



UMR6026
Interactions Cellulaires
et Moléculaires



CoBaltDB

**Complete Bacterial and Archaeal Orfeomes Subcellular
Localization Database and Associated Resources**

(BMC Microbiol 2010 Mar 23;10:88.)

David Goudenège

Frédérique Barloy-Hubler

UMR-CNRS 6026

Equipe SP@RTE

Functional annotation of proteomes

Link between the sequence and biological informations

Biochemical function

Biological role

