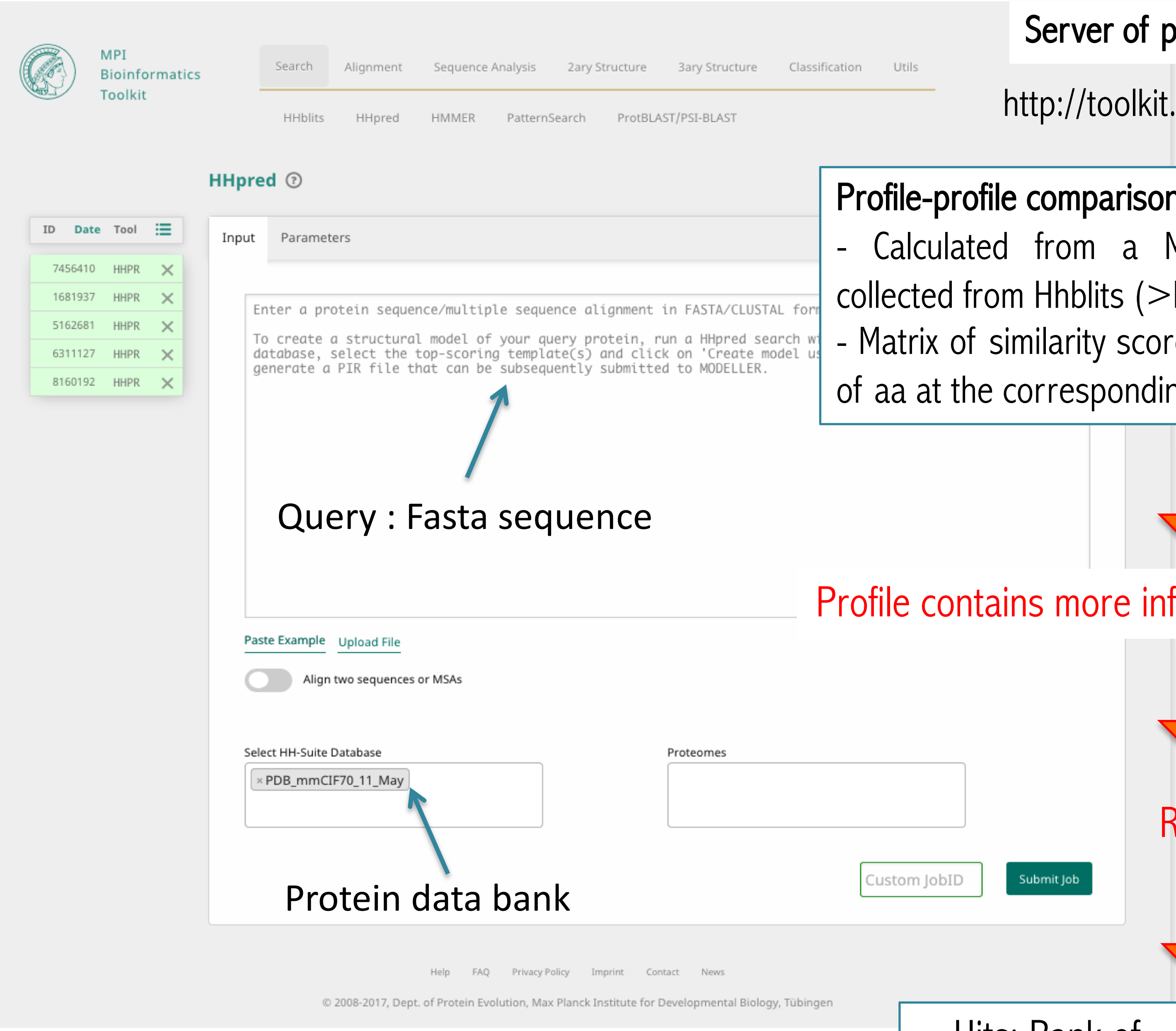
Coprostanone

KEGG:C03091

Coprostanol

KEGG: ?

HHpred homology Detection & Structure Prediction



ID	Date	Tool
7456410		HHPR
1681937		HHPR
5162681		HHPR
6311127		HHPR
8160192		HHPR

Enter a protein sequence/multiple sequence alignment in FASTA/CLUSTAL for

To create a structural model of your query protein, run a HHpred search wi database, select the top-scoring template(s) and click on 'Create model us generate a PIR file that can be subsequently submitted to MODELLER.

Select HH-Suite Database: PDB_mmCIF70_11_May

© 2008-2017, Dept. of Protein Evolution, Max Planck Institute for Developmental Biology, Tübingen

Server of protein structure prediction

http://toolkit.tuebingen.mpg.de/hhpred

Profile-profile comparison tools:

- Calculated from a MSA of related sequences collected from Hhblits (>Psi-blast+).
- Matrix of similarity scores calculated from frequency of aa at the corresponding positions in the MSA.

Query : Fasta sequence

Profile contains more information than a sequence.

Protein data bank

Result

Hits: Rank of « matching « pdb proteins

J. Söding, NAR 2005

Inverse strategy: setting up of standalone commands

Véronique Martin

- **Prediction of MSA profile with SS for each protein of *B. Dorei* 175^T & *B. dorei D8* genomes**
- **Compare each protein with a bank of cholesterol oxidase « Enzyme A » to be assessed**

1- Split of each fasta sequence in the genome -> more than 3,400 for each genome of *B. dorei*

2- Multiple sequence alignment for each fasta: 'HHblits' > PSI-Blast

3- Prediction & addition of SS elements: 'adssl.pl'

4- Profiling & comparaison with the 3D template « bank »:'HHBlitsdb.pl' & 'HHsearch'

5- Extraction, ranking and analyse of scores.

6- Analyse of selected sequences of proteins. Annotation? Fonction?

7- *In fine*, a list of cholesterol oxidase putative proteins to test *in vitro* / *in vivo*

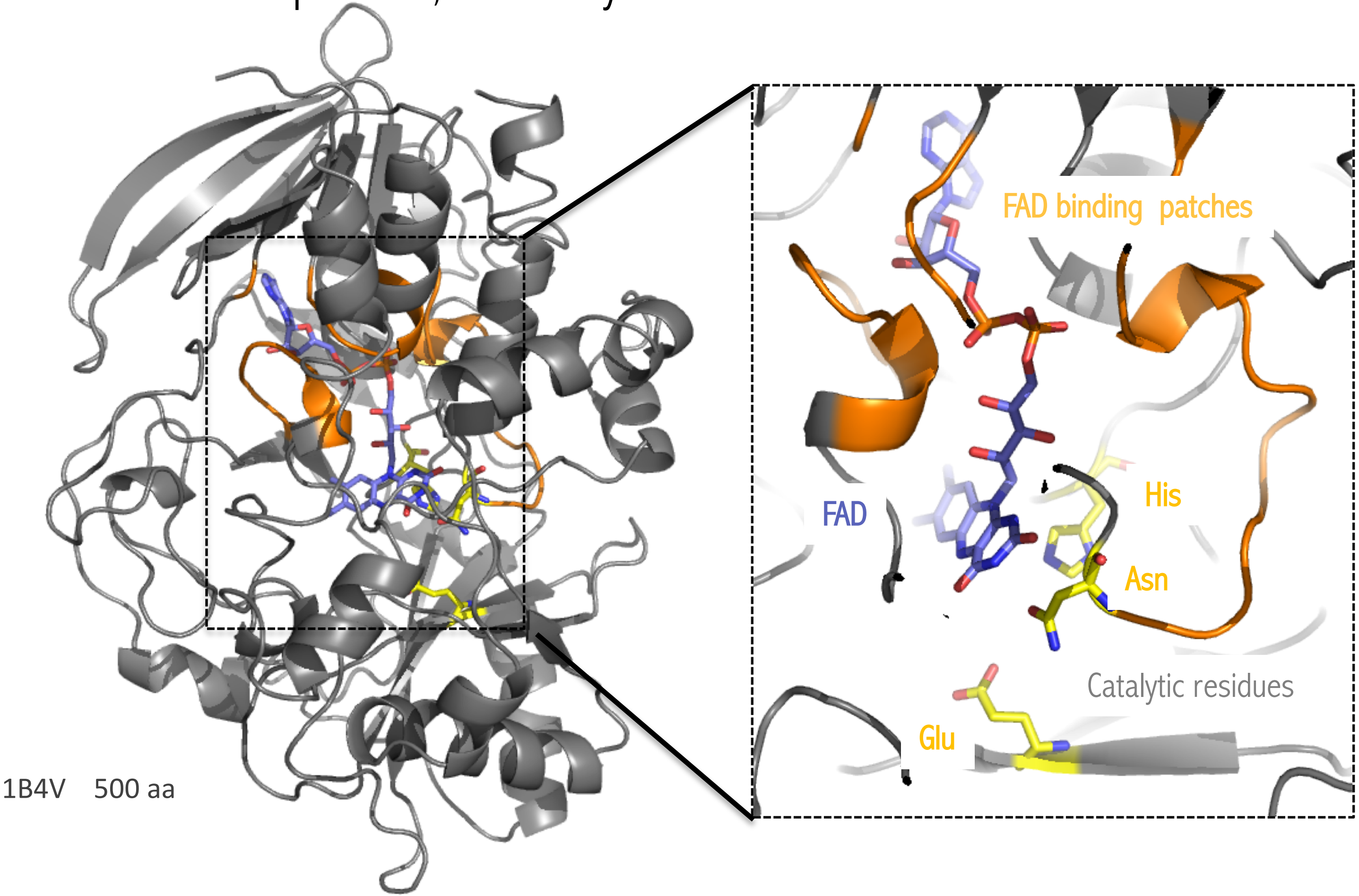
→ **The 3D template « bank » is critical**

Choosing the bank of cholesterol oxidases – Bank 1

Cholesterol oxidase:

- monomeric flavoenzyme catalyzes oxidation and isomerization into cholest-4-en-3-one.
- two folds exist, cofactor-FAD- dependent, covalently bound or not.

Catalytic site
Asn, Glu, His

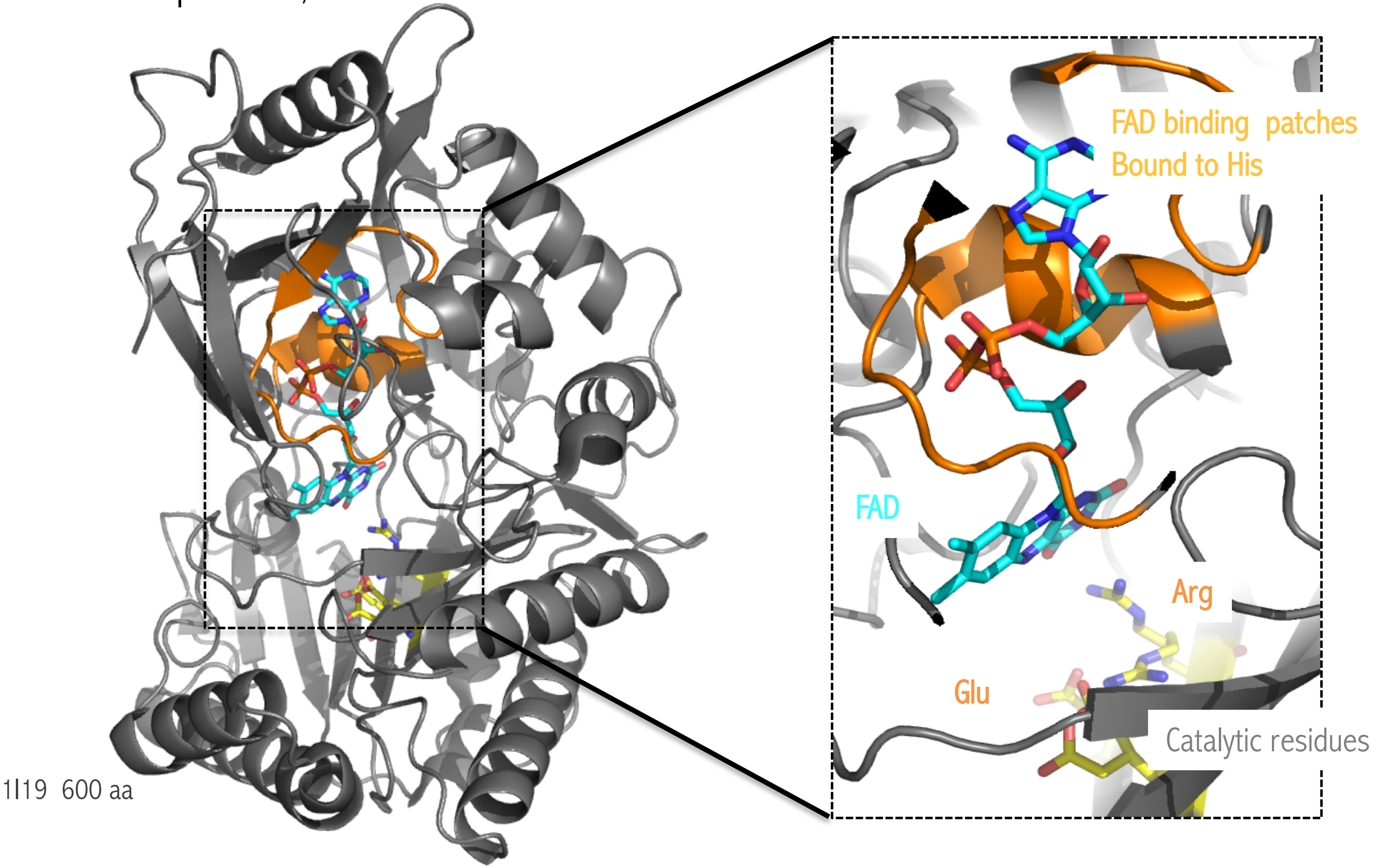


1B4V. pdb & fasta	Cholesterol oxydase from *Streptomyces* Yue et al, 1999 - 1.5 Å
1COY.pdb & fasta	Cholesterol oxydase from *R. hoagii ex B. steroliticum* Li et al, 1993 - 1.8 Å

Choosing the bank of cholesterol oxidases – Bank 2

Cholesterol oxidase:

- monomeric flavoenzyme catalyzes oxidation and isomerization into cholest-4-en-3-one.
- two folds exist, cofactor- dependent, covalent or not.



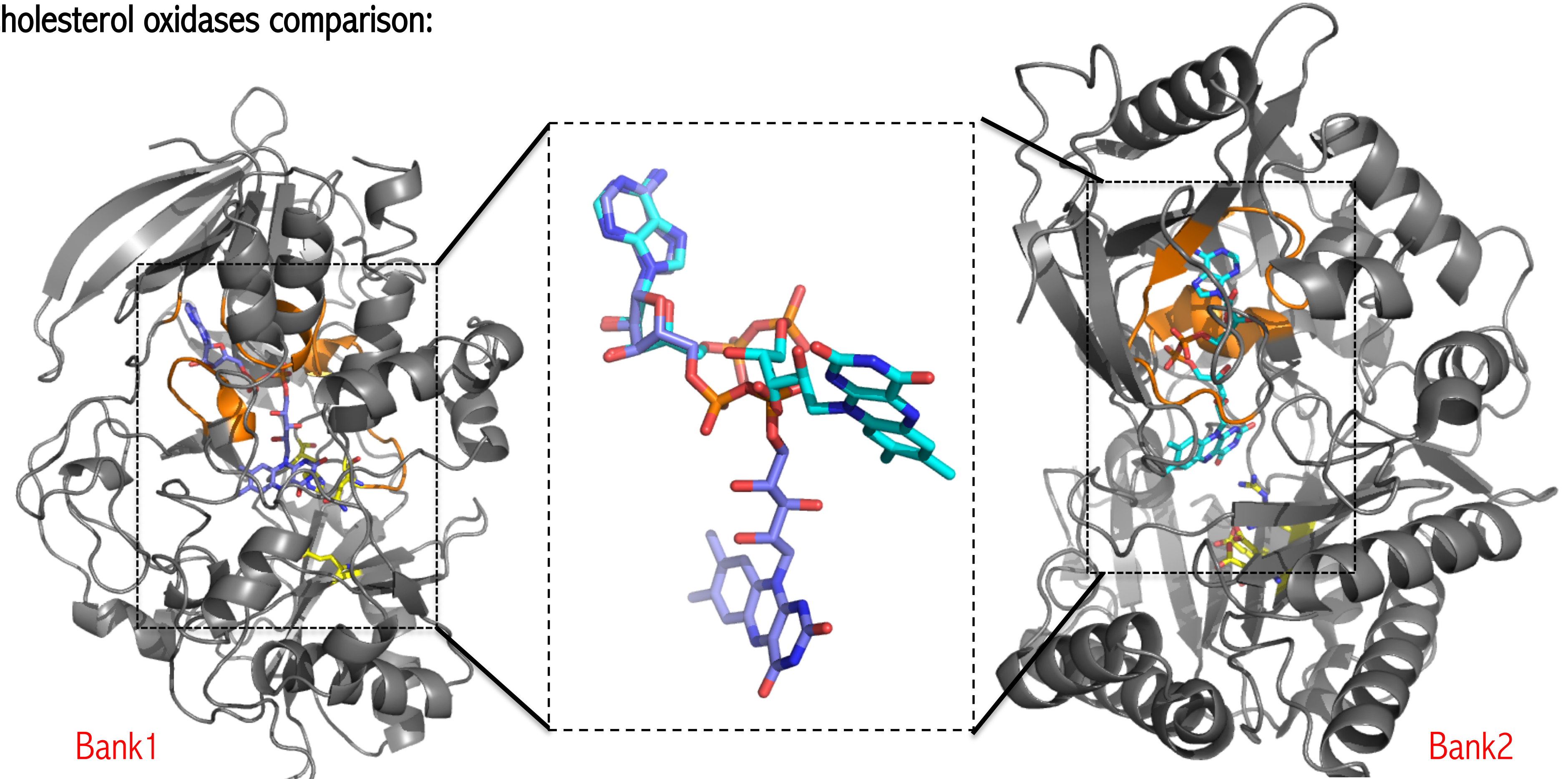
Catalytic site
Glu, Arg

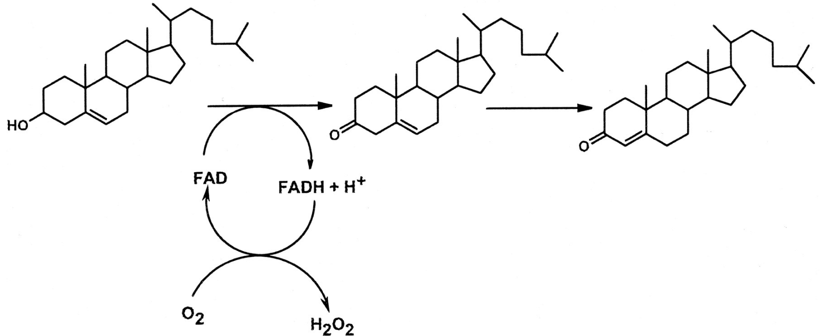
3JS8. pdb & fasta — Cholesterol oxydase from *Chromobacterium sp. DS-1* Sagermann et al, 2010 -1.5 Å

1I19.pdb & fasta — Cholesterol oxydase from *B. sterolíticum* Coulombe et al, 2001 - 1.7 Å

Choosing the bank of cholesterol oxidases – taking one ? Which ? Both ?

Cholesterol oxidases comparison:

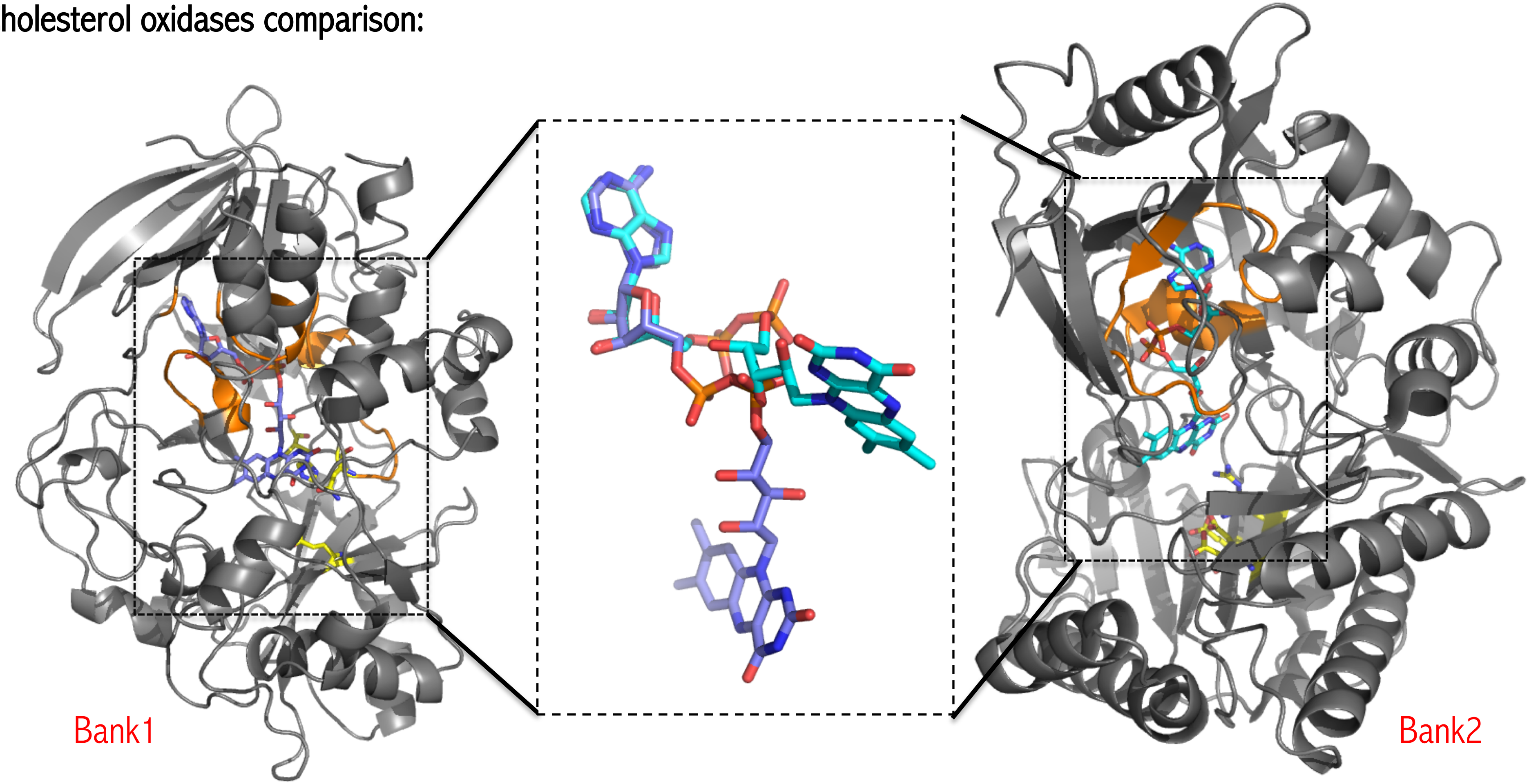




- ✪ 3D structures distinct
- ✪ Catalytic sites different
- ✪ Binding patches divergent
- ✪ FAD binding topologies contrasting despite same redox role

Taking both cholesterol oxidases : 2 banks

Cholesterol oxidases comparison:



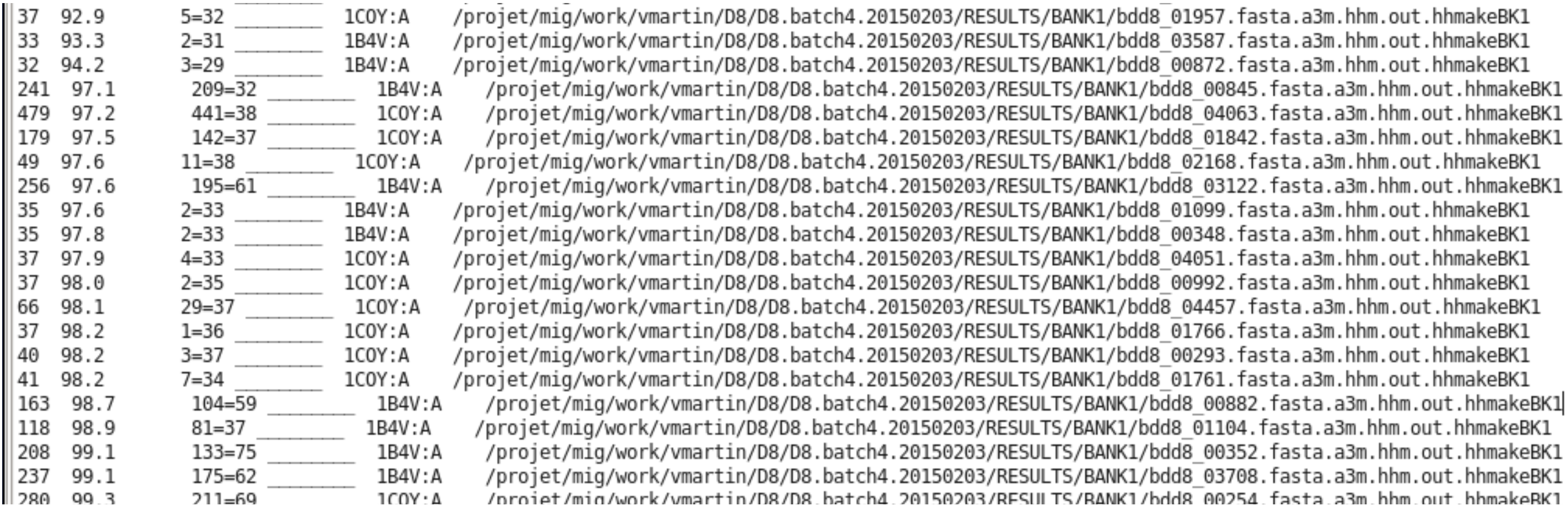
- ✪ 3D structures distinct
- ✪ Catalytic sites different
- ✪ Binding patches divergent
- ✪ FAD binding topologies contrasting despite same redox role

Structural detection with 2 banks required

Identification of hits for *B. dorei* D8 with Bank1 and Bank2

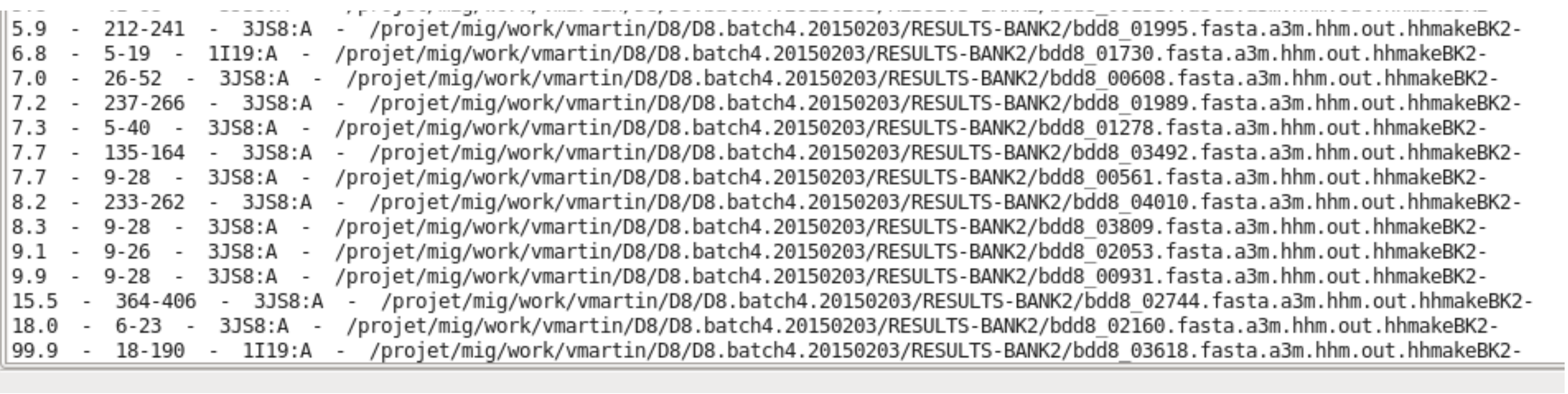
Véronique Martin

✪ *B. dorei* D8 Sorted by Probability – Bank1



```
37   92.9       5=32 ________   1COY:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01957.fasta.a3m.hhm.out.hhmakeBK1
33   93.3       2=31 ________   1B4V:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_03587.fasta.a3m.hhm.out.hhmakeBK1
32   94.2       3=29 ________   1B4V:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00872.fasta.a3m.hhm.out.hhmakeBK1
241   97.1       209=32 ________   1B4V:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00845.fasta.a3m.hhm.out.hhmakeBK1
479   97.2       441=38 ________   1COY:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_04063.fasta.a3m.hhm.out.hhmakeBK1
179   97.5       142=37 ________   1COY:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01842.fasta.a3m.hhm.out.hhmakeBK1
49   97.6       11=38 ________   1COY:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_02168.fasta.a3m.hhm.out.hhmakeBK1
256   97.6       195=61 ________   1B4V:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_03122.fasta.a3m.hhm.out.hhmakeBK1
35   97.6       2=33 ________   1B4V:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01099.fasta.a3m.hhm.out.hhmakeBK1
35   97.8       2=33 ________   1B4V:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00348.fasta.a3m.hhm.out.hhmakeBK1
37   97.9       4=33 ________   1COY:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_04051.fasta.a3m.hhm.out.hhmakeBK1
37   98.0       2=35 ________   1COY:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00992.fasta.a3m.hhm.out.hhmakeBK1
66   98.1       29=37 ________   1COY:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_04457.fasta.a3m.hhm.out.hhmakeBK1
37   98.2       1=36 ________   1COY:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01766.fasta.a3m.hhm.out.hhmakeBK1
40   98.2       3=37 ________   1COY:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00293.fasta.a3m.hhm.out.hhmakeBK1
41   98.2       7=34 ________   1COY:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01761.fasta.a3m.hhm.out.hhmakeBK1
163   98.7       104=59 ________   1B4V:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00882.fasta.a3m.hhm.out.hhmakeBK1
118   98.9       81=37 ________   1B4V:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01104.fasta.a3m.hhm.out.hhmakeBK1
208   99.1       133=75 ________   1B4V:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00352.fasta.a3m.hhm.out.hhmakeBK1
237   99.1       175=62 ________   1B4V:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_03708.fasta.a3m.hhm.out.hhmakeBK1
280   99.3       211=69 ________   1COY:A    /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00254.fasta.a3m.hhm.out.hhmakeBK1
```

✪ *B. dorei* D8 Sorted by Probability – Bank2



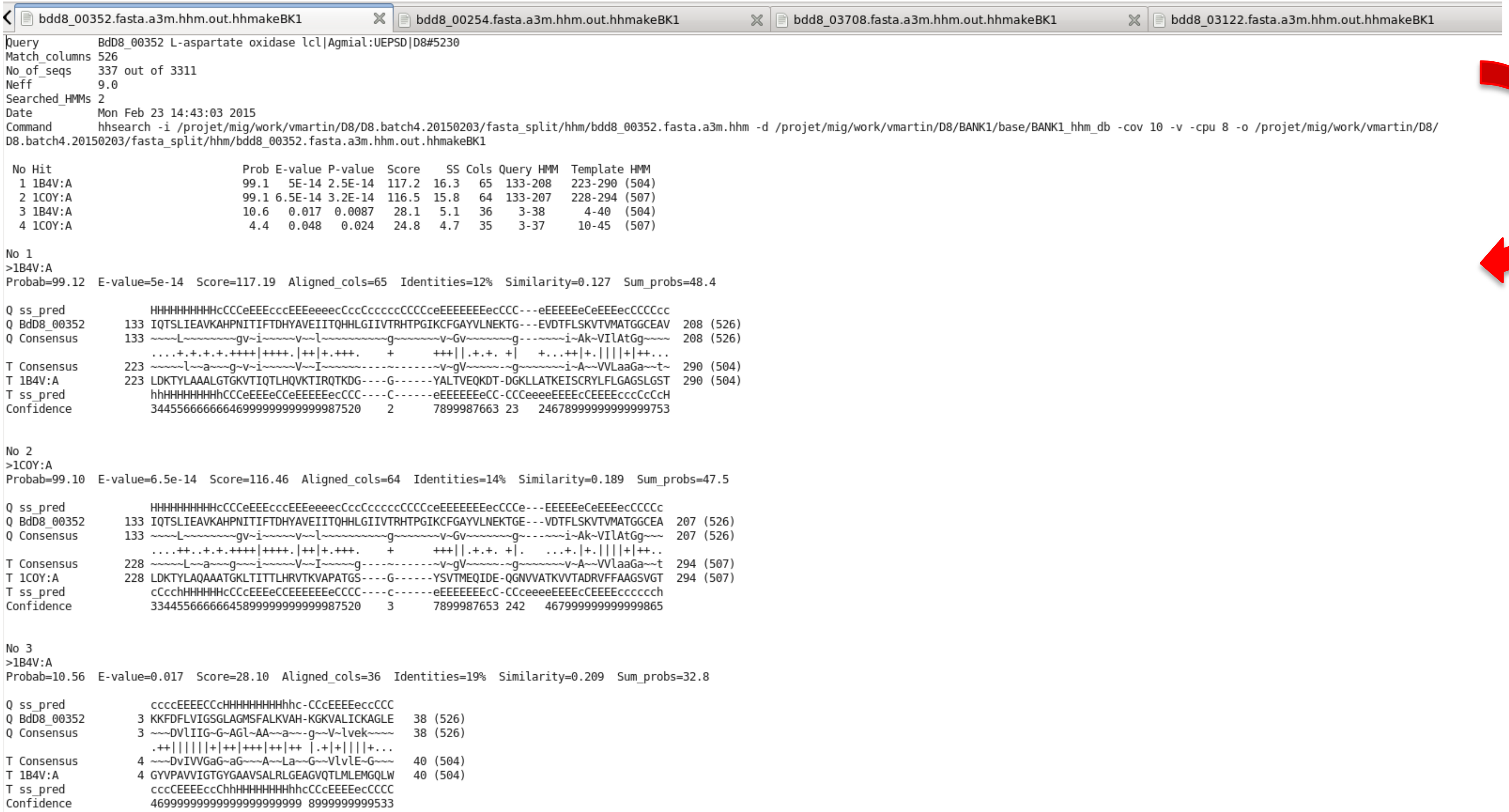
```
5.9  -  212-241  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_01995.fasta.a3m.hhm.out.hhmakeBK2-
6.8  -  5-19  -  1I19:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_01730.fasta.a3m.hhm.out.hhmakeBK2-
7.0  -  26-52  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_00608.fasta.a3m.hhm.out.hhmakeBK2-
7.2  -  237-266  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_01989.fasta.a3m.hhm.out.hhmakeBK2-
7.3  -  5-40  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_01278.fasta.a3m.hhm.out.hhmakeBK2-
7.7  -  135-164  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_03492.fasta.a3m.hhm.out.hhmakeBK2-
7.7  -  9-28  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_00561.fasta.a3m.hhm.out.hhmakeBK2-
8.2  -  233-262  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_04010.fasta.a3m.hhm.out.hhmakeBK2-
8.3  -  9-28  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_03809.fasta.a3m.hhm.out.hhmakeBK2-
9.1  -  9-26  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_02053.fasta.a3m.hhm.out.hhmakeBK2-
9.9  -  9-28  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_00931.fasta.a3m.hhm.out.hhmakeBK2-
15.5  -  364-406  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_02744.fasta.a3m.hhm.out.hhmakeBK2-
18.0  -  6-23  -  3JS8:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_02160.fasta.a3m.hhm.out.hhmakeBK2-
99.9  -  18-190  -  1I19:A  -  /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_03618.fasta.a3m.hhm.out.hhmakeBK2-
```

Analysis of hits for *B. dorei* D8 with Bank1 and Bank2

✪ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

Véronique Martin

99.9	75 aa	bdd8_00352.fasta.a3m.hhm.out.hhmakeBK1	→	L-aspartate oxidase
99.3	69 aa	bdd8_00254.fasta.a3m.hhm.out.hhmakeBK1	→	Putative uncharacterized protein
99.0	62 aa	bdd8_03708.fasta.a3m.hhm.out.hhmakeBK1	→	Succinate Dehydrogenase flavoprotein subunit
97.6	61 aa	bdd8_03122.fasta.a3m.hhm.out.hhmakeBK1	→	Pyridine nucleotide-disulfide oxidoreductase



```
Query         BdD8_00352 L-aspartate oxidase lcl|Agmial:UEPSD|D8#5230
Match_columns 526
No_of_seqs    337 out of 3311
Neff          9.0
Searched_HMMs 2
Date          Mon Feb 23 14:43:03 2015
Command       hhsearch -i /projet/mig/work/vmartin/D8/D8.batch4.20150203/fasta_split/hhm/bdd8_00352.fasta.a3m.hhm -d /projet/mig/work/vmartin/D8/BANK1/base/BANK1_hhm_db -cov 10 -v -cpu 8 -o /projet/mig/work/vmartin/D8/
D8.batch4.20150203/fasta_split/hhm/bdd8_00352.fasta.a3m.hhm.out.hhmakeBK1

 No Hit                             Prob E-value P-value  Score    SS Cols Query HMM  Template HMM
  1 1B4V:A                          99.1   5E-14 2.5E-14  117.2  16.3   65  133-208   223-290 (504)
  2 1COY:A                          99.1 6.5E-14 3.2E-14  116.5  15.8   64  133-207   228-294 (507)
  3 1B4V:A                          10.6   0.017  0.0087   28.1   5.1   36    3-38      4-40  (504)
  4 1COY:A                           4.4   0.048   0.024   24.8   4.7   35    3-37     10-45  (507)

No 1
>1B4V:A
Probab=99.12  E-value=5e-14  Score=117.19  Aligned_cols=65  Identities=12%  Similarity=0.127  Sum_probs=48.4

Q ss_pred              HHHHHHHHHHcCCCeEEEcccEEEeeeecCccCccccCCCCceEEEEEEEEecCCC---eEEEEEeCeEEEecCCCCcc
Q BdD8_00352     133 IQTSLIEAVKAHPNITIFTDHYAVEIITQHHLGIIVTRHTPGIKCFGAYVLNEKTG---EVDTFLSKVTVMATGGCEAV  208 (526)
Q Consensus      133 ~~~~L~~~~~~~~~gv~i~~~~~v~~l~~~~~~~~~~~g~~~~~~~v~Gv~~~~~~~g---~~~~i~Ak~VIlAtGg~~~~  208 (526)
                    ....+.+.+.+.++++|++++.|++|+.+++.    +      +++||.+.+. +|   +...++|+.||||+|++...
T Consensus      223 ~~~~~l~~a~~~g~v~i~~~~~V~~I~~~~~~----~------~v~gV~~~~~-~g~~~~~~~~i~A~~VVLaaGa~~t~  290 (504)
T 1B4V:A         223 LDKTYLAAALGTGKVTIQTLHQVKTIRQTKDG----G------YALTVEQKDT-DGKLLATKEISCRYLFLGAGSLGST  290 (504)
T ss_pred              hhHHHHHHHHhCCCeEEEeCCeEEEEEecCCC----C------eEEEEEEeCC-CCCeeeeEEEEcCEEEEcccCcCcH
Confidence          34455666666646999999999999987520    2      7899987663 23   24678999999999999753


No 2
>1COY:A
Probab=99.10  E-value=6.5e-14  Score=116.46  Aligned_cols=64  Identities=14%  Similarity=0.189  Sum_probs=47.5

Q ss_pred              HHHHHHHHHHcCCCeEEEcccEEEeeeecCccCccccCCCCceEEEEEEEEecCCCe---EEEEEeCeEEEecCCCCc
Q BdD8_00352     133 IQTSLIEAVKAHPNITIFTDHYAVEIITQHHLGIIVTRHTPGIKCFGAYVLNEKTGE---VDTFLSKVTVMATGGCEA  207 (526)
Q Consensus      133 ~~~~L~~~~~~~~~gv~i~~~~~v~~l~~~~~~~~~~~g~~~~~~~v~Gv~~~~~~~g~---~~~i~Ak~VIlAtGg~~~  207 (526)
                    ....++..+.+.++++|++++.|++|+.+++.    +      +++||.+.+. +|.   ...+.|+.||||+|++..
T Consensus      228 ~~~~~L~~a~~~g~~~i~~~~~V~~I~~~~~g----~------~v~gV~~~~~-~g~~~~~~~~v~A~~VVlaaGa~~t  294 (507)
T 1COY:A         228 LDKTYLAQAAATGKLTITTLHRVTKVAPATGS----G------YSVTMEQIDE-QGNVVATKVVTADRVFFAAGSVGT  294 (507)
T ss_pred              cCcchHHHHHHcCCcEEEeCCEEEEEEeCCCC----c------eEEEEEEEcC-CCceeeeEEEEcCEEEEcccccch
Confidence          33445566666645899999999999987520    3      7899987653 242  4679999999999999865


No 3
>1B4V:A
Probab=10.56  E-value=0.017  Score=28.10  Aligned_cols=36  Identities=19%  Similarity=0.209  Sum_probs=32.8

Q ss_pred              ccccEEEECCcHHHHHHHHHhhc-CCcEEEEeccCCC
Q BdD8_00352       3 KKFDFLVIGSGLAGMSFALKVAH-KGKVALICKAGLE   38 (526)
Q Consensus        3 ~~~DVlIIG~G~AGl~AA~~a~~-g~~V~lvek~~~~   38 (526)
                    .++||||||+|++|+++|++|++ |.+|+||||+...
T Consensus        4 ~~~DvIVVGaG~aG~~~A~~La~~G~~VlvlE~G~~~   40 (504)
T 1B4V:A          4 GYVPAVVIGTGYGAAVSALRLGEAGVQTLMLEMGQLW   40 (504)
T ss_pred              cccCEEEEccChhHHHHHHHHhhcCCcEEEEecCCCC
Confidence          46999999999999999999999 8999999999533
```

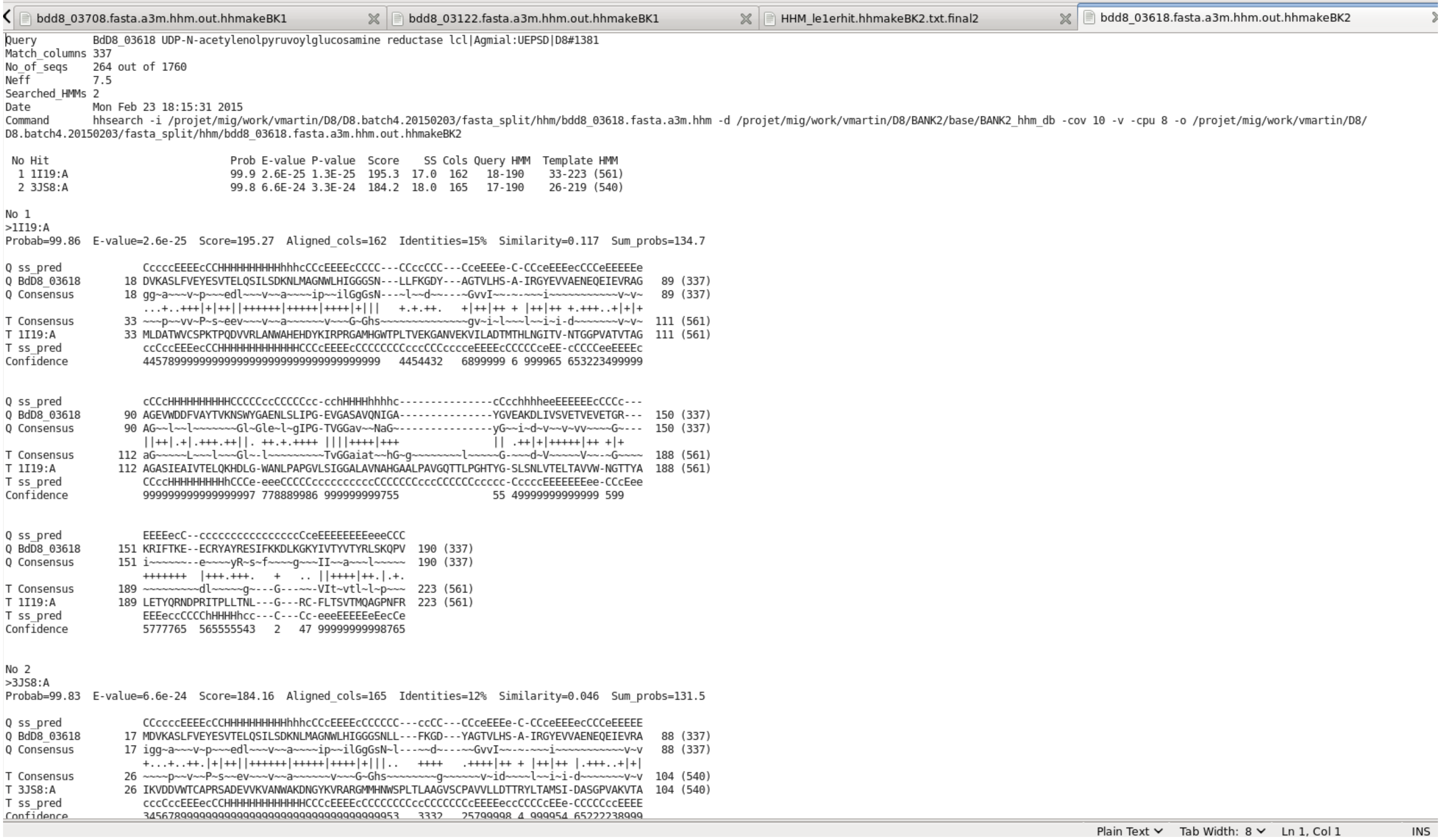
Analysis of hits for *B. dorei* D8 with Bank1 and Bank2

✪ *B. dorei* D8

Sorted by Prob > 95 and size cut off > 50 amino acid residues

Véronique Martin

99.9 172 aa bdd8_03618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



```
Query         BdD8_03618 UDP-N-acetylenolpyruvoylglucosamine reductase lcl|Agmial:UEPSD|D8#1381
Match_columns 337
No_of_seqs    264 out of 1760
Neff          7.5
Searched_HMMs 2
Date          Mon Feb 23 18:15:31 2015
Command       hhsearch -i /projet/mig/work/vmartin/D8/D8.batch4.20150203/fasta_split/hhm/bdd8_03618.fasta.a3m.hhm -d /projet/mig/work/vmartin/D8/BANK2/base/BANK2_hhm_db -cov 10 -v -cpu 8 -o /projet/mig/work/vmartin/D8/
D8.batch4.20150203/fasta_split/hhm/bdd8_03618.fasta.a3m.hhm.out.hhmakeBK2

 No Hit                             Prob E-value P-value  Score    SS Cols Query HMM  Template HMM
  1 1I19:A                          99.9 2.6E-25 1.3E-25  195.3  17.0  162   18-190    33-223 (561)
  2 3JS8:A                          99.8 6.6E-24 3.3E-24  184.2  18.0  165   17-190    26-219 (540)

No 1
>1I19:A
Probab=99.86  E-value=2.6e-25  Score=195.27  Aligned_cols=162  Identities=15%  Similarity=0.117  Sum_probs=134.7

Q ss_pred             CccccEEEEcCCHHHHHHHHHHhhhcCCcEEEEcCCCC---CCccCCC---CceEEEe-C-CCceEEEecCCCeEEEEEe
Q BdD8_03618      18 DVKASLFVEYESVTELQSILSDKNLMAGNWLHIGGGSN---LLFKGDY---AGTVLHS-A-IRGYEVVAENEQEIEVRAG   89 (337)
Q Consensus       18 gg~a~~~v~p~~~edl~~~v~~a~~~~ip~~ilGgGsN---~l~~d~~----GvvI~~-~-~~~i~~~~~~~~~~~~~v~v~   89 (337)
                    ...+..+++|+|++||++++++|+++++|++++|+|||   +.+.++.   +|++|++ + |++|++ +.+++..+|+|+
T Consensus       33 ~~~p~~vv~P~s~eev~~~v~~a~~~~~~v~~~G~Ghs~~~~~~~~~~~~~~gv~i~l~~~l~~i~i~d~~~~~~~~v~v~  111 (561)
T 1I19:A         33 MLDATWVCSPKTPQDVVRLANWAHEHDYKIRPRGAMHGWTPLTVEKGANVEKVILADTMTHLNGITV-NTGGPVATVTAG  111 (561)
T ss_pred             ccCccEEEecCCHHHHHHHHHHHHHHCCCcEEEEcCCCCCCCCcccCCCccccEEEEcCCCCCceEE-cCCCCeeEEEEc
Confidence           445789999999999999999999999999999999   4454432   6899999 6 999965 653223499999


Q ss_pred             cCCcHHHHHHHHHHCCCCCccCCCCCcc-cchHHHHhhhhc---------------cCcchhhheeEEEEEEcCCCc---
Q BdD8_03618      90 AGEVWDDFVAYTVKNSWYGAENLSLIPG-EVGASAVQNIGA---------------YGVEAKDLIVSVETVEVETGR---  150 (337)
Q Consensus       90 AG~~l~~l~~~~~~~Gl~Gle~l~gIPG-TVGGav~~NaG~---------------yG~~i~d~v~~v~vv~~~G~---  150 (337)
                    ||++|.+|.+++.++||. ++.+.++++ ||||++++|+++               || .++|+|++++++|++ +|+
T Consensus      112 aG~~~~~L~~~l~~~Gl~~l~~~~~~~~~TvGGaiat~~hG~g~~~~~~~~~l~~~~~~G~~~~d~V~~~~~~V~~~~G~~~~  188 (561)
T 1I19:A        112 AGASIEAIVTELQKHDLG-WANLPAPGVLSIGGALAVNAHGAALPAVGQTTLPGHTYG-SLSNLVTELTAVVW-NGTTYA  188 (561)
T ss_pred             CCccHHHHHHHHHHhCCCe-eeeCCCCCccccccccCCCCCCcccCCCCCCcccc-CccccEEEEEEee-CCcEee
Confidence           99999999999999997 778889986 999999999755               55 49999999999 599


Q ss_pred             EEEEecC--ccccccccccccccCceEEEEEEEEeeeCCC
Q BdD8_03618     151 KRIFTKE--ECRYAYRESIFKKDLKGKYIVTYVTYRLSKQPV  190 (337)
Q Consensus      151 i~~~~~~~--e~~~~yR~s~f~~~~g~~~II~~a~~~l~~~~~  190 (337)
                    +++++++  |+++.+++.   +   ..  ||++++|++.|.+.
T Consensus      189 ~~~~~~~~~~~dl~~~~~g~---G---~~~-VIt~vtl~l~p~~~  223 (561)
T 1I19:A        189 LETYQRNDPRITPLLTNL---G---RC-FLTSVTMQAGPNFR  223 (561)
T ss_pred             EEEeccCCCChHHHhcc---C---Cc-eeeEEEEEeEecCe
Confidence           5777765  565555543   2   47 999999998765


No 2
>3JS8:A
Probab=99.83  E-value=6.6e-24  Score=184.16  Aligned_cols=165  Identities=12%  Similarity=0.046  Sum_probs=131.5

Q ss_pred             CCccccEEEEcCCHHHHHHHHHHhhhcCCcEEEEcCCCCCC---ccCC---CCceEEEe-C-CCceEEEecCCCeEEEEE
Q BdD8_03618      17 MDVKASLFVEYESVTELQSILSDKNLMAGNWLHIGGGSNLL---FKGD---YAGTVLHS-A-IRGYEVVAENEQEIEVRA   88 (337)
Q Consensus       17 igg~a~~~v~p~~~edl~~~v~~a~~~~ip~~ilGgGsN~l---~~d~---~~GvvI~~~-~-~~~i~~~~~~~~~~~~v~v   88 (337)
                    +...+..++.|+|++||++++++|+++++|++++|+|||..   ++++   .++++|++ + |++|++ |.+++..+|+|
T Consensus       26 ~~~p~~v~~P~s~~ev~~~v~~a~~~~~~v~~~G~Ghs~~~~~~~~g~~~~~v~id~~~~l~~i~i~d~~~~~~~~v~v  104 (540)
T 3JS8:A         26 IKVDDVWTCAPRSADEVVKVANWAKDNGYKVRARGMMHNWSPLTLAAGVSCPAVVLLDTTRYLTAMSI-DASGPVAKVTA  104 (540)
T ss_pred             cccCccEEEecCCHHHHHHHHHHHHHHCCCcEEEEcCCCCCCCCccCCCCCCCcEEEEeccCCCCcEEe-CCCCCccEEEE
Confidence           345678999999999999999999999999999999953   3332   25799998 4 999954 65222238999
```

Identification of hits for *B. dorei 175^T* with Bank1 and Bank2

✪ *B. dorei* 175^T Sorted by Prob > 95 and size cut off > 50 amino acid residues

Véronique Martin

99.4 % 70 aa bacdor_01751.fasta.a3m.hhm.out.hhmakeBK1 → Hypothetical protein
99.1 % 75 aa bacdor_01928.fasta.a3m.hhm.out.hhmakeBK1 → L-aspartate oxidase
99.1 % 62 aa bacdor_02089.fasta.a3m.hhm.out.hhmakeBK1 → Succinate Dehydrogenase flavoprotein
97.5 % 61 aa bacdor_03886.fasta.a3m.hhm.out.hhmakeBK1 → Pyridine nucleotide oxido reductase

✪ *B. dorei* 175^T Sorted by Prob > 95 and size > 50 amino acid residues

99.9 % 172 aa bacdor_03524.fasta.a3m.hhm.out.hhmakeBK2 →UDP-N-acetyl muramate dehydrogenase

Comparison of hits for *B. dorei 175^T* and D8 with Bank1 and Bank2

✪ *B. dorei* 175^T Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.4 % 70 aa bacdor_01751.fasta.a3m.hhm.out.hhmakeBK1 → Hypothetical protein
99.1 % 75 aa bacdor_01928.fasta.a3m.hhm.out.hhmakeBK1 → L-aspartate oxidase
99.1 % 62 aa bacdor_02089.fasta.a3m.hhm.out.hhmakeBK1 → Succinate Dehydrogenase flavoprotein
97.5 % 61 aa bacdor_03886.fasta.a3m.hhm.out.hhmakeBK1 → Pyridine nucleotide oxido reductase

✪ *B. dorei* 175^T Sorted by Prob > 95 and size > 50 amino acid residues

99.9 % 172 aa bacdor_03524.fasta.a3m.hhm.out.hhmakeBK2 →UDP-N-acetyl muramate dehydrogenase

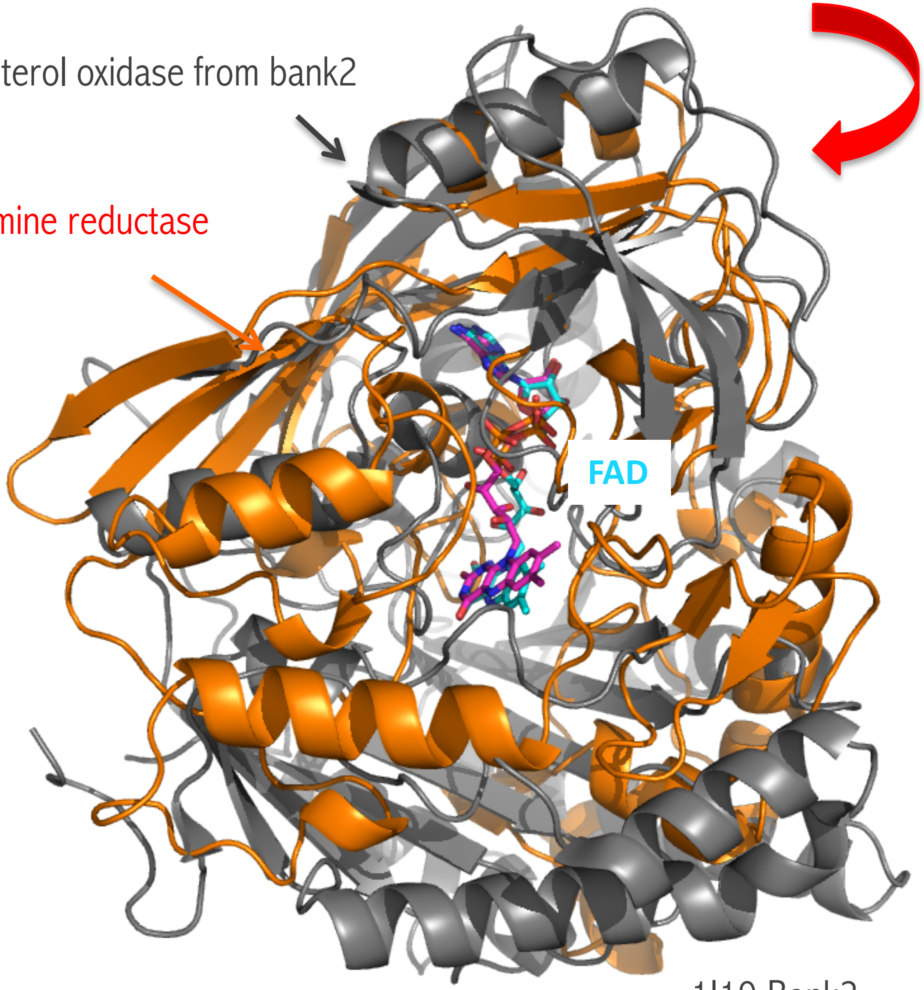
=

99.9 % 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase

Analysis of hits with bank2 two for the two strains

✪ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



1I19 Cholesterol oxidase from bank2

UDP-N-acetylenolpyruvoylglucosamine reductase

Oxidoreductase MurB

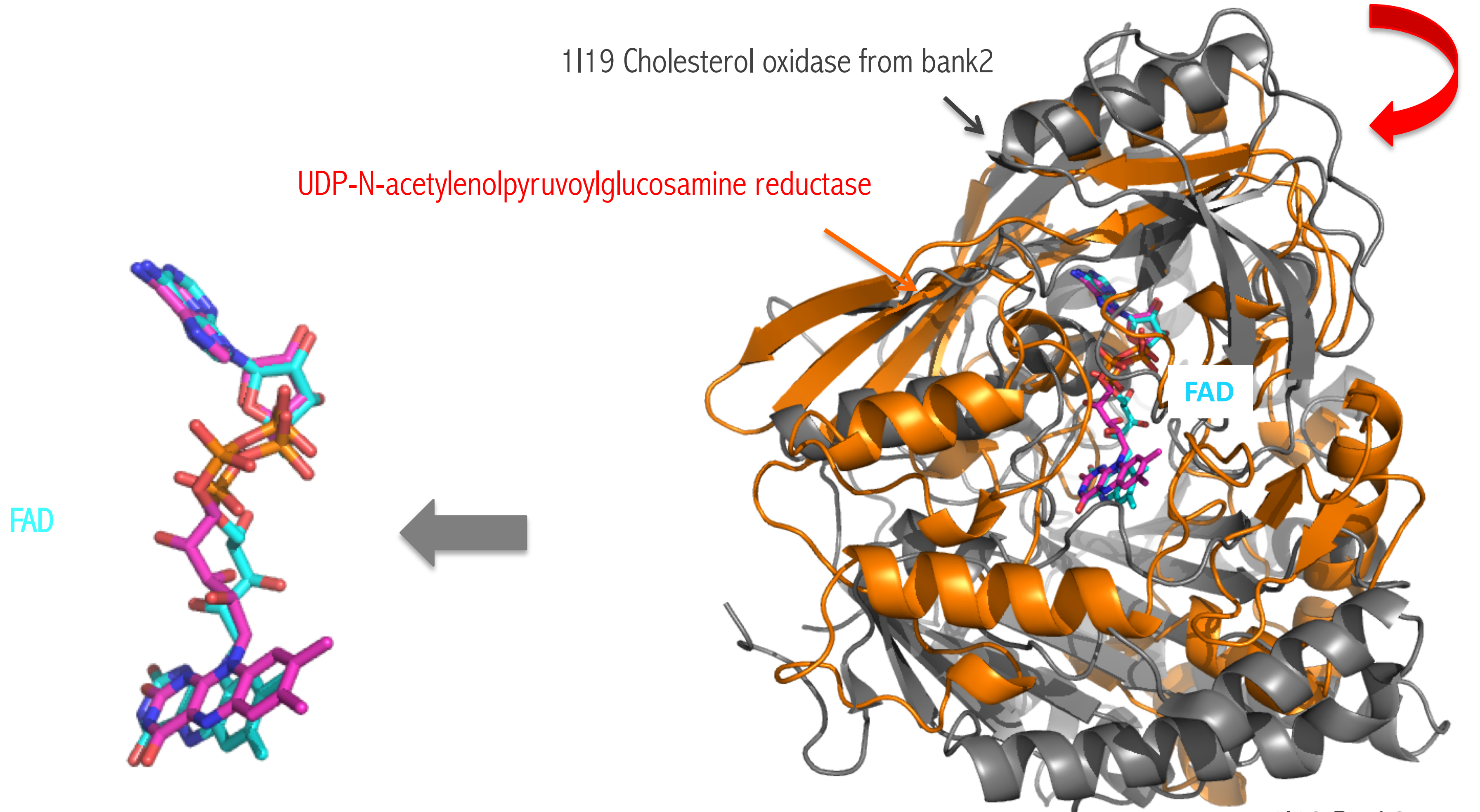
1I19 Bank2

1MBB E.coli Gram-

Analysis of hits with bank2 two for the two strains

✪ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



1I19 Cholesterol oxidase from bank2

UDP-N-acetylenolpyruvoylglucosamine reductase

FAD

FAD

1I19 Bank2

1MBB E.coli Gram-

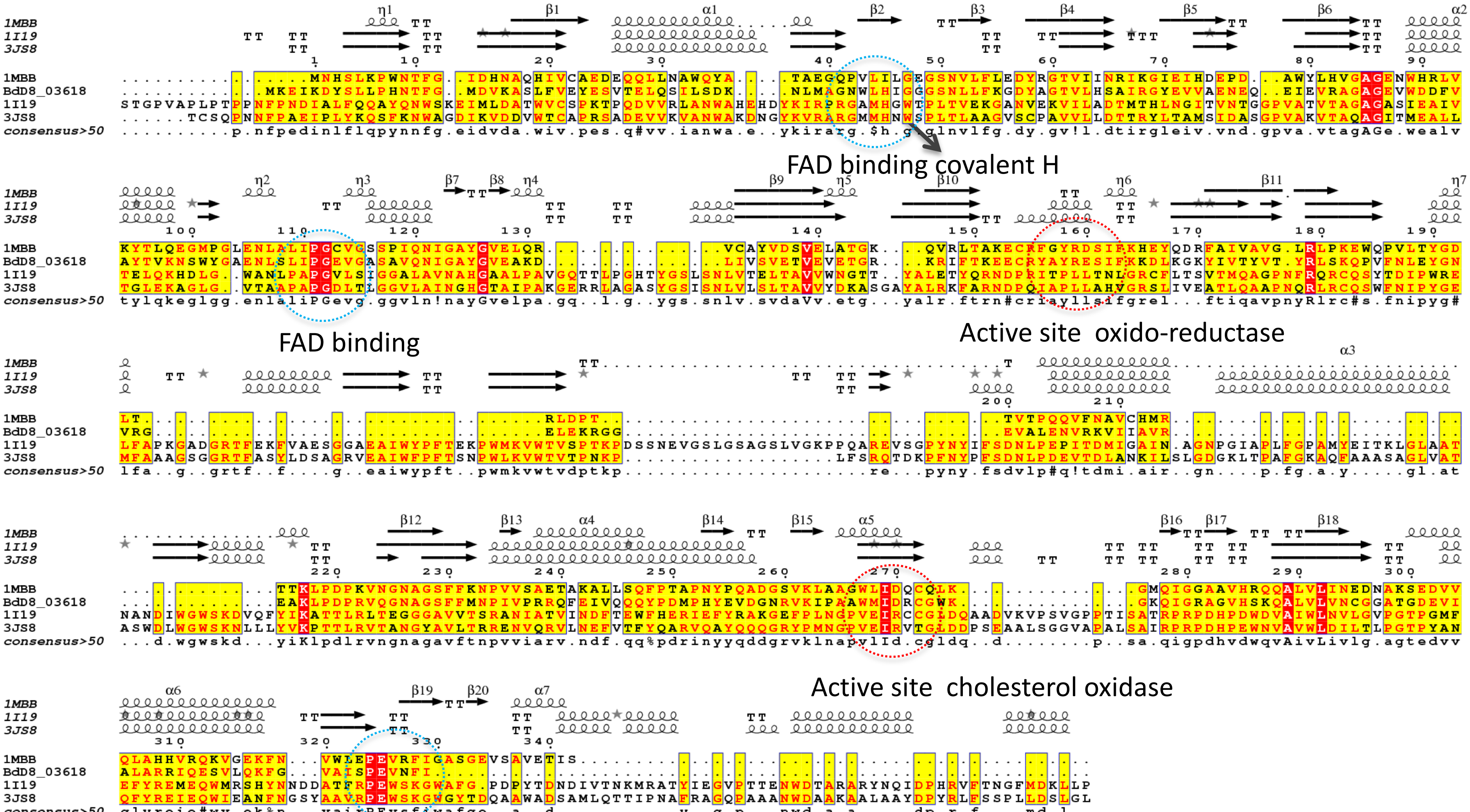
Oxidoreductase MurB

One major hit for *B. dorei* D8 and D^{175T} with Bank2 : possibly an enzyme

✪ *B. dorei* D8

Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



FAD binding covalent H

FAD binding

Active site oxido-reductase

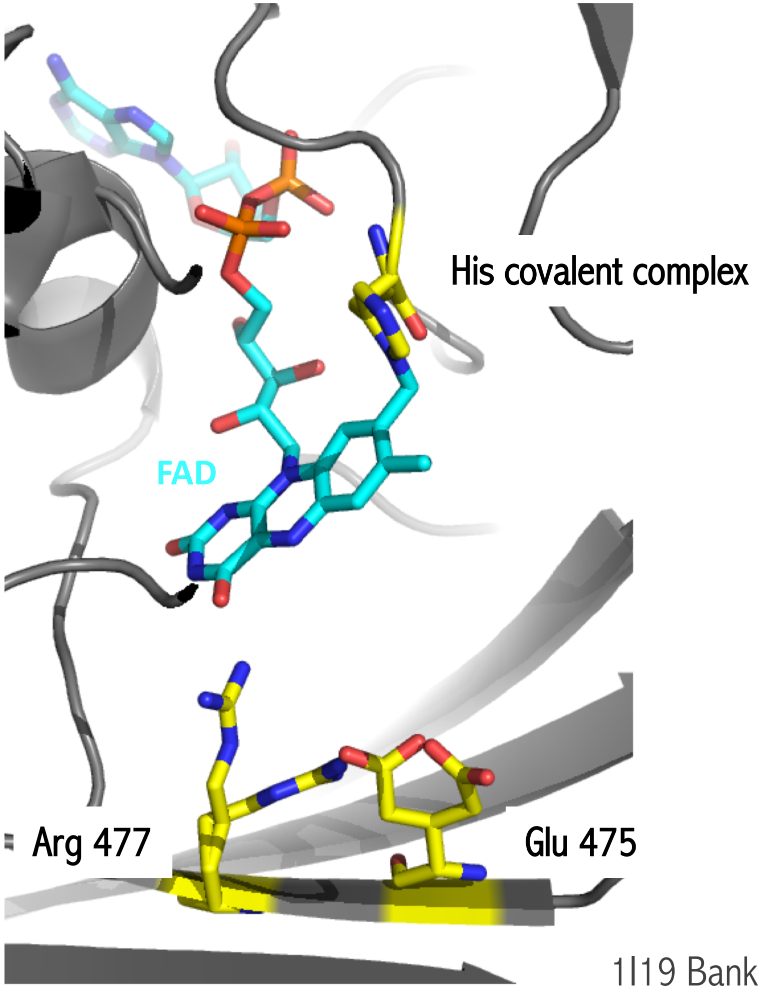
Active site cholesterol oxidase

FAD binding

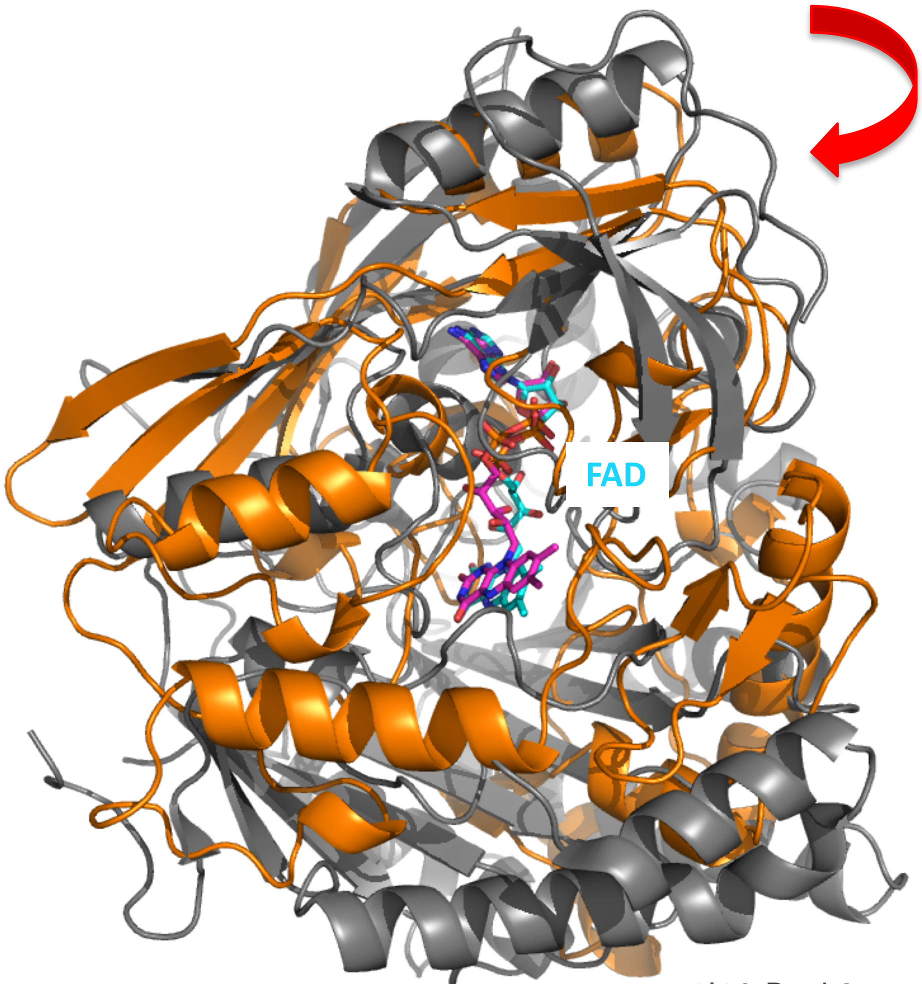
The hit for *B. dorei* D8 and D^{175T} with Bank2 : possibly a flavo-enzyme

✪ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



1I19 Bank2



Oxidoreductase MurB

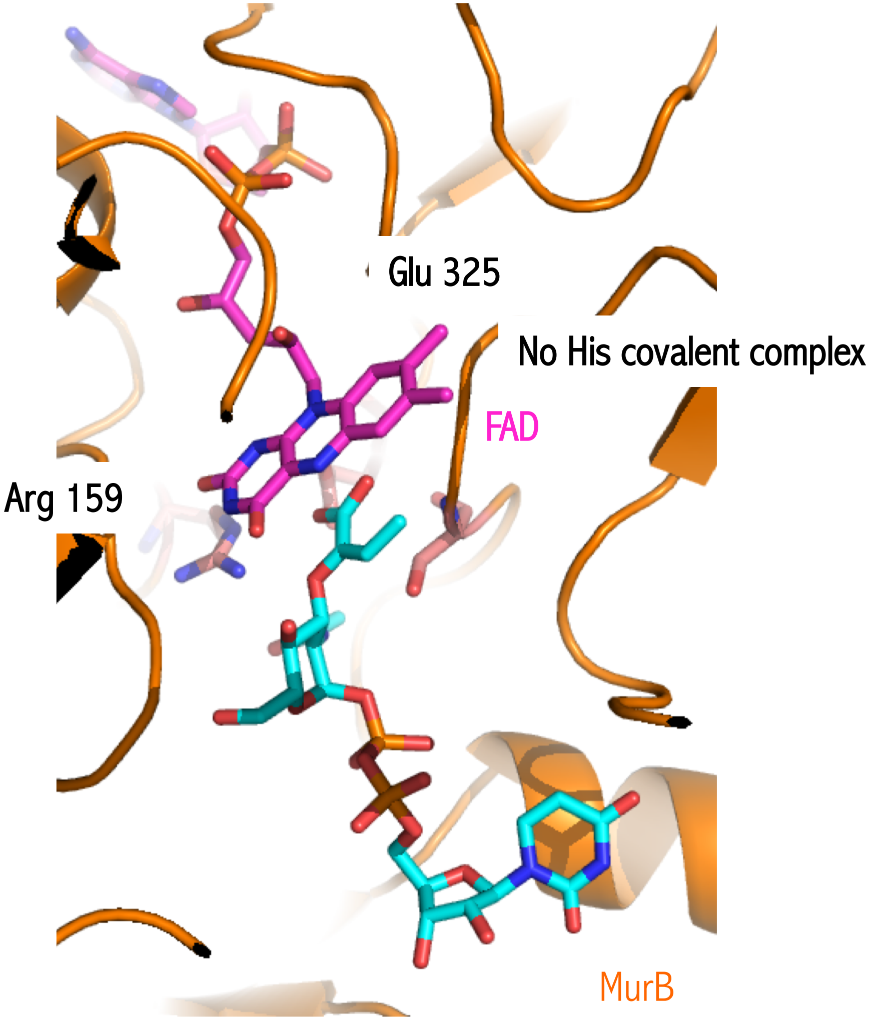
1I19 Bank2

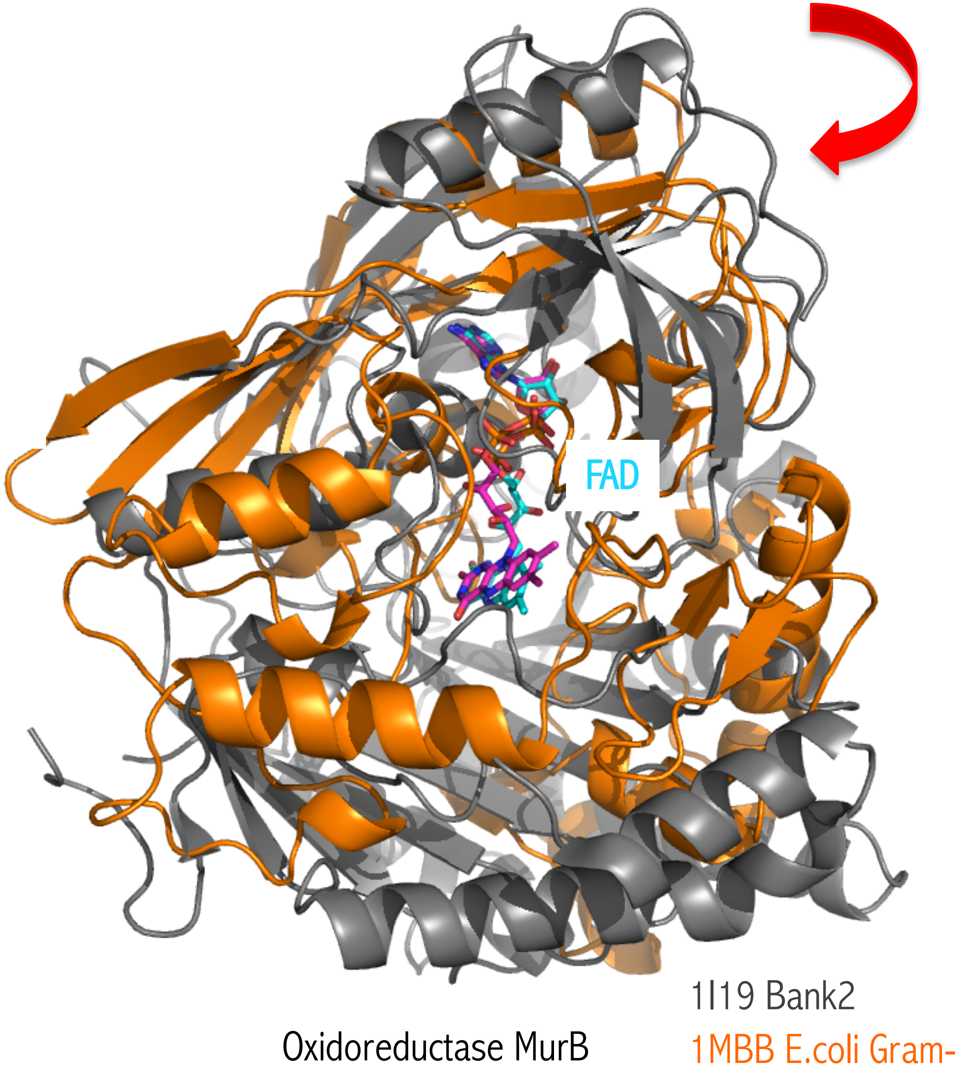
1MBB E.coli Gram-

The hit for *B. dorei* D8 and D^{175T} with Bank2 : possibly a flavo-enzyme

✪ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase

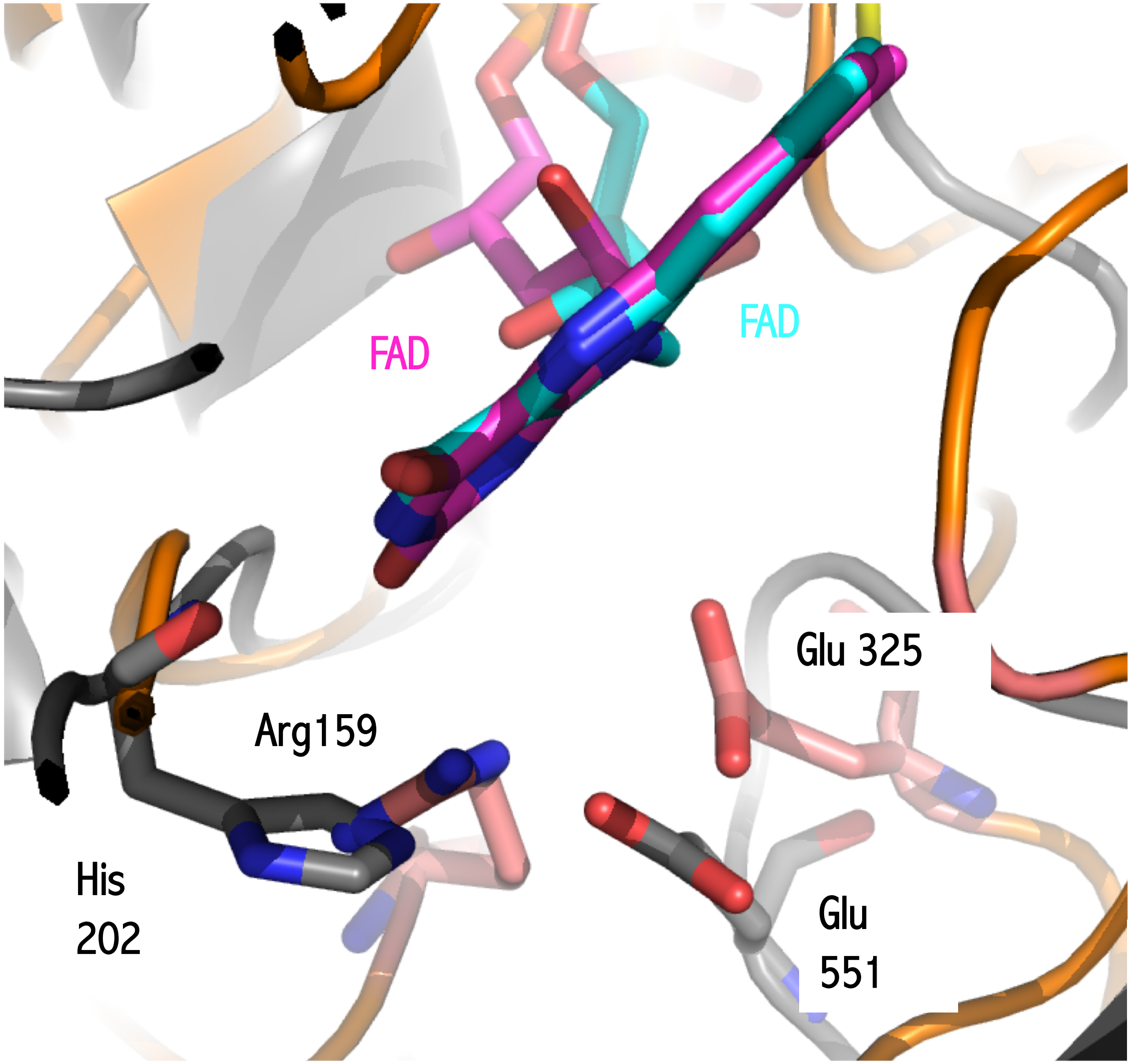




The hit for *B. dorei* D8 and D^{175T} with Bank2 : possibly a flavo-enzyme

✪ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



1I19 Bank2

1MBB MurB

Conclusion from screening of *B. dorei 175ᵀ* with Bank1 and Bank2

✪ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.4 % 70 aa bacdor_01751.fasta.a3m.hhm.out.hhmakeBK1 → Hypothetical protein
99.1 % 75 aa bacdor_01928.fasta.a3m.hhm.out.hhmakeBK1 → L-aspartate oxidase
99.1 % 62 aa bacdor_02089.fasta.a3m.hhm.out.hhmakeBK1 → Succinate Dehydrogenase flavoprotein
97.5 % 61 aa bacdor_03886.fasta.a3m.hhm.out.hhmakeBK1 → Pyridine nucleotide oxido reductase

✪ *B. dorei* 175^T Sorted by Prob > 95 and size > 50 amino acid residues

99.9 % 172 aa bacdor_03524.fasta.a3m.hhm.out.hhmakeBK2 →UDP-N-acetyl muramate dehydrogenase

=

99.9 % 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase

⇩

Cloning and expression tests in progress.

Then, purification, activity test and crystallization assays

Cholesterol oxidases are secreted. This contains a signal peptide.

Conclusion from screening of *B. dorei* 175^T with Bank1 and Bank2

- *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.4 % 70 aa bacdor_01751.fasta.a3m.hhm.out.hhmakeBK1 → Hypothetical protein
99.1 % 75 aa bacdor_01928.fasta.a3m.hhm.out.hhmakeBK1 → L-aspartate oxidase
99.1 % 62 aa bacdor_02089.fasta.a3m.hhm.out.hhmakeBK1 → Succinate Dehydrogenase flavoprotein
97.5 % 61 aa bacdor_03886.fasta.a3m.hhm.out.hhmakeBK1 → Pyridine nucleotide oxido reductase

- *B. dorei* 175^T Sorted by Prob > 95 and size > 50 amino acid residues

99.9 % 172 aa bacdor_03524.fasta.a3m.hhm.out.hhmakeBK2 →UDP-N-acetyl muramate dehydrogenase

=

99.9 % 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase

↓

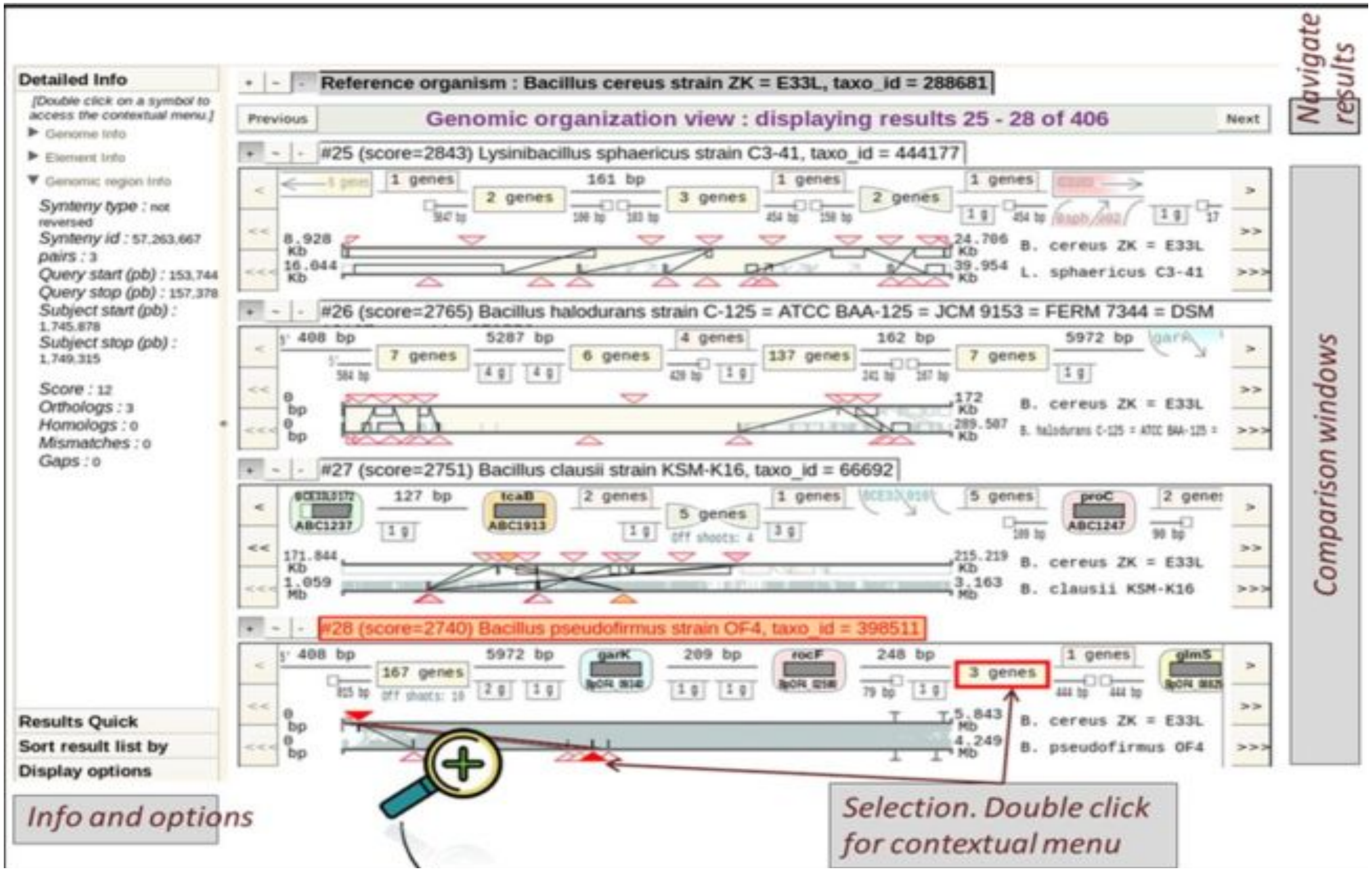
Analysis of synthenies by Thomas Lacroix using Insyght.

Genome comparison using Insyght

http://genome.jouy.inra.fr/Insyght

Thomas Lacroix

→ Powerful tool for genes comparison and syntheny inferences



Detailed Info

[Double click on a symbol to access the contextual menu.]

- ▶ Genome info
- ▶ Element info
- ▼ Genomic region info

Synteny type : not reversed
Synteny id : 57,263,667
pairs : 3
Query start (pb) : 153,744
Query stop (pb) : 157,378
Subject start (pb) : 1,745,878
Subject stop (pb) : 1,749,315

Score : 12
Orthologs : 3
Homologs : 0
Mismatches : 0
Gaps : 0

Reference organism : Bacillus cereus strain ZK = E33L, taxo_id = 288681

Genomic organization view : displaying results 25 - 28 of 406

#25 (score=2843) Lysinibacillus sphaericus strain C3-41, taxo_id = 444177

#26 (score=2765) Bacillus halodurans strain C-125 = ATCC BAA-125 = JCM 9153 = FERM 7344 = DSM

#27 (score=2751) Bacillus clausii strain KSM-K16, taxo_id = 66692

#28 (score=2740) Bacillus pseudofirmus strain OF4, taxo_id = 398511

Lacroix *et al*, Nar 2014

Analysis of the 5 genes from 3D, using Insyght

✪ http://genome.jouy.inra.fr/Insyght

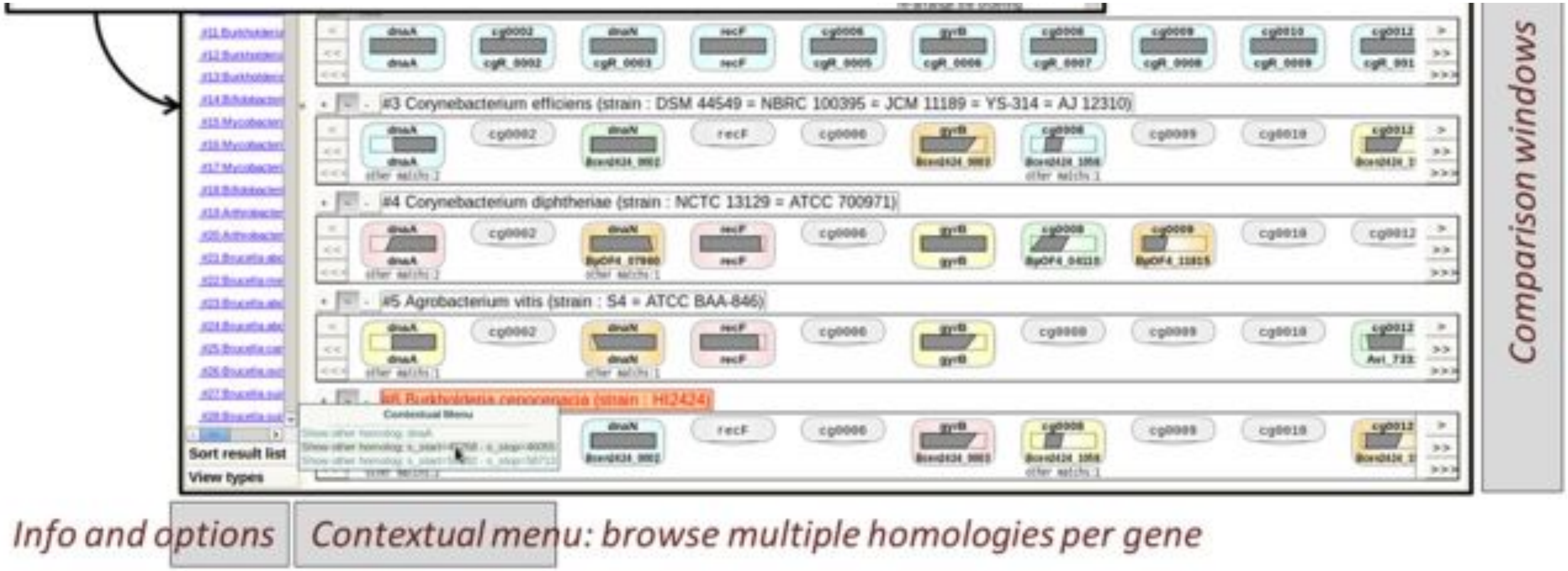
Thomas Lacroix

→ **Powerful tool for genes comparison and syntheny inferences**

Synthenies : co-localization of homologuous loci accross species

Comparison 2 vs 2 with a reference genome.

No syntheny. No clear information



#3 Corynebacterium efficiens (strain : DSM 44549 = NBRC 100395 = JCM 11189 = YS-314 = AJ 12310)

#4 Corynebacterium diphtheriae (strain : NCTC 13129 = ATCC 700971)

#5 Agrobacterium vitis (strain : S4 = ATCC BAA-846)

#6 Burkholderia cenocepacia (strain : HI2424)

Lacroix *et al*, Nar 2014

E. Coprostanoligenes

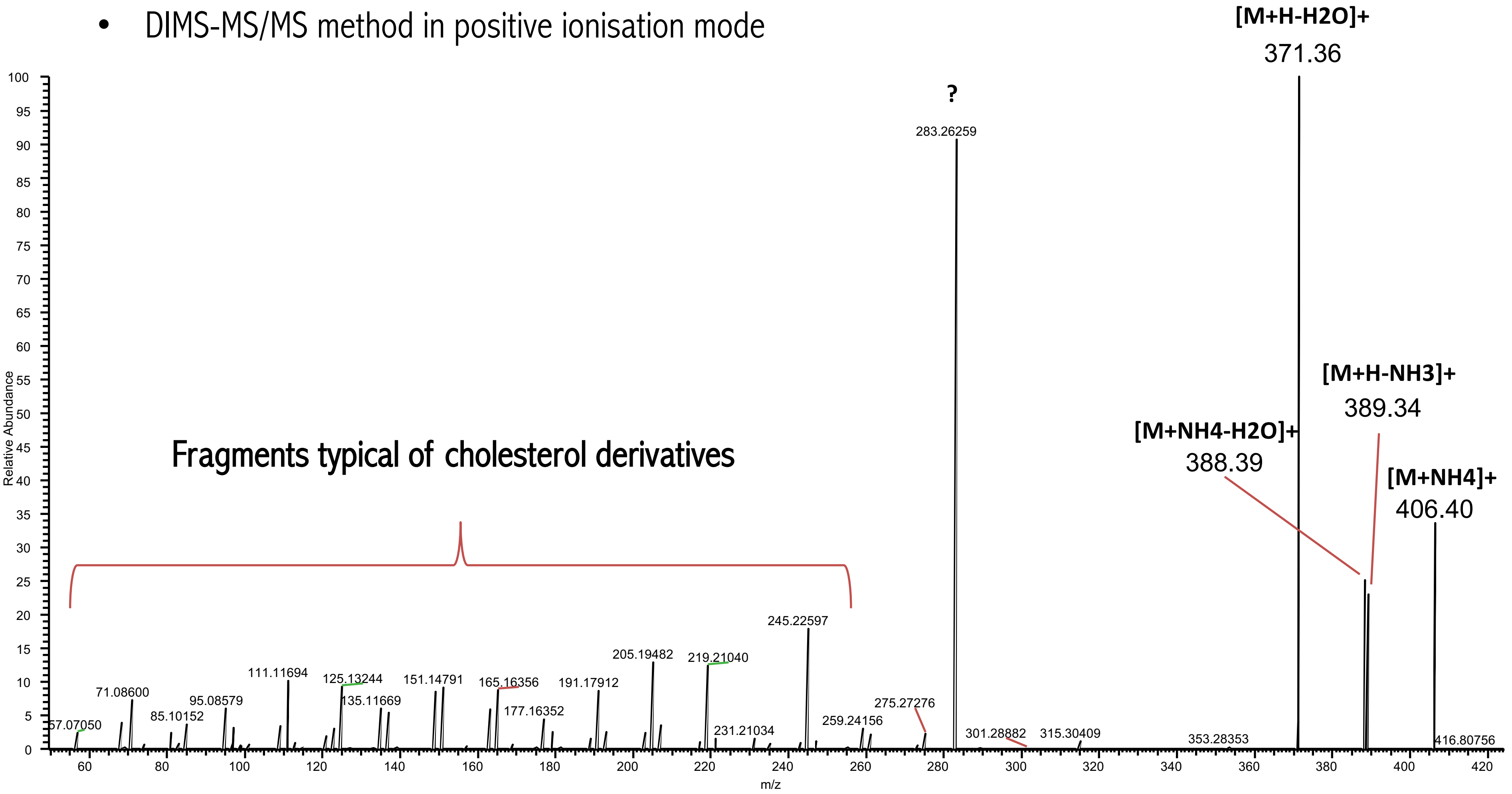
Catherine. Juste
Fabienne Beguet-Crespel

✪ Direct introduction of the COPROSTANOL standard by DIMS-MS/MS

Samples resuspended in 500 mL of Chloroform/ Methanol (50/50)
Diluted to 1/100eme before analysis

- Ions detection corresponding to COPROSTANOL
- DIMS-MS/MS method in positive ionisation mode

Coprostanol

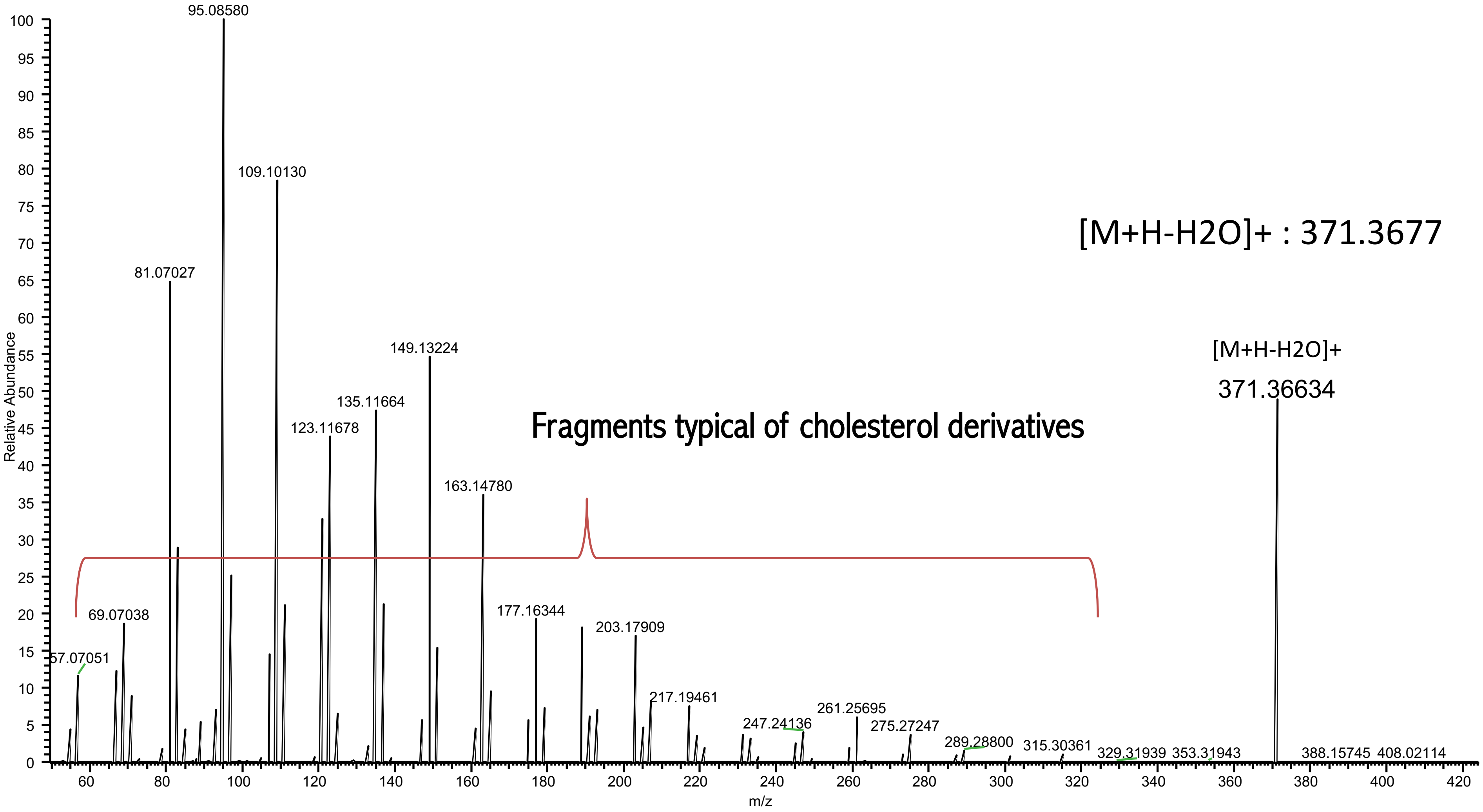


Fragments typical of cholesterol derivatives

E. Coprostanoligenes samples

Catherine. Juste
Fabienne Beguet-Crespel

- Ions detection corresponding to COPROSTANOL
- DIMS-MS/MS method in positive ionisation mode



[M+H-H2O]+ : 371.3677

Fragments typical of cholesterol derivatives

E. Coprostanoligenes samples

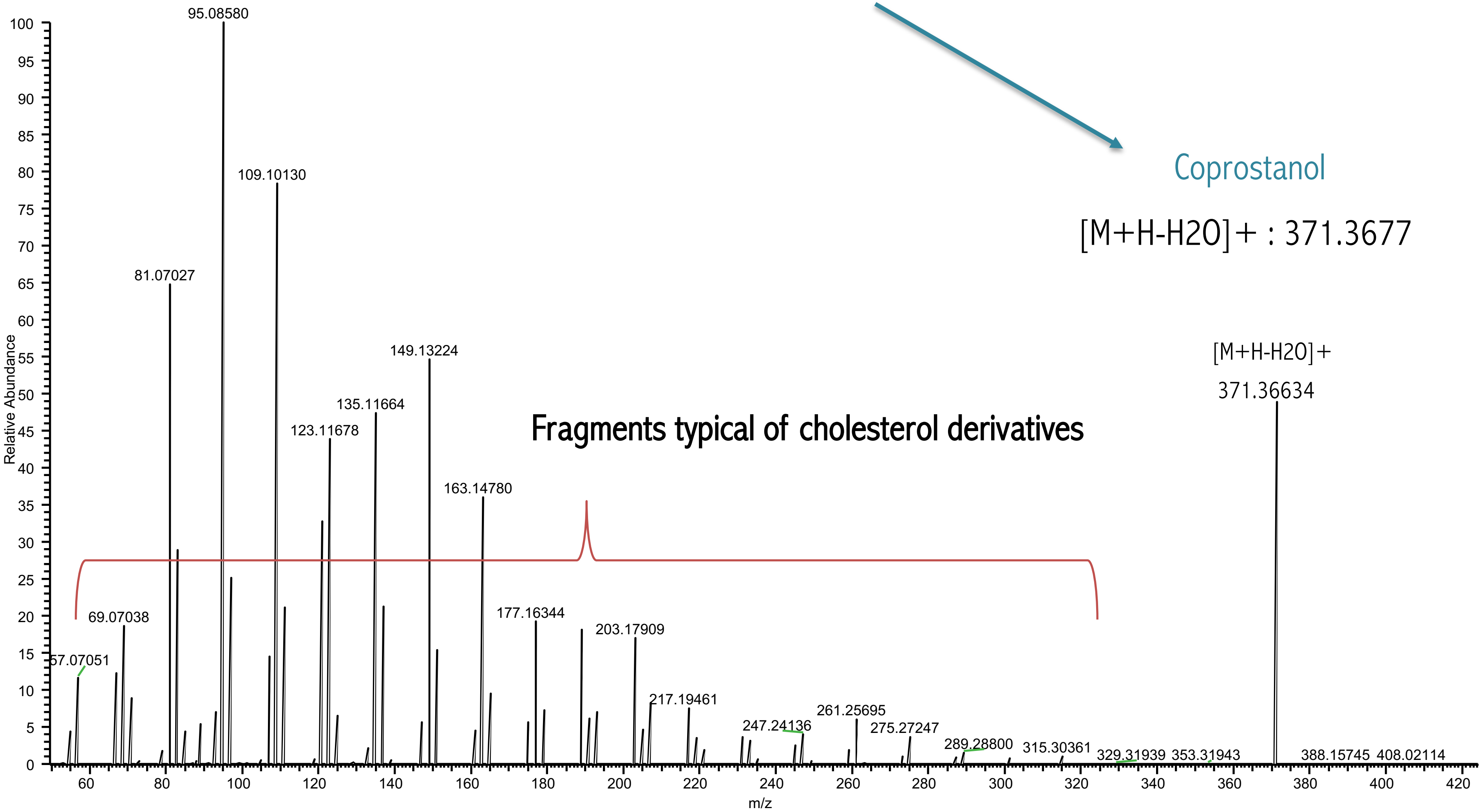
Catherine. Juste
Fabienne Beguet-Crespel

- Ions detection corresponding to COPROSTANOL
- DIMS-MS/MS method in positive ionisation mode

Coprostanol is present & preferentially detected

Coprostanol

$[M+H-H_2O]^+$: 371.3677



Fragments typical of cholesterol derivatives

Whole genomes comparison, using Insyght

- Analysis of candidates- Comparison within cholesterol degrading genomes and others
  Comparison between *E. coprostanoligenes (2017), B. Dorei D8, Parabacteroides distasonis* shown to lower cholesterolemia while producing coprostanol. PhD thesis Aline Potiron.

- Annotation comparator's analysis

- Genomic organisation's analysis

→ One hit Interesting: Analysis in progress with Thomas Lacroix.
Present in degrading bacteria, absent in non degrading bacteria

→ Annotated as lipolytic protein G-D-S-L family, phospholipase, major facilitator superfamily

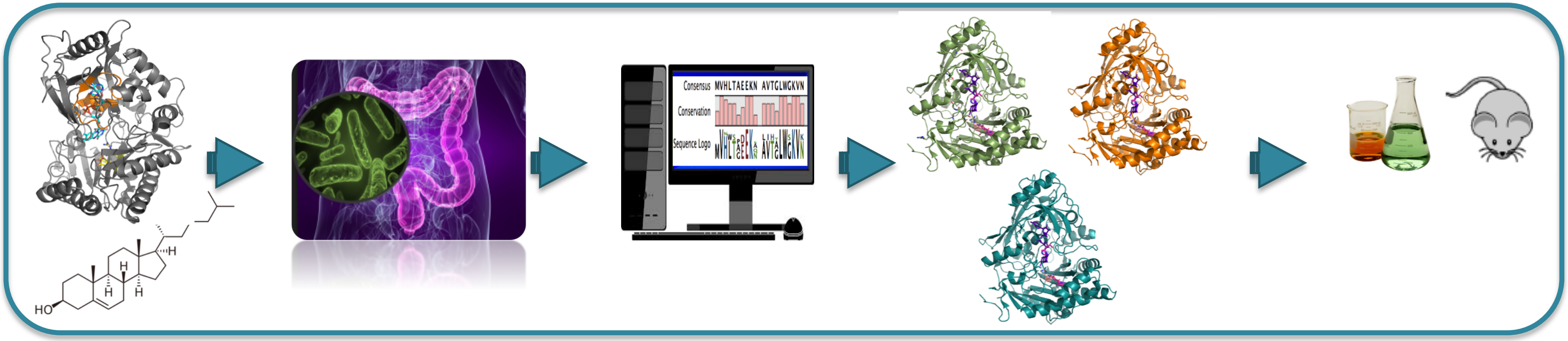
→ Cholesterol oxidase with a Ser His Asp conserved catalytic site plus the oxyanionic hole identified

Homology modeling and docking of cholesterol in progress

Cholesterol conversion in the gut microbiota

- No significant difference between the two genomes. Similar enzymes are found.
- Transcriptomic analysis of both chromosome and plasmids of *B. dorei 175*^{T} & *B. dorei D8?*
- Activity test and 3D structure validation of our cholesterol-oxidase hit in bank2.
- Genomic & transcriptomic analysis of both chromosome & plasmids of *E. coprostanoligenes*.
- Can we rule out some synergy among bacteria of the microbiota?
- Promising hit coming from entire genomes comparison: cholesterol esterase function and pathway.

Towards the Structural Screening of Microbial Ecosystems



Outline

- Cholesterol conversion in the gut microbiota: the mystery enzyme(s) ?

- MetaFoldScan project : 3D screening of gut microbiota

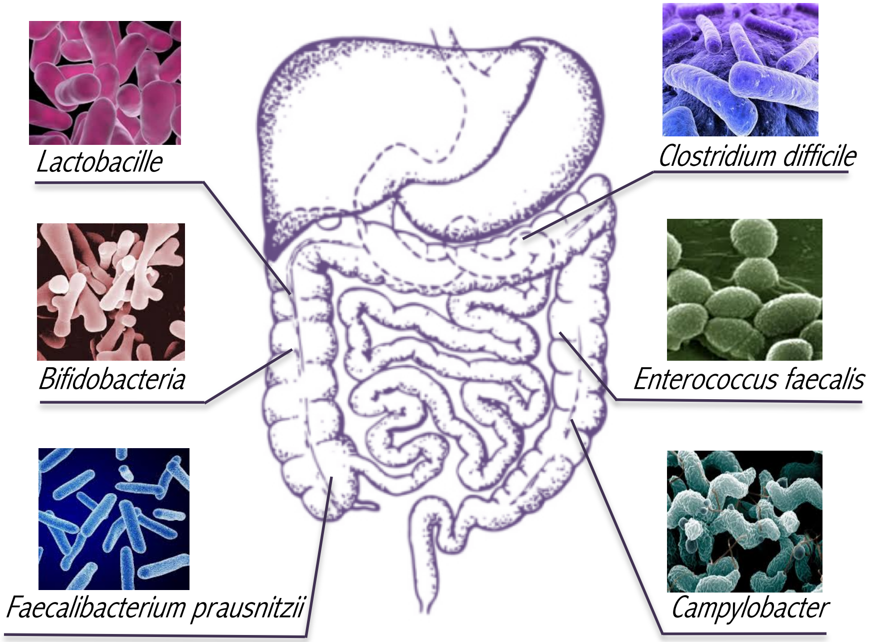
- Conclusions and perspectives

The gut microbiota and our cells: intimate interaction

- 100 000 billions of bacteria
- 4 phyla
-~1000 species

INRA SCIENCE & IMPACT

Pionneer



« Symbiontes»

« Pathobiontes »

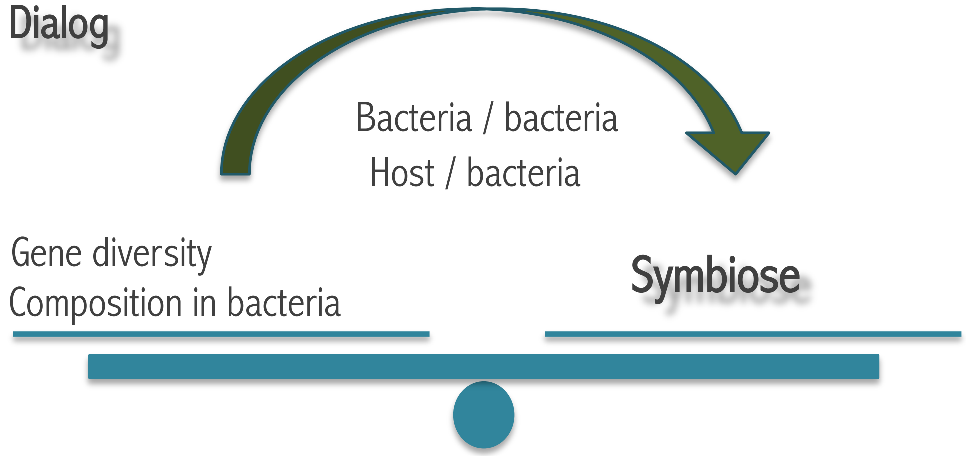
Catalog of 10 millions genes

Universel : core genome

Identity card

Bidirectional dialog: language based on signaling

The gut microbiota and our cells: intimate interaction



Dialog

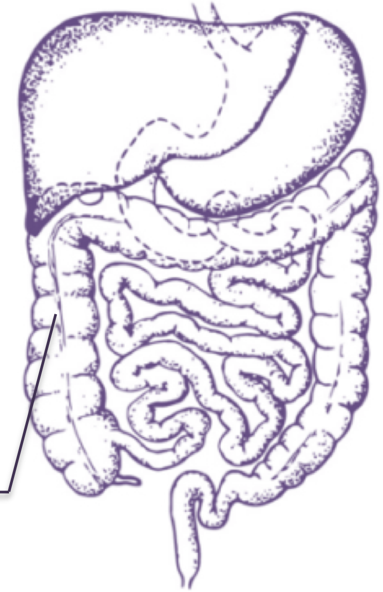
Bacteria / bacteria
Host / bacteria

Gene diversity
Composition in bacteria

Symbiose

Health markers

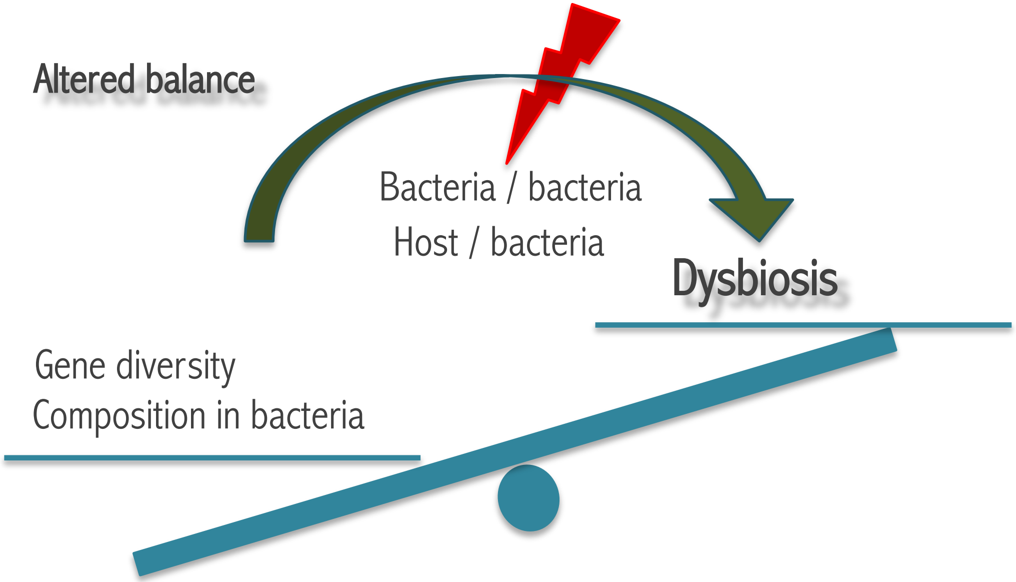
Homeostasis



Gut ecosystem

Key interface cells/food/bacteria

Decipher the molecular mechanisms responsible for this dialog

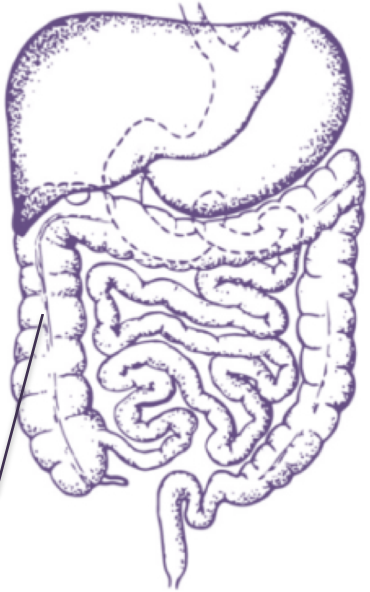


Altered balance

Bacteria / bacteria
Host / bacteria

Dysbiosis

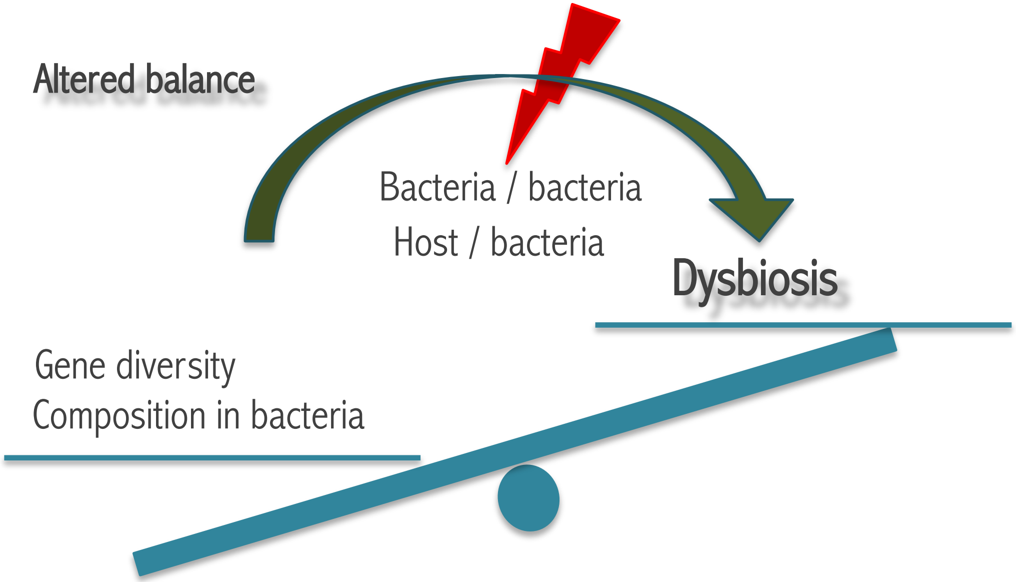
Gene diversity
Composition in bacteria



Gut ecosystem

Gut ecosystem alteration → **Inflammation** → Chronic intestinal bowel diseases

Decipher the molecular mechanisms responsible for this dialog

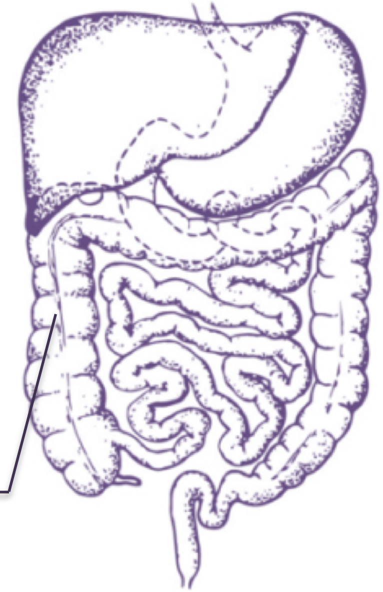


Altered balance

Bacteria / bacteria
Host / bacteria

Dysbiosis

Gene diversity
Composition in bacteria



Gut ecosystem

Gut ecosystem alteration → **Inflammation** → Chronic intestinal bowel diseases

Catalogue of 10 millions of genes

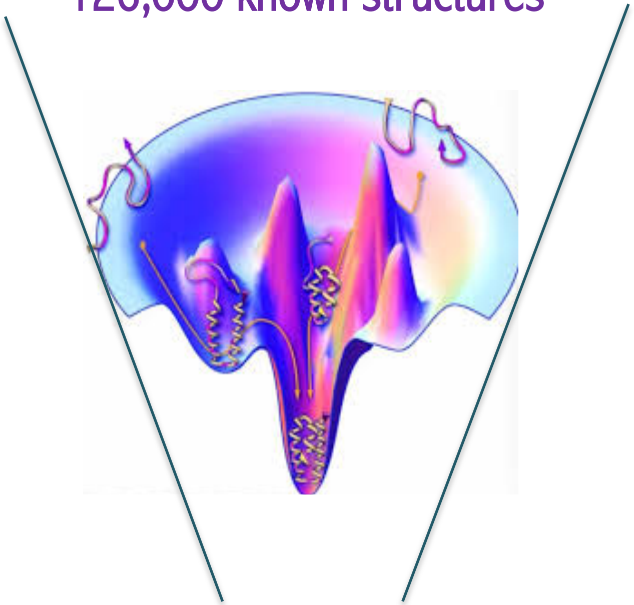
Metadata to explore and target

Investigate the molecular mechanisms of this balance

Investigate MetaOmics through structural bioinformatic

- - More than 10 millions prokaryotic genes
- - 40% protein sequences of unknown function
- - Function can be inferred from structure
- - Assessing the fold is possible
- - Fishing out the right fold is crucial !

126,000 known structures



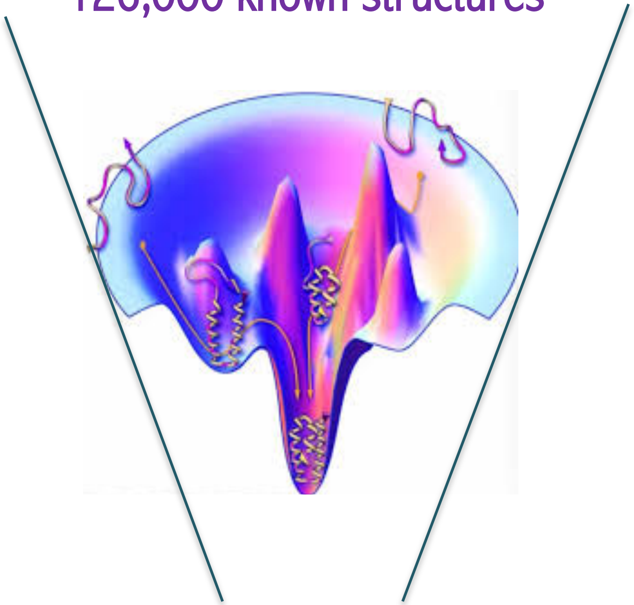
Only 1400 3D folds

Investigate MetaOmics through structural bioinformatic

- - More than 10 millions prokaryotic genes
- - 40% protein sequences of unknown function
- - Function can be inferred from structure
- - Assessing the fold is possible
- - Fishing out the right fold is crucial !

Fold prediction informs on the function

126,000 known structures



Only 1400 3D folds

Investigate MetaOmics through structural bioinformatic

- - More than 10 millions prokaryotic genes
- - 40% protein sequences of unknown function
- - Function can be inferred from structure
- - Assessing the fold is possible
- - Fishing out the right fold is crucial !

Fold prediction informs on the function

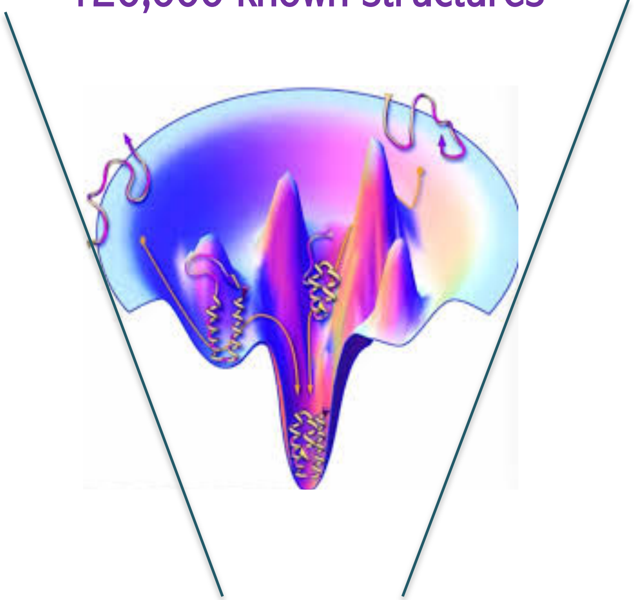
Metagenomic → **Functional metagenomic**

Catalogues of sequences

Fold detection and prediction accurate

Computer compatible with big data

126,000 known structures

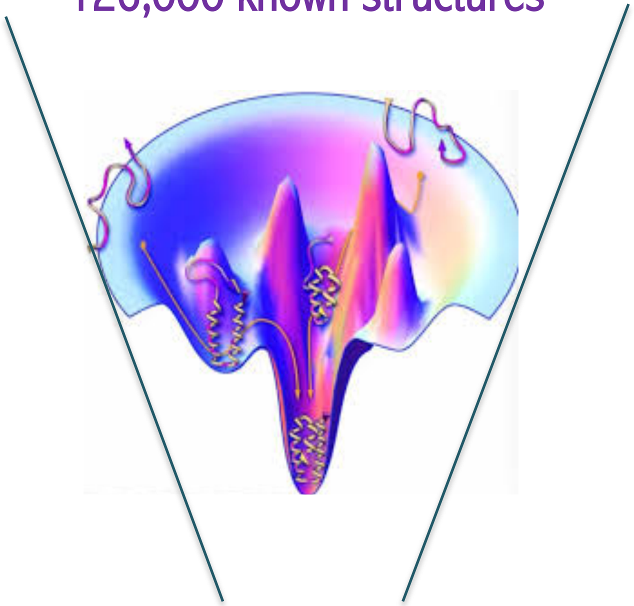


Only 1400 3D folds

Investigate MetaOmics through structural bioinformatic

- - More than 10 millions prokaryotic genes
- - 40% protein sequences of unknown function
- - Function can be inferred from structure
- - Assessing the fold is possible
- - Fishing out the right fold is crucial !

126,000 known structures



Only 1400 3D folds

Fold prediction informs on the function

Metagenomic → Functional metagenomic

Catalogues of sequences

Fold detection and prediction accurate

Computer compatible with big data

Comprehensive detection and 3D modeling = StructurOmics

Strategy

Target protein

Relevance

Scan microbiota

Hits identification

Validation

Biomarker

- Biological property
- Therapeutical interest
- Structural challenge

Strategy

Target protein

Relevance

Scan microbiota

Hits identification

Validation

Biomarker

- Biological property
- Therapeutical interest
- Structural challenge

Preliminary data:

- Histone-like Nucleotide Structuring protein in *Enterococcus faecalis*
- Penicillin binding protein in *Bacillus subtilis*
- Cholesterol oxidase for *Bacteroides dorei* and *dorei D8*

❖ Experimental data in progress

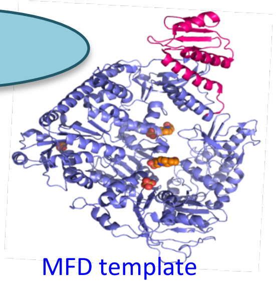
❖ Scaling up to full microbiota requested

MetaFoldScan Pipeline

Target

3D Fold bank

MAM-template



MFD template

HMM profile of the target bank

Query

Genome x_i

Genome x_j

…

Genome x_n

Coarse-grained filtering of sequence candidates.

Multiple Sequence Alignment MSA of filtered candidates.
Secondary structure addition

HMM profile of input MSA

Profile (candidate) *vs* profile (bank) Comparison

Scoring & ranking

Score analysis

Hits prediction

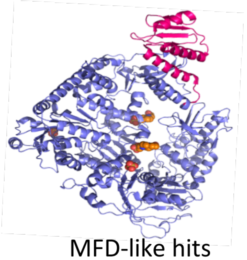
Abundance classification…
Prevalence in MGS

Crystal structure …

In vitro/in vivo testing…

Biomarker ?

+



MAM-like hits

MFD-like hits

Collaborations

Collaborations

Targets of the microbiota

To tackle both methodological and biological issues

MAM

JM. Chatel

Microbial Anti-inflammatory Molecule

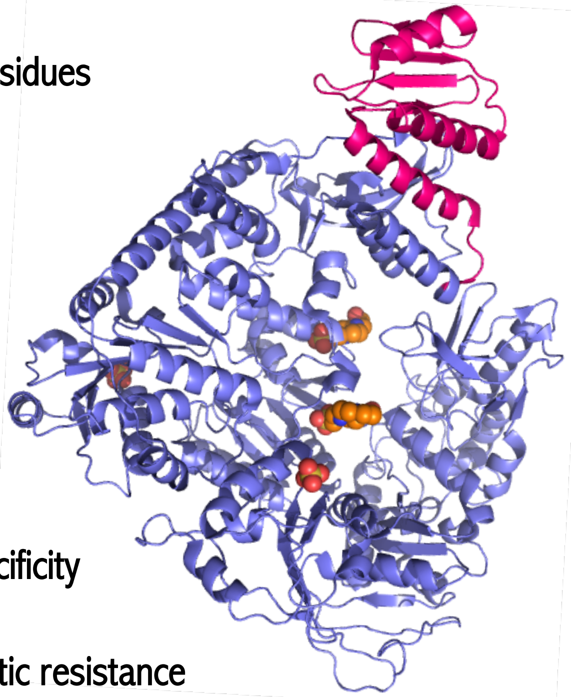
- Unknown biological function
- Small 180 residues
- Structurally not characterized
- Rare, found in *F. prausnitzii*
- F. Prausnitzii 5% of the microbiota, IBD Crohn Disease, Ulcerative Colitis,
- Fishing from homology modeling
- Associated with probiotic properties

MFD

N. Rama Rao

Mutation Frequency Decline

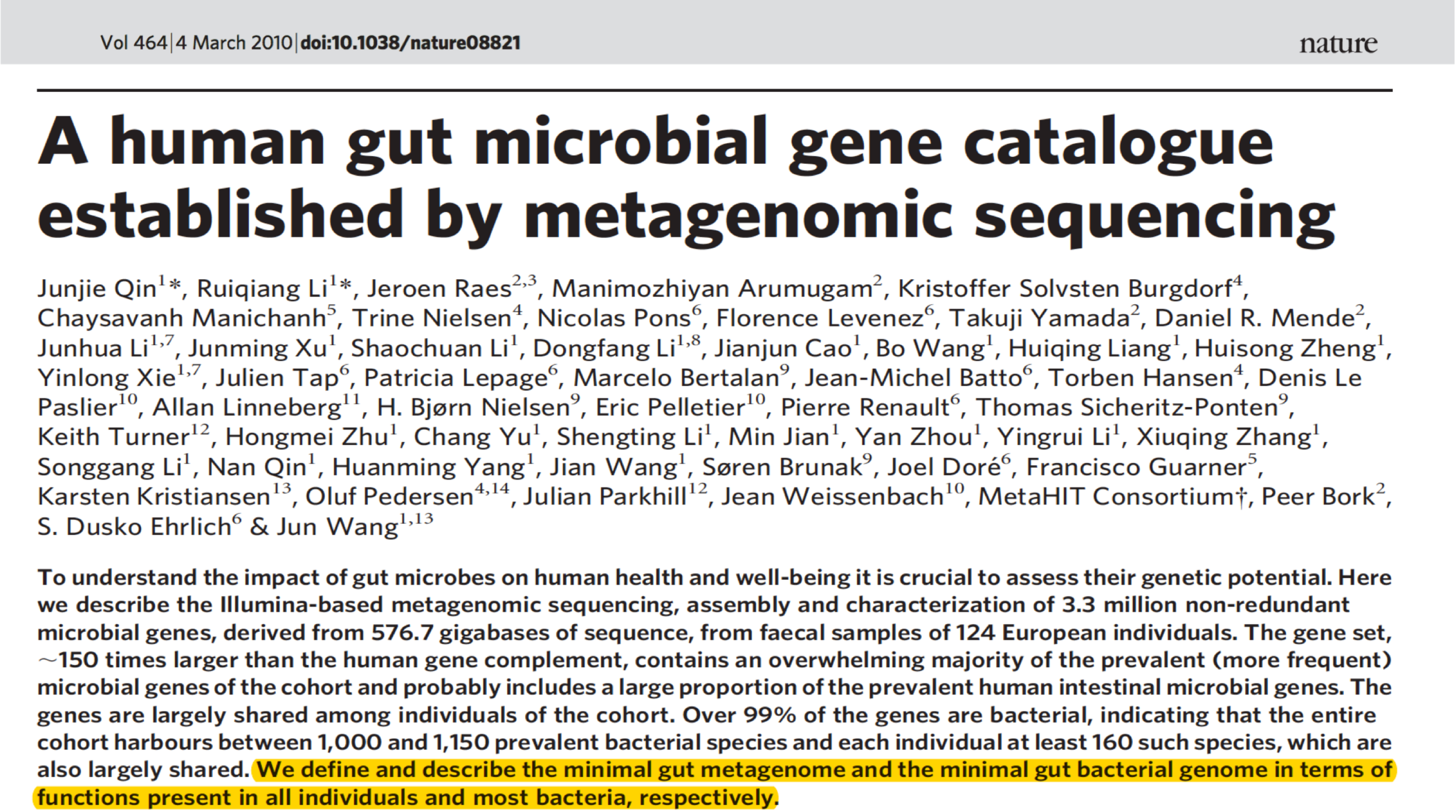
- Involved in bacterial DNA repair, in virulence
- Multi-domains 1150 residues
- Solved structure
- Ubiquitous, Gram+/-
- Promiscuity
- Molecular basis of specificity
- Associated with antibiotic resistance



Before screening MetaHit…

Cohort of 124 European individuals, healthy and obese, danish and spanish

→ Nordic & Mediterranean faecal DNA



Vol 464|4 March 2010|**doi:10.1038/nature08821**

nature

A human gut microbial gene catalogue established by metagenomic sequencing

Junjie Qin[1]*, Ruiqiang Li[1]*, Jeroen Raes[2,3], Manimozhiyan Arumugam[2], Kristoffer Solvsten Burgdorf[4], Chaysavanh Manichanh[5], Trine Nielsen[4], Nicolas Pons[6], Florence Levenez[6], Takuji Yamada[2], Daniel R. Mende[2], Junhua Li[1,7], Junming Xu[1], Shaochuan Li[1], Dongfang Li[1,8], Jianjun Cao[1], Bo Wang[1], Huiqing Liang[1], Huisong Zheng[1], Yinlong Xie[1,7], Julien Tap[6], Patricia Lepage[6], Marcelo Bertalan[9], Jean-Michel Batto[6], Torben Hansen[4], Denis Le Paslier[10], Allan Linneberg[11], H. Bjørn Nielsen[9], Eric Pelletier[10], Pierre Renault[6], Thomas Sicheritz-Ponten[9], Keith Turner[12], Hongmei Zhu[1], Chang Yu[1], Shengting Li[1], Min Jian[1], Yan Zhou[1], Yingrui Li[1], Xiuqing Zhang[1], Songgang Li[1], Nan Qin[1], Huanming Yang[1], Jian Wang[1], Søren Brunak[9], Joel Doré[6], Francisco Guarner[5], Karsten Kristiansen[13], Oluf Pedersen[4,14], Julian Parkhill[12], Jean Weissenbach[10], MetaHIT Consortium†, Peer Bork[2], S. Dusko Ehrlich[6] & Jun Wang[1,13]

To understand the impact of gut microbes on human health and well-being it is crucial to assess their genetic potential. Here we describe the Illumina-based metagenomic sequencing, assembly and characterization of 3.3 million non-redundant microbial genes, derived from 576.7 gigabases of sequence, from faecal samples of 124 European individuals. The gene set, ~150 times larger than the human gene complement, contains an overwhelming majority of the prevalent (more frequent) microbial genes of the cohort and probably includes a large proportion of the prevalent human intestinal microbial genes. The genes are largely shared among individuals of the cohort. Over 99% of the genes are bacterial, indicating that the entire cohort harbours between 1,000 and 1,150 prevalent bacterial species and each individual at least 160 such species, which are also largely shared. We define and describe the minimal gut metagenome and the minimal gut bacterial genome in terms of functions present in all individuals and most bacteria, respectively.

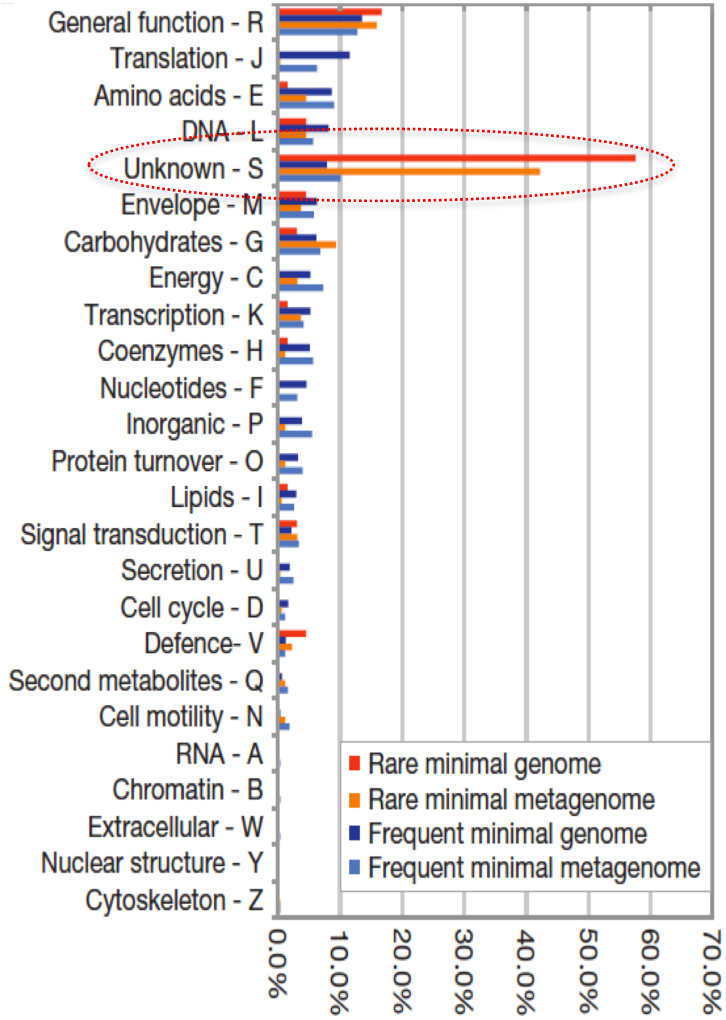
➢ - 3,3 millions of genes , 99,1 % of bacterial origin

➢ - Minimal gut bacterial genome : 1,000 to 1,150 prevalent bacterial species

➢ - Each individual harbors at least 160 such species

Test case : the Gut core genome

Minimal gut genome: functions necessary for a bacterium to thrive in a gut context

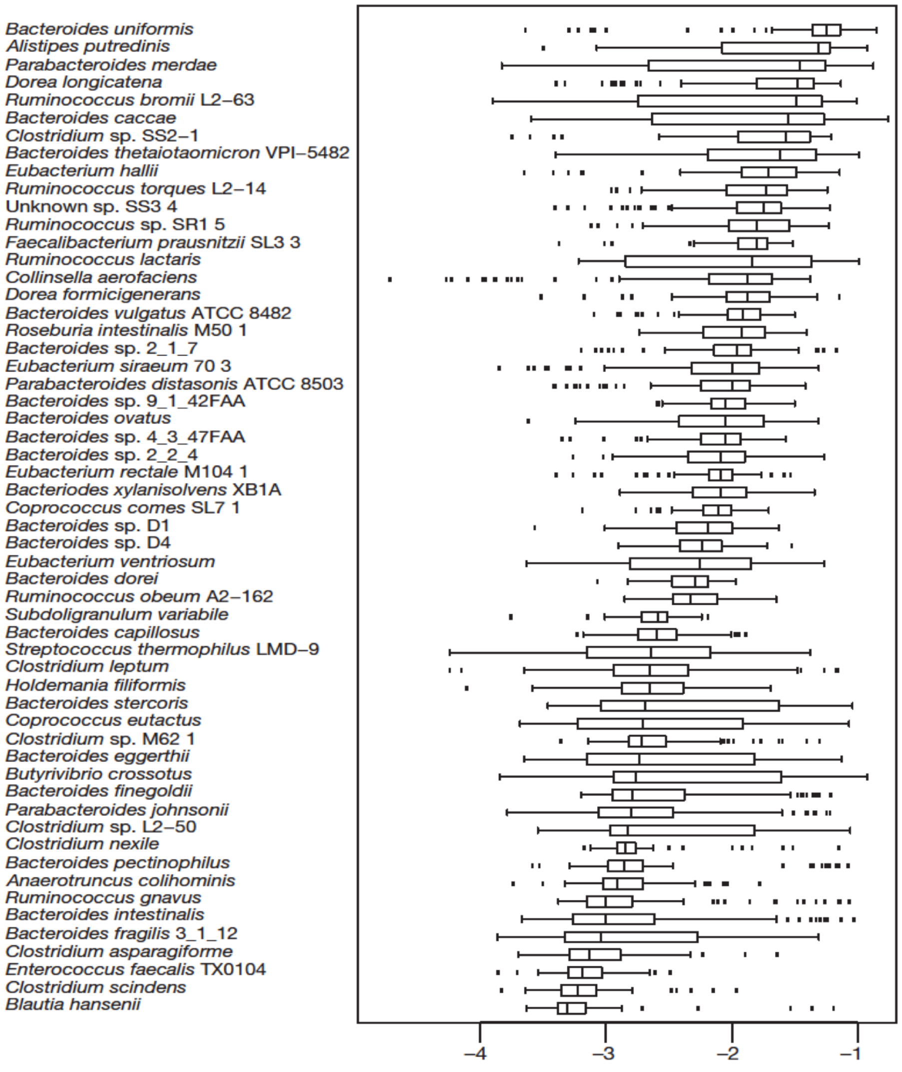


Qin et al, Nature 2010 **Functional composition**

Test case : the Gut core genome

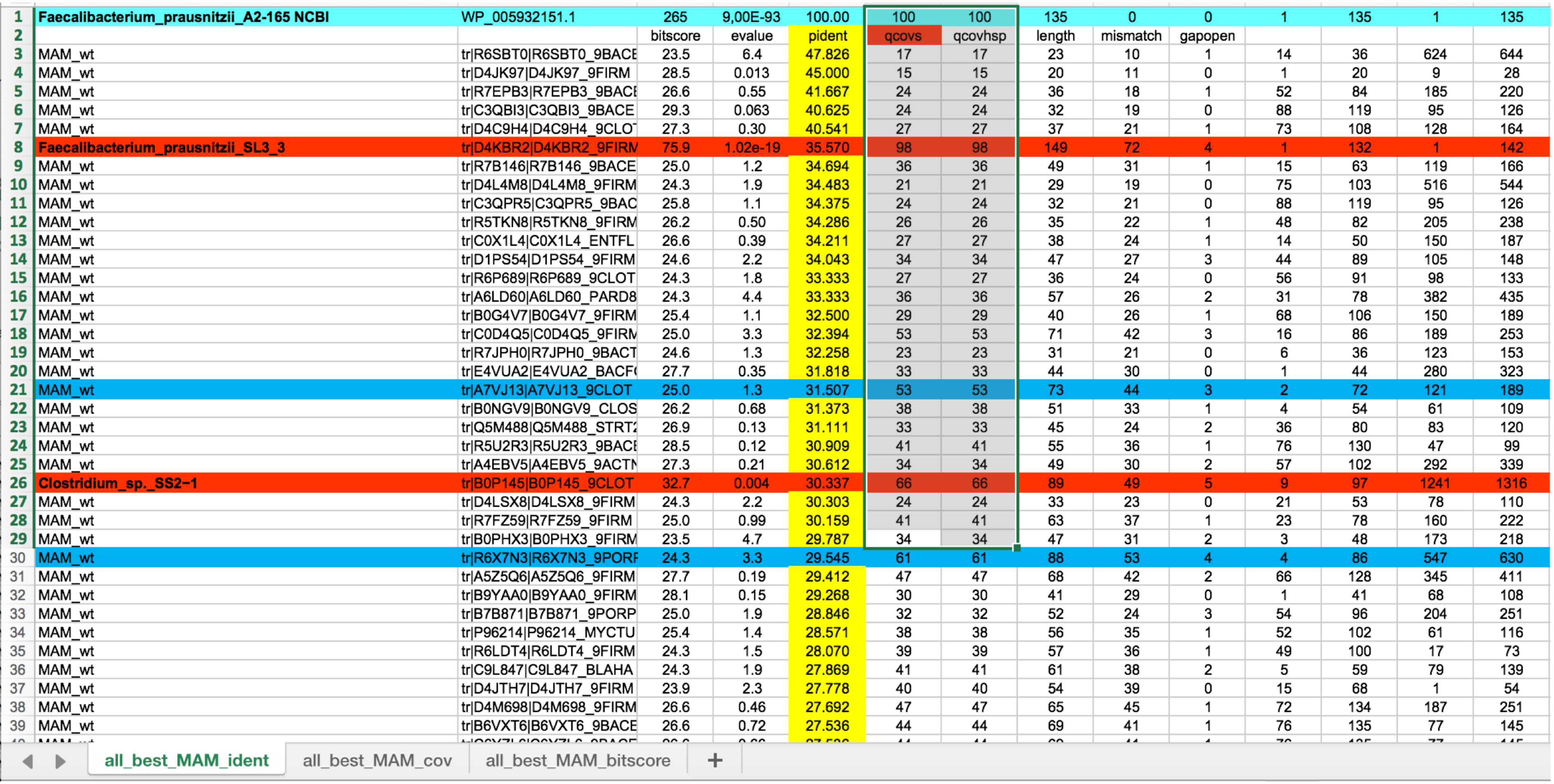
Core genome: 57 bacterial species present in > 90 of individuals (prevalence and consensus)

→ Mostly Bacteroidetes & Firmicutes



Qin et al, Nature 2010

Structural hits of MAM in the core genome



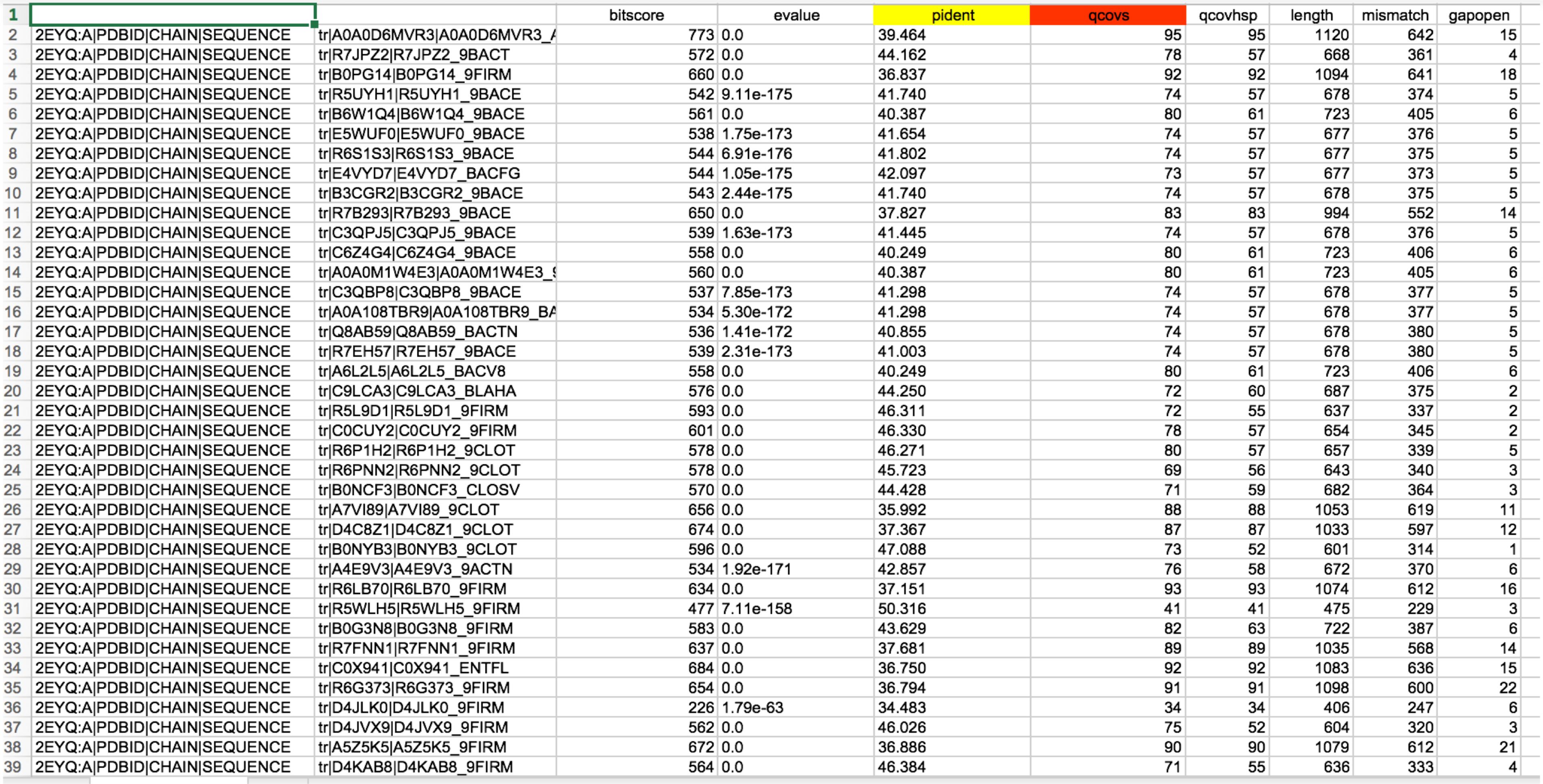
1	**Faecalibacterium_prausnitzii_A2-165 NCBI**	WP_005932151.1	265	9,00E-93	100.00	100	100	135	0	0	1	135	1	135
2			bitscore	evalue	pident	qcovs	qcovhsp	length	mismatch	gapopen				
3	MAM_wt	tr|R6SBT0|R6SBT0_9BACE	23.5	6.4	47.826	17	17	23	10	1	14	36	624	644
4	MAM_wt	tr|D4JK97|D4JK97_9FIRM	28.5	0.013	45.000	15	15	20	11	0	1	20	9	28
5	MAM_wt	tr|R7EPB3|R7EPB3_9BACE	26.6	0.55	41.667	24	24	36	18	1	52	84	185	220
6	MAM_wt	tr|C3QBI3|C3QBI3_9BACE	29.3	0.063	40.625	24	24	32	19	0	88	119	95	126
7	MAM_wt	tr|D4C9H4|D4C9H4_9CLOT	27.3	0.30	40.541	27	27	37	21	1	73	108	128	164
8	**Faecalibacterium_prausnitzii_SL3_3**	tr|D4KBR2|D4KBR2_9FIRM	75.9	1.02e-19	35.570	98	98	149	72	4	1	132	1	142
9	MAM_wt	tr|R7B146|R7B146_9BACE	25.0	1.2	34.694	36	36	49	31	1	15	63	119	166
10	MAM_wt	tr|D4L4M8|D4L4M8_9FIRM	24.3	1.9	34.483	21	21	29	19	0	75	103	516	544
11	MAM_wt	tr|C3QPR5|C3QPR5_9BAC	25.8	1.1	34.375	24	24	32	21	0	88	119	95	126
12	MAM_wt	tr|R5TKN8|R5TKN8_9FIRM	26.2	0.50	34.286	26	26	35	22	1	48	82	205	238
13	MAM_wt	tr|C0X1L4|C0X1L4_ENTFL	26.6	0.39	34.211	27	27	38	24	1	14	50	150	187
14	MAM_wt	tr|D1PS54|D1PS54_9FIRM	24.6	2.2	34.043	34	34	47	27	3	44	89	105	148
15	MAM_wt	tr|R6P689|R6P689_9CLOT	24.3	1.8	33.333	27	27	36	24	0	56	91	98	133
16	MAM_wt	tr|A6LD60|A6LD60_PARD8	24.3	4.4	33.333	36	36	57	26	2	31	78	382	435
17	MAM_wt	tr|B0G4V7|B0G4V7_9FIRM	25.4	1.1	32.500	29	29	40	26	1	68	106	150	189
18	MAM_wt	tr|C0D4Q5|C0D4Q5_9FIRM	25.0	3.3	32.394	53	53	71	42	3	16	86	189	253
19	MAM_wt	tr|R7JPH0|R7JPH0_9BACT	24.6	1.3	32.258	23	23	31	21	0	6	36	123	153
20	MAM_wt	tr|E4VUA2|E4VUA2_BACF	27.7	0.35	31.818	33	33	44	30	0	1	44	280	323
21	MAM_wt	tr|A7VJ13|A7VJ13_9CLOT	25.0	1.3	31.507	53	53	73	44	3	2	72	121	189
22	MAM_wt	tr|B0NGV9|B0NGV9_CLOS	26.2	0.68	31.373	38	38	51	33	1	4	54	61	109
23	MAM_wt	tr|Q5M488|Q5M488_STRT	26.9	0.13	31.111	33	33	45	24	2	36	80	83	120
24	MAM_wt	tr|R5U2R3|R5U2R3_9BACE	28.5	0.12	30.909	41	41	55	36	1	76	130	47	99
25	MAM_wt	tr|A4EBV5|A4EBV5_9ACTN	27.3	0.21	30.612	34	34	49	30	2	57	102	292	339
26	**Clostridium_sp._SS2−1**	tr|B0P145|B0P145_9CLOT	32.7	0.004	30.337	66	66	89	49	5	9	97	1241	1316
27	MAM_wt	tr|D4LSX8|D4LSX8_9FIRM	24.3	2.2	30.303	24	24	33	23	0	21	53	78	110
28	MAM_wt	tr|R7FZ59|R7FZ59_9FIRM	25.0	0.99	30.159	41	41	63	37	1	23	78	160	222
29	MAM_wt	tr|B0PHX3|B0PHX3_9FIRM	23.5	4.7	29.787	34	34	47	31	2	3	48	173	218
30	MAM_wt	tr|R6X7N3|R6X7N3_9PORF	24.3	3.3	29.545	61	61	88	53	4	4	86	547	630
31	MAM_wt	tr|A5Z5Q6|A5Z5Q6_9FIRM	27.7	0.19	29.412	47	47	68	42	2	66	128	345	411
32	MAM_wt	tr|B9YAA0|B9YAA0_9FIRM	28.1	0.15	29.268	30	30	41	29	0	1	41	68	108
33	MAM_wt	tr|B7B871|B7B871_9PORP	25.0	1.9	28.846	32	32	52	24	3	54	96	204	251
34	MAM_wt	tr|P96214|P96214_MYCTU	25.4	1.4	28.571	38	38	56	35	1	52	102	61	116
35	MAM_wt	tr|R6LDT4|R6LDT4_9FIRM	24.3	1.5	28.070	39	39	57	36	1	49	100	17	73
36	MAM_wt	tr|C9L847|C9L847_BLAHA	24.3	1.9	27.869	41	41	61	38	2	5	59	79	139
37	MAM_wt	tr|D4JTH7|D4JTH7_9FIRM	23.9	2.3	27.778	40	40	54	39	0	15	68	1	54
38	MAM_wt	tr|D4M698|D4M698_9FIRM	26.6	0.46	27.692	47	47	65	45	1	72	134	187	251
39	MAM_wt	tr|B6VXT6|B6VXT6_9BACE	26.6	0.72	27.536	44	44	69	41	1	76	135	77	145

Analysis

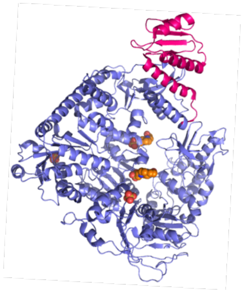
→ No other hit except in *Faecalibacterium Prausnitzii SL3_3*

→ MAM strictly restricted to a niche of *F. Praunitzii species*

Test case : the Gut core genome



1			bitscore	evalue	pident	qcovs	qcovhsp	length	mismatch	gapopen
2	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|A0A0D6MVR3|A0A0D6MVR3_	773	0.0	39.464	95	95	1120	642	15
3	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R7JPZ2|R7JPZ2_9BACT	572	0.0	44.162	78	57	668	361	4
4	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|B0PG14|B0PG14_9FIRM	660	0.0	36.837	92	92	1094	641	18
5	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R5UYH1|R5UYH1_9BACE	542	9.11e-175	41.740	74	57	678	374	5
6	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|B6W1Q4|B6W1Q4_9BACE	561	0.0	40.387	80	61	723	405	6
7	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|E5WUF0|E5WUF0_9BACE	538	1.75e-173	41.654	74	57	677	376	5
8	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R6S1S3|R6S1S3_9BACE	544	6.91e-176	41.802	74	57	677	375	5
9	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|E4VYD7|E4VYD7_BACFG	544	1.05e-175	42.097	73	57	677	373	5
10	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|B3CGR2|B3CGR2_9BACE	543	2.44e-175	41.740	74	57	678	375	5
11	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R7B293|R7B293_9BACE	650	0.0	37.827	83	83	994	552	14
12	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|C3QPJ5|C3QPJ5_9BACE	539	1.63e-173	41.445	74	57	678	376	5
13	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|C6Z4G4|C6Z4G4_9BACE	558	0.0	40.249	80	61	723	406	6
14	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|A0A0M1W4E3|A0A0M1W4E3_	560	0.0	40.387	80	61	723	405	6
15	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|C3QBP8|C3QBP8_9BACE	537	7.85e-173	41.298	74	57	678	377	5
16	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|A0A108TBR9|A0A108TBR9_BA	534	5.30e-172	41.298	74	57	678	377	5
17	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|Q8AB59|Q8AB59_BACTN	536	1.41e-172	40.855	74	57	678	380	5
18	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R7EH57|R7EH57_9BACE	539	2.31e-173	41.003	74	57	678	380	5
19	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|A6L2L5|A6L2L5_BACV8	558	0.0	40.249	80	61	723	406	6
20	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|C9LCA3|C9LCA3_BLAHA	576	0.0	44.250	72	60	687	375	2
21	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R5L9D1|R5L9D1_9FIRM	593	0.0	46.311	72	55	637	337	2
22	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|C0CUY2|C0CUY2_9FIRM	601	0.0	46.330	78	57	654	345	2
23	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R6P1H2|R6P1H2_9CLOT	578	0.0	46.271	80	57	657	339	5
24	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R6PNN2|R6PNN2_9CLOT	578	0.0	45.723	69	56	643	340	3
25	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|B0NCF3|B0NCF3_CLOSV	570	0.0	44.428	71	59	682	364	3
26	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|A7VI89|A7VI89_9CLOT	656	0.0	35.992	88	88	1053	619	11
27	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|D4C8Z1|D4C8Z1_9CLOT	674	0.0	37.367	87	87	1033	597	12
28	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|B0NYB3|B0NYB3_9CLOT	596	0.0	47.088	73	52	601	314	1
29	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|A4E9V3|A4E9V3_9ACTN	534	1.92e-171	42.857	76	58	672	370	6
30	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R6LB70|R6LB70_9FIRM	634	0.0	37.151	93	93	1074	612	16
31	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R5WLH5|R5WLH5_9FIRM	477	7.11e-158	50.316	41	41	475	229	3
32	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|B0G3N8|B0G3N8_9FIRM	583	0.0	43.629	82	63	722	387	6
33	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R7FNN1|R7FNN1_9FIRM	637	0.0	37.681	89	89	1035	568	14
34	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|C0X941|C0X941_ENTFL	684	0.0	36.750	92	92	1083	636	15
35	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|R6G373|R6G373_9FIRM	654	0.0	36.794	91	91	1098	600	22
36	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|D4JLK0|D4JLK0_9FIRM	226	1.79e-63	34.483	34	34	406	247	6
37	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|D4JVX9|D4JVX9_9FIRM	562	0.0	46.026	75	52	604	320	3
38	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|A5Z5K5|A5Z5K5_9FIRM	672	0.0	36.886	90	90	1079	612	21
39	2EYQ:A|PDBID|CHAIN|SEQUENCE	tr|D4KAB8|D4KAB8_9FIRM	564	0.0	46.384	71	55	636	333	4



Post-analysis

- Mfd ubiquitous as expected
- Modeling of Mfd complex functional cycle that associates distinct functions with conformations
- Focus on B. cereus strains clinical and not to identify some mark of pathogeny-NO stress

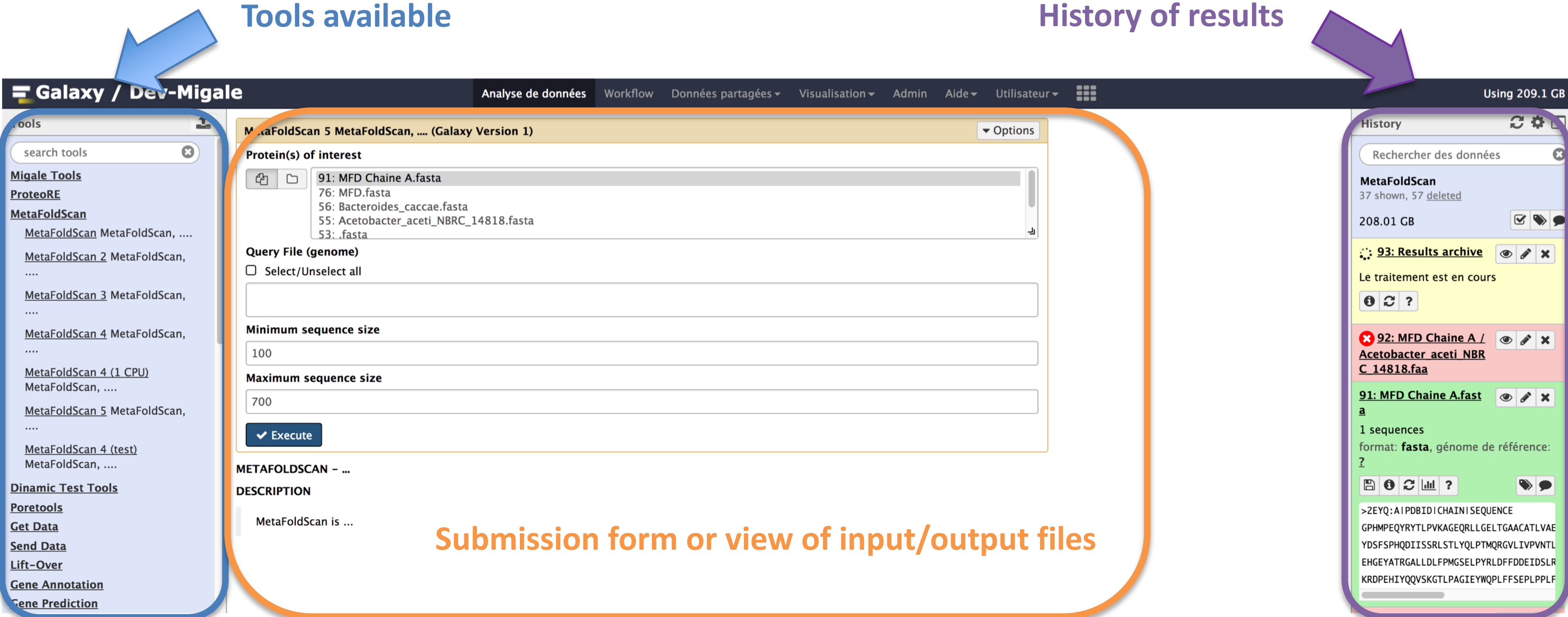
MetaFoldScan implementation into Galaxy

S. Derozier

Galaxy is a scientific workflow, data integration and data analysis platform to render computational biology accessible to research scientists who do not have computer programming or system administration knowledge.

Galaxy is an open, web-based platform:

- **Accessible:** users without programming experience can easily specify parameters and run tools and workflows.
- **Reproducible:** galaxy captures information so that any user can reproduce a complete computational analysis.
- **Transparent:** users share and publish analyses *via* the web, compute interactive, web-based a complete analysis.



Migale Tools
ProteoRE
MetaFoldScan
MetaFoldScan MetaFoldScan, ...
MetaFoldScan 2 MetaFoldScan, ...
MetaFoldScan 3 MetaFoldScan, ...
MetaFoldScan 4 MetaFoldScan, ...
MetaFoldScan 4 (1 CPU) MetaFoldScan, ...
MetaFoldScan 5 MetaFoldScan, ...
MetaFoldScan 4 (test) MetaFoldScan, ...
Dinamic Test Tools
Poretools
Get Data
Send Data
Lift-Over
Gene Annotation
Gene Prediction

MetaFoldScan 5 MetaFoldScan, ... (Galaxy Version 1)

Protein(s) of interest
91: MFD Chaine A.fasta
76: MFD.fasta
56: Bacteroides_caccae.fasta
55: Acetobacter_aceti_NBRC_14818.fasta
53: .fasta

Query File (genome)
☐ Select/Unselect all

Minimum sequence size
100

Maximum sequence size
700

METAFOLDSCAN – ...

DESCRIPTION

MetaFoldScan is ...

MetaFoldScan
37 shown, 57 deleted
208.01 GB

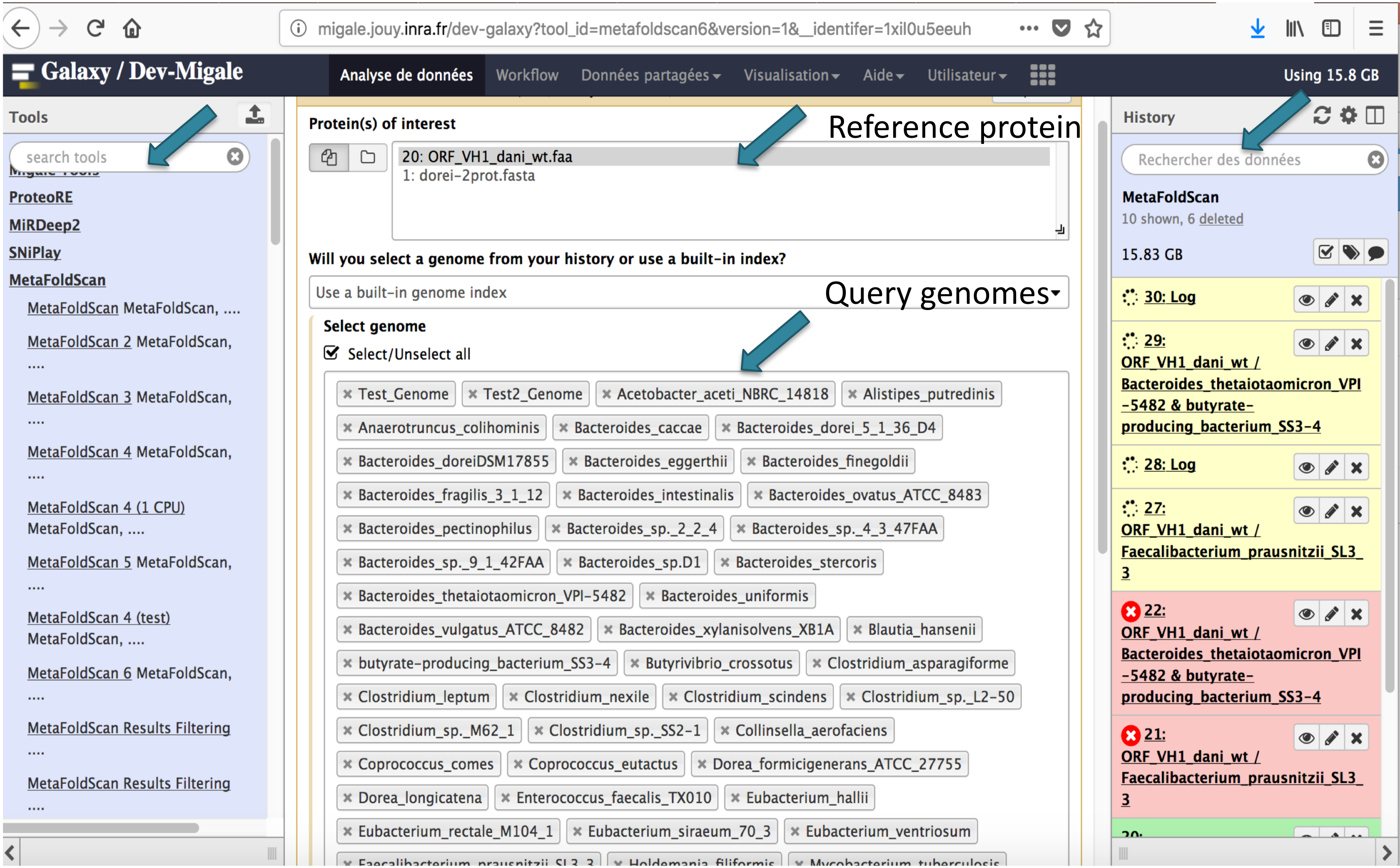
93: Results archive
Le traitement est en cours

92: MFD Chaine A / Acetobacter_aceti_NBRC_14818.faa

91: MFD Chaine A.fasta
1 sequences
format: **fasta**, génome de référence: ?

```
>2EYQ:A|PDBID|CHAIN|SEQUENCE
GPHMPEQYRYTLPVKAGEQRLLGELTGAACATLVAE
YDSFSPHQDIISSRLSTLYQLPTMQRGVLIVPVNTL
EHGEYATRGALLDLFPMGSELPYRLDFFDDEIDSLR
KRDPEHIYQQVSKGTLPAGIEYWQPLFFSEPLPPLF
```

MetaFoldScan implemented in Galaxy



Reference protein

Query genomes

Tools

- ProteoRE
- MiRDeep2
- SNiPlay
- MetaFoldScan
  - MetaFoldScan MetaFoldScan, ...
  - MetaFoldScan 2 MetaFoldScan, ...
  - MetaFoldScan 3 MetaFoldScan, ...
  - MetaFoldScan 4 MetaFoldScan, ...
  - MetaFoldScan 4 (1 CPU) MetaFoldScan, ...
  - MetaFoldScan 5 MetaFoldScan, ...
  - MetaFoldScan 4 (test) MetaFoldScan, ...
  - MetaFoldScan 6 MetaFoldScan, ...
  - MetaFoldScan Results Filtering ....
  - MetaFoldScan Results Filtering ....

Protein(s) of interest

20: ORF_VH1_dani_wt.faa
1: dorei-2prot.fasta

Will you select a genome from your history or use a built-in index?

Use a built-in genome index

Select genome

☑ Select/Unselect all

× Test_Genome × Test2_Genome × Acetobacter_aceti_NBRC_14818 × Alistipes_putredinis × Anaerotruncus_colihominis × Bacteroides_caccae × Bacteroides_dorei_5_1_36_D4 × Bacteroides_doreiDSM17855 × Bacteroides_eggerthii × Bacteroides_finegoldii × Bacteroides_fragilis_3_1_12 × Bacteroides_intestinalis × Bacteroides_ovatus_ATCC_8483 × Bacteroides_pectinophilus × Bacteroides_sp._2_2_4 × Bacteroides_sp._4_3_47FAA × Bacteroides_sp._9_1_42FAA × Bacteroides_sp.D1 × Bacteroides_stercoris × Bacteroides_thetaiotaomicron_VPI-5482 × Bacteroides_uniformis × Bacteroides_vulgatus_ATCC_8482 × Bacteroides_xylanisolvens_XB1A × Blautia_hansenii × butyrate-producing_bacterium_SS3-4 × Butyrivibrio_crossotus × Clostridium_asparagiforme × Clostridium_leptum × Clostridium_nexile × Clostridium_scindens × Clostridium_sp._L2-50 × Clostridium_sp._M62_1 × Clostridium_sp._SS2-1 × Collinsella_aerofaciens × Coprococcus_comes × Coprococcus_eutactus × Dorea_formicigenerans_ATCC_27755 × Dorea_longicatena × Enterococcus_faecalis_TX010 × Eubacterium_hallii × Eubacterium_rectale_M104_1 × Eubacterium_siraeum_70_3 × Eubacterium_ventriosum × Faecalibacterium_prausnitzii_SL3_3 × Holdemania_filiformis × Mycobacterium_tuberculosis

History

MetaFoldScan

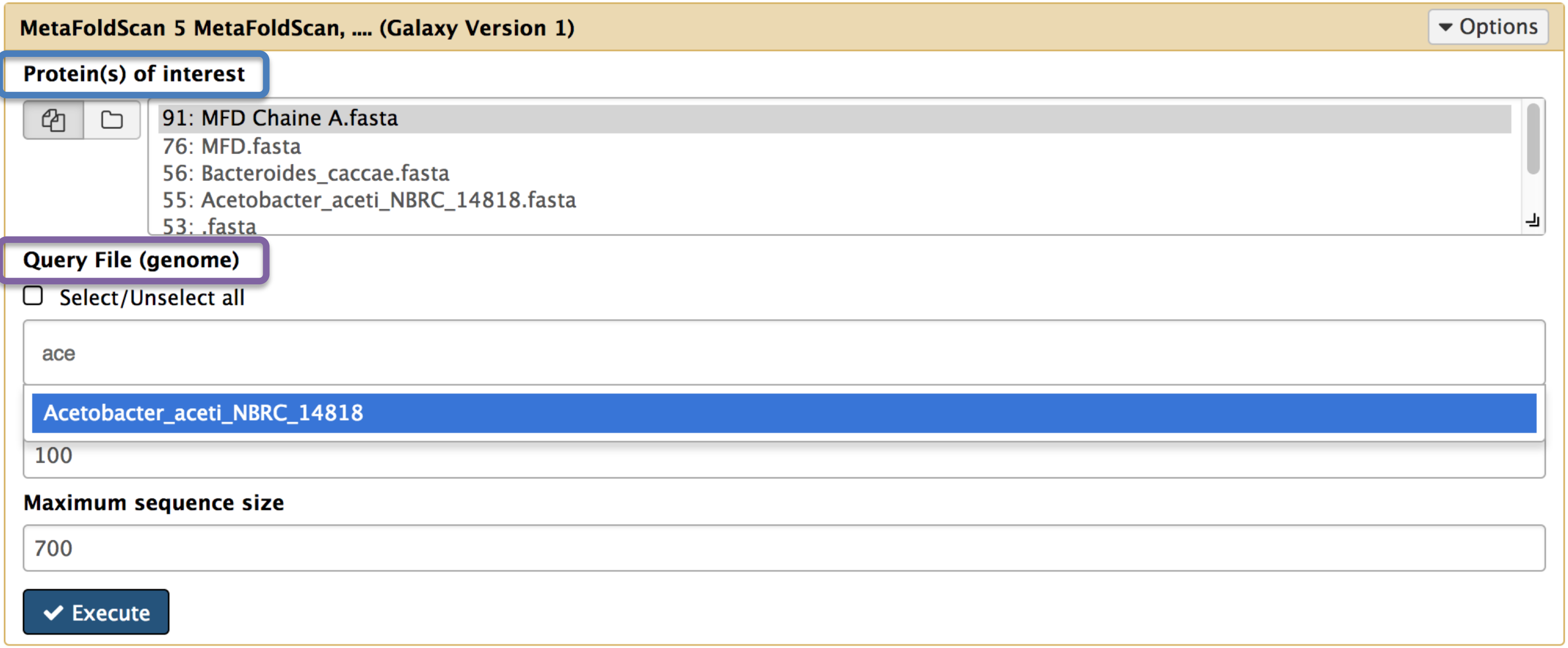
10 shown, 6 deleted

15.83 GB

- 30: Log
- 29: ORF_VH1_dani_wt / Bacteroides_thetaiotaomicron_VPI-5482 & butyrate-producing_bacterium_SS3-4
- 28: Log
- 27: ORF_VH1_dani_wt / Faecalibacterium_prausnitzii_SL3_3
- 22: ORF_VH1_dani_wt / Bacteroides_thetaiotaomicron_VPI-5482 & butyrate-producing_bacterium_SS3-4
- 21: ORF_VH1_dani_wt / Faecalibacterium_prausnitzii_SL3_3

MetaFoldScan implementation into Galaxy

S. Derozier



MetaFoldScan 5 MetaFoldScan, …. (Galaxy Version 1)

Protein(s) of interest

91: MFD Chaine A.fasta
76: MFD.fasta
56: Bacteroides_caccae.fasta
55: Acetobacter_aceti_NBRC_14818.fasta
53: .fasta

Query File (genome)

☐ Select/Unselect all

ace

Acetobacter_aceti_NBRC_14818

100

Maximum sequence size

700

METAFOLDSCAN – …

DESCRIPTION

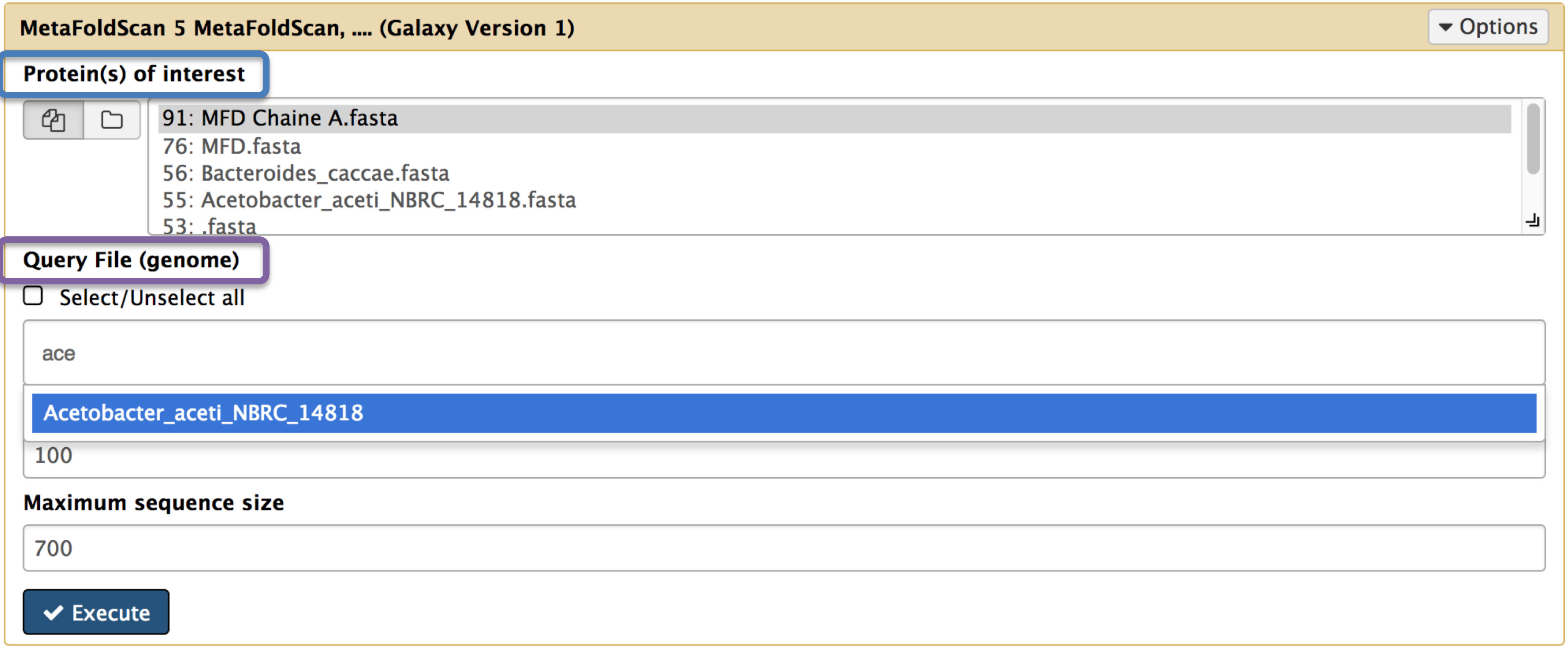
Work in progress

Promising as a web tool to screen a couple of genomes or more

Associates accuracy and autonomy

MetaFoldScan implementation into Galaxy

S. Derozier



MetaFoldScan 5 MetaFoldScan, ... (Galaxy Version 1)

Protein(s) of interest

91: MFD Chaine A.fasta
76: MFD.fasta
56: Bacteroides_caccae.fasta
55: Acetobacter_aceti_NBRC_14818.fasta
53: .fasta

Query File (genome)

☐ Select/Unselect all

ace

Acetobacter_aceti_NBRC_14818

100

Maximum sequence size

700

METAFOLDSCAN – ...

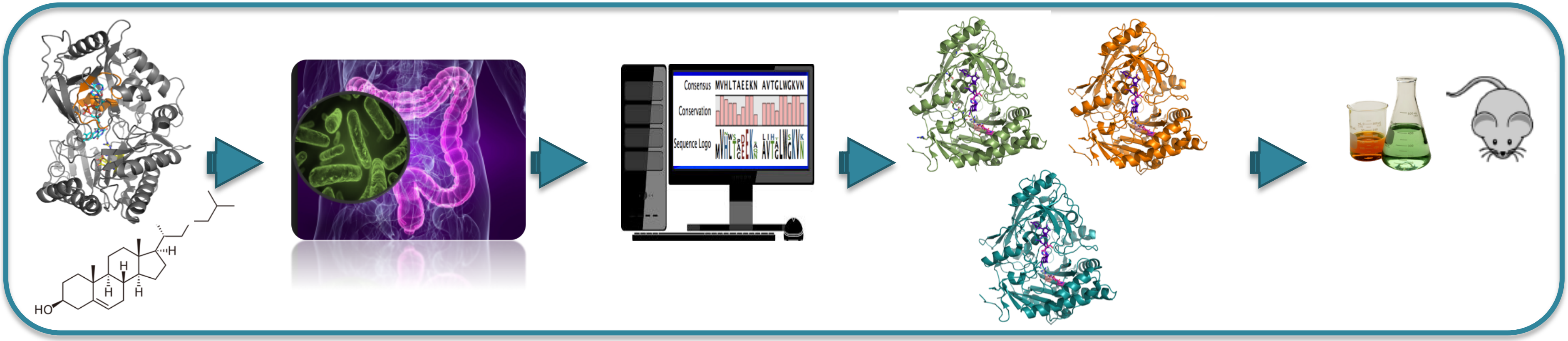
DESCRIPTION

Work in progress

Main hurdle : scaling up from core genome to 10 millions of genes

→ **Filtering**: size (number of residues), signal peptide, secretome, etc.
→ **Data storage**, including the temporary data generated crash after 22 genomes and 2 months of calculations

MetaFoldScan scaling up



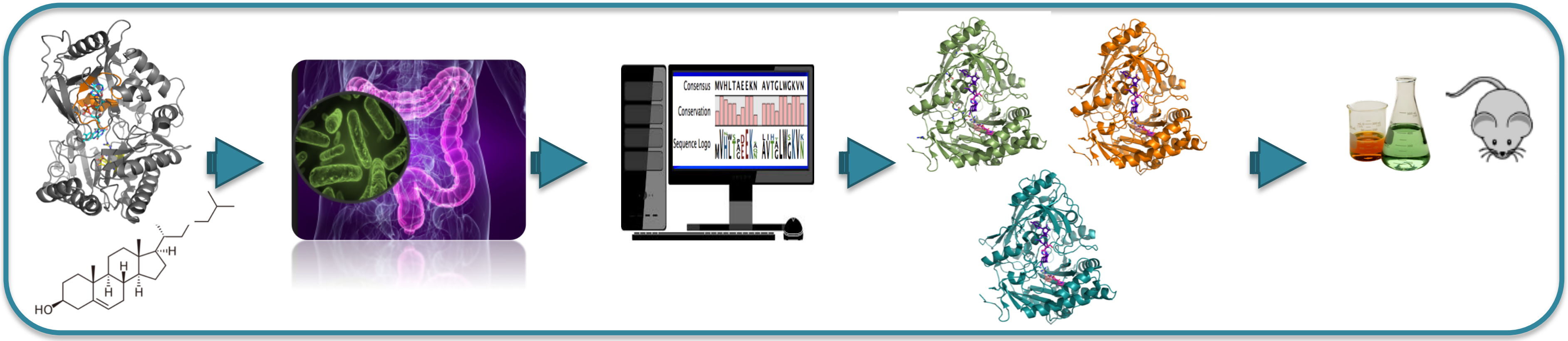
- Collaboration with B. Offmann –University of Nantes, France, for FORSA
- Structural annotation of protein sequences: from Hidden Markov Model profiling to alphabet protein
- FORSA scores the compatibility of any amino acid sequence with any fold represented in the form of protein blocks sequences (1D representation of 3D information).
- FORSA is highly scalable to large datasets
- FORSA uses both local and global threading through parallelising large amount of analyses.

→ One genome one protein one hour on a laptop

↓

Deployable, scalable to 1,000 genomes and more, early 2019

Towards the Structural Screening of Microbial Ecosystems



Outline

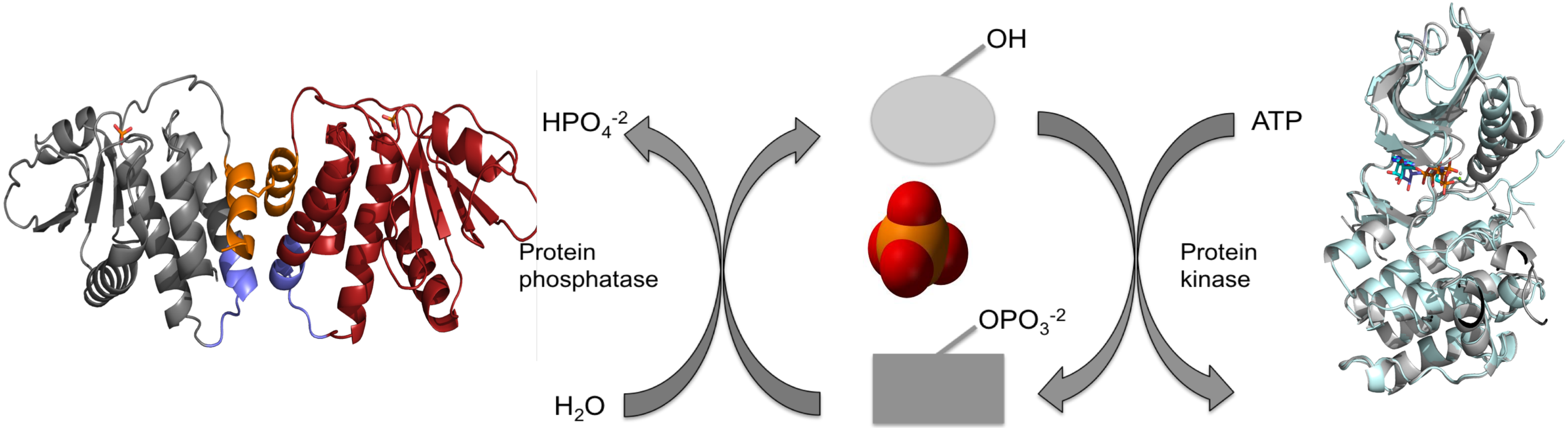
- Cholesterol conversion in the gut microbiota: the mystery enzyme(s) ?

- MetaFoldScan project : 3D screening of the gut microbiota

- **Conclusions and perspectives**

Collaborative projects: actors of the dialog gut /host

Cell signaling proteins → Responsible for the dialog bacteria/bacteria et host/bacteria

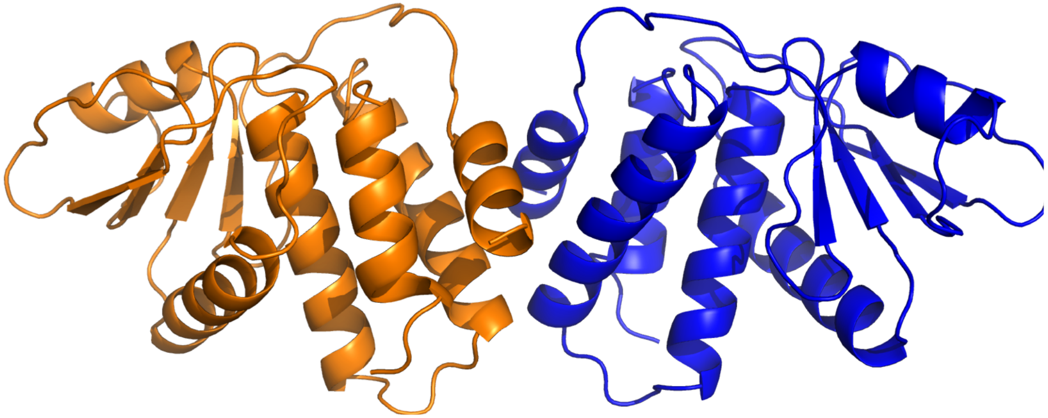


Dialog between in and outside: Tyr-phosphatase & kinases amplify a signal from the membrane

Collaborative project 1: identify and characterize Tyr-phosphatases

Dr. A. Villarino & Dr. M. Berois, Universitad de Montevideo, Uruguay

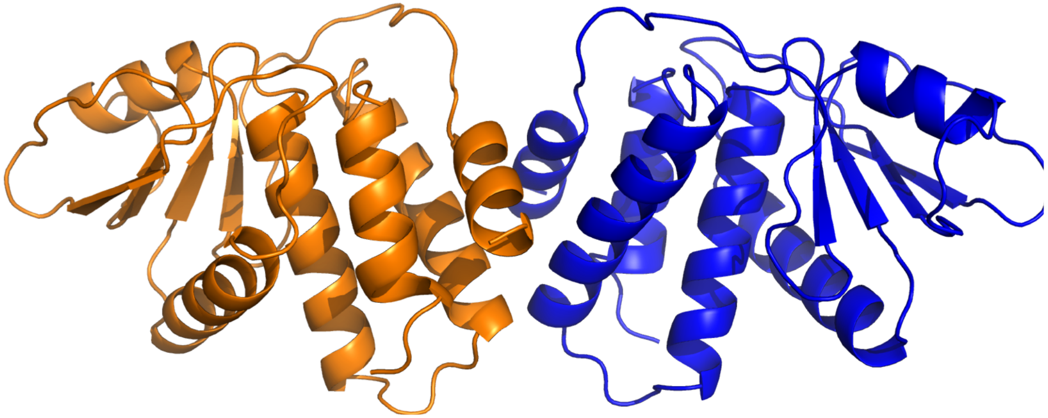
Tyr-Phosphatase



Collaborative project 1: identify and characterize Tyr-phosphatases

Dr. A. Villarino & Dr. M. Berois, Universitad de Montevideo, Uruguay

Tyr-Phosphatase



Intracellulaire pathogen ORF

- Actif as a dimer
- Hijacks the immune system
- Dephosphorylates an essential cell mediator cellulaire essentiel

2 CR1, 1AI, 2 Master

MaIAGE, Migale

2 PR, 2 Master

Facultad de Ciencas, Montevideo, Uruguay

Ecos-Sud Uruguay

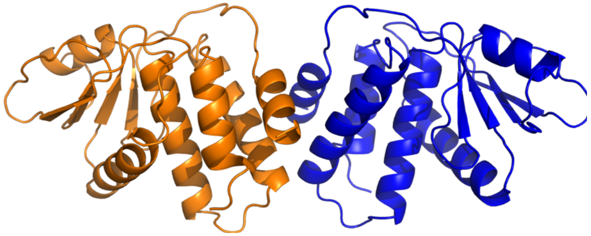
grants

Screen the gut microbiota to target Tyr-phosphatases

Dr. A. Villarino & Dr. M. Berois, Universitad de Montevideo, Uruguay

Antigene in the new Crohn detection kit: Tyr-phosphatase PtpA

Mycobacterium avium ssp paratuberculosis



viral Tyr-phosphatase

Bacteria

?

- Conservation
- Distribution
- Structure/fonction

Structural homologs of Tyr-phosphatases

1DR2,1 CR1, 1AI, 2 Master

MaIAGE, Migale

2 PR, 2 Master

Faculté des Sciences, Montevideo, Uruguay

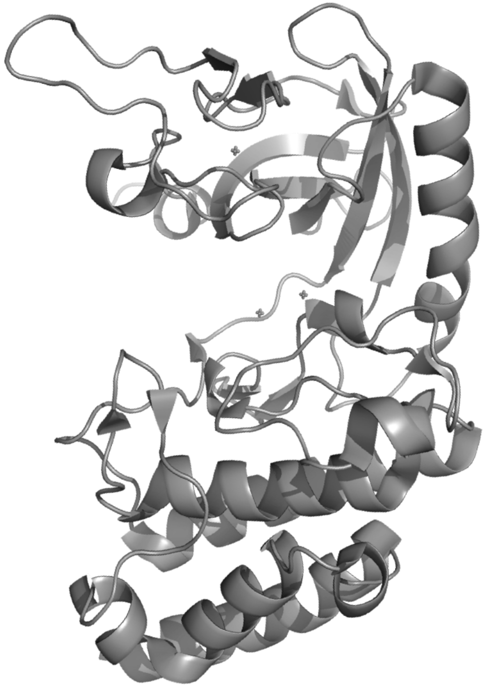
Ecos-Sud Uruguay

Grants

Collaborative project 2: identify a pathway involved in metabolism

Dr. M.N. Lisa, IBR-Conicet, Rosario Argentina

Ser/Thr Protein Kinase



Mycobacterium tuberculosis

- Central metabolism
- Operon of 7 proteins
- Detects nutrients availability

1DR2,1 CR1, 1AI
MaIAGe, Migale

1PI
IBR, Rosario, Argentine

MEM Grants
CONICET Deposited

Screen and target kinase PknG and its upstream and downstream partners

Dr. M.N. Lisa, IBR-Conicet, Rosario Argentina

Antigene in the new Crohn detection kit: Ser/Thr Protein Kinase PknG

Mycobacterium avium ssp paratuberculosis

Mycobacterium tuberculosis

Actinobactéries

?

- ➢ Conservation
- ➢ Distribution
- ➢ Structure/fonction

Structural homologs of Ser/Thr kinases and the 6 partners of the signaling pathway

Actinobacteria: 250 genomes, 3 months computation time, heat-map with R analysis

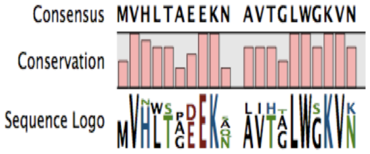
1DR2,1 CR1, 1AI
MaIAGe, Migale

1PI
IBR, Rosario, Argentine

MEM Grants
CONICET Deposited

Conclusions and Perspectives

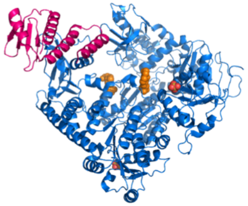
❖**Strong positioning** at the interface structural bioinformatics & Biology



❖**Innovative webserver** to achieve gut ecosystem exploration.

Needs scaling up & tuning (filters)

Experimental validation of the structural hits



❖**High valorisation** potential

HTS of structures

Structure –Function –Annotation

Functional clusters and Pathways

Originality /prevalence of folds

Disease/health -associated patterns of diversity

… Towards new targets with Health impact

… Towards other microbiotas

Acknowledgments

Mathématiques et Informatique Appliquées
MaIAGE
du Génome à l'Environnement

Véronique Martin

Jean-François Gibrat

Thomas Lacroix

migale
Plateforme de BioInformatique - INRA Jouy en Josas

Sandra Derozier

Valentin Loux

Micalis
Microbiologie de l'Alimentation au service de la Santé

Fabienne Béguet-Crespel

Jean-Marc Chatel

Philippe Gérard

Catherine Juste

Nalini Rama Rao

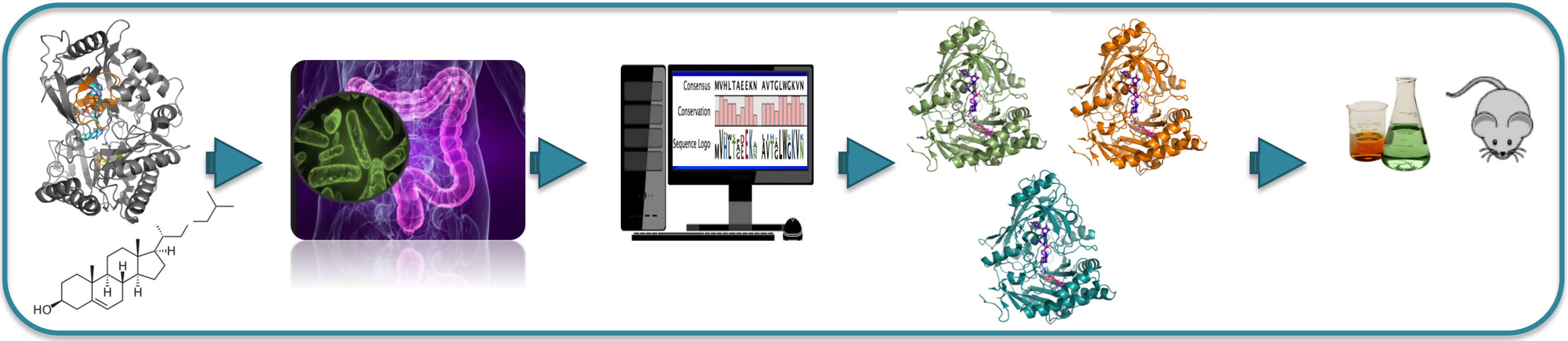
FACULTAD DE CIENCIAS
UDELAR fcien.edu.uy

Andrea Villarino

Mabel Berois

Danilo Segovia

Dario Porley



UNIVERSITÉ DE NANTES

Bernard Offmann

Surbhi Dingra

MEM Metaprogram

INRA
SCIENCE & IMPACT

Institut Pasteur

Mariano Martinez

Pedro Alzari

CONICET

María Natalia Lisa

Acknowledgments

MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES MaIAGE DU GÉNOME À L'ENVIRONNEMENT

Véronique Martin

Jean-François Gibrat

Thomas Lacroix

migale

Plateforme de BioInformatique - INRA Jouy en Josas

Sandra Derozier

Valentin Loux

Micalis

Fabienne Béguet-Crespel

Jean-Marc Chatel

Philippe Gérard

Catherine Juste

Nalini Rama Rao

FACULTAD DE CIENCIAS

UDELAR fcien.edu.uy

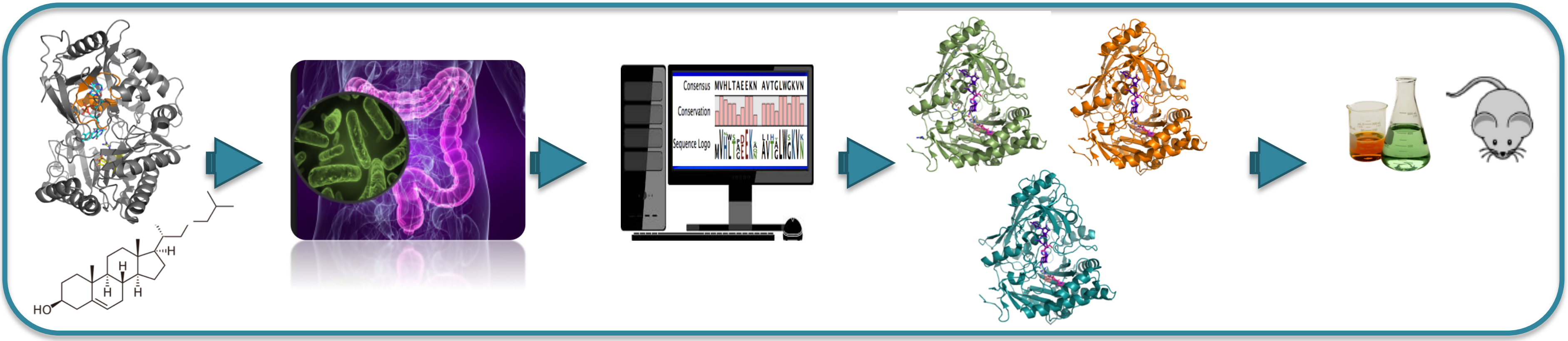
Andrea Villarino

Mabel Berois

Danilo Segovia

Dario Porley

And you for your kind attention . . .



UNIVERSITÉ DE NANTES

Bernard Offmann

Surbhi Dingra

MEM Metaprogram

INRA SCIENCE & IMPACT

Institut Pasteur

Mariano Martinez

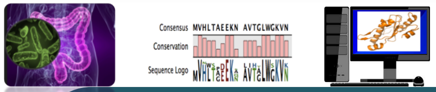
Pedro Alzari

CONICET

María Natalia Lisa

MetaFoldScan workflow

WP1



Bioinformatics development

Integration of MetaHit datasets, filtering, scanning and 3D fold detection

G. André-Leroux

- ✧ Set up of Metahit* browsing. Splitting of genomes. Filtering. Connection to pdb
- ✧ HMM profiling of MAM and MFD targets 3D candidates: from hits to leads
- ✧ Analysis of specificity, promiscuity, identification of associated pathways

Experimental validation of targets

WP2

MAM hits

JM. Chatel

- ✧ *In vitro* validation of MAM –like leads using HEK293 NFκB luciferase reporter assay
- ✧ *In vivo* validation of MAM-like leads using IBD colitis in mouse model
- ✧ MAM production, purification, crystallization, structural characterization

WP3

MFD hits

N. Rama Rao

- ✧ Clone MFD-like 10 best hits in *E.coli* deficient MFD mutant
- ✧ Test resistance to NO for “Complemented *E.coli* strains”
- ✧ Assess the function and specificity of MFD leads in prevalent species such as *Bacteroidetes* & *Firmicutes*

WP4

Applications - Impacts

Biologist friendly web-tool interface

V. Loux

- ✧ Production of command line pipeline distribution.
- ✧ Integration into Galaxy.
- ✧ Integration into training sessions within Migale Bioinformatics cycle
- ✧ Extension to other microbiotas

Li *et al*, Nat. Biotech. 2014

Analysis of the 5 genes from 3D, using Insyght

Thomas Lacroix

✪ http://genome.jouy.inra.fr/Insyght

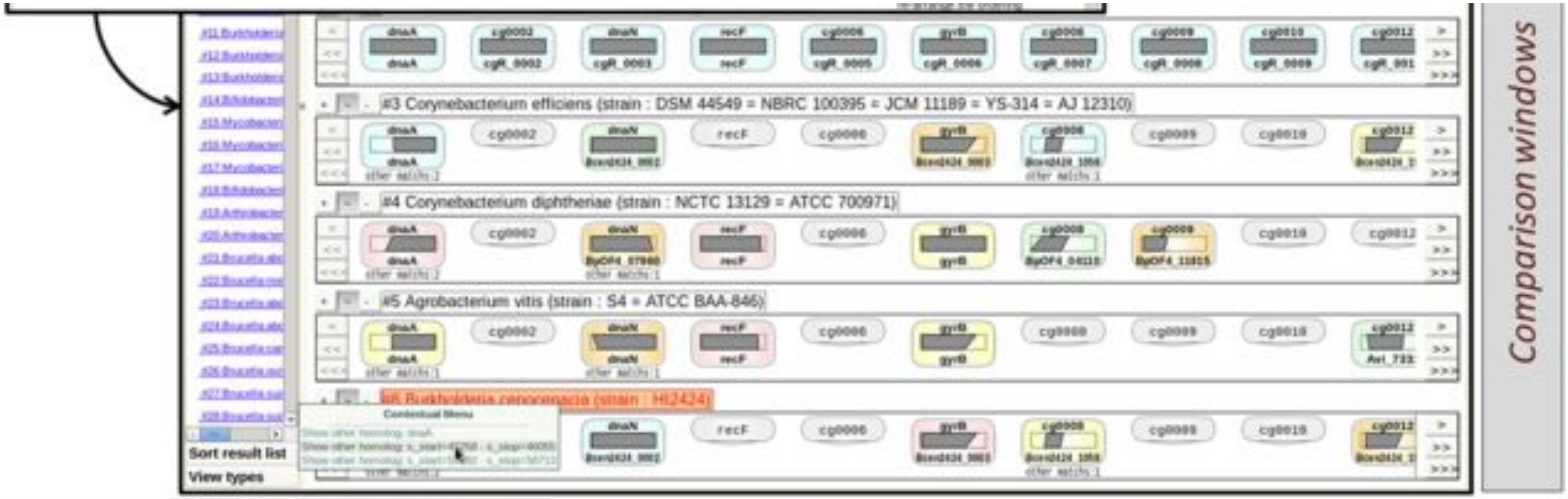
Powerful tool for genes comparison and syntheny inferences

Synthenies : co-localization of homologuous loci accross species

Comparison 2 vs 2 with a reference genome.

BDBH method Bi-Directional Best Hit i.e best reciprocal match between two proteomes.

Gene comparison. Synthenies visible



Info and options

Contextual menu: browse multiple homologies per gene

Comparison windows

Lacroix *et al*, Nar 2014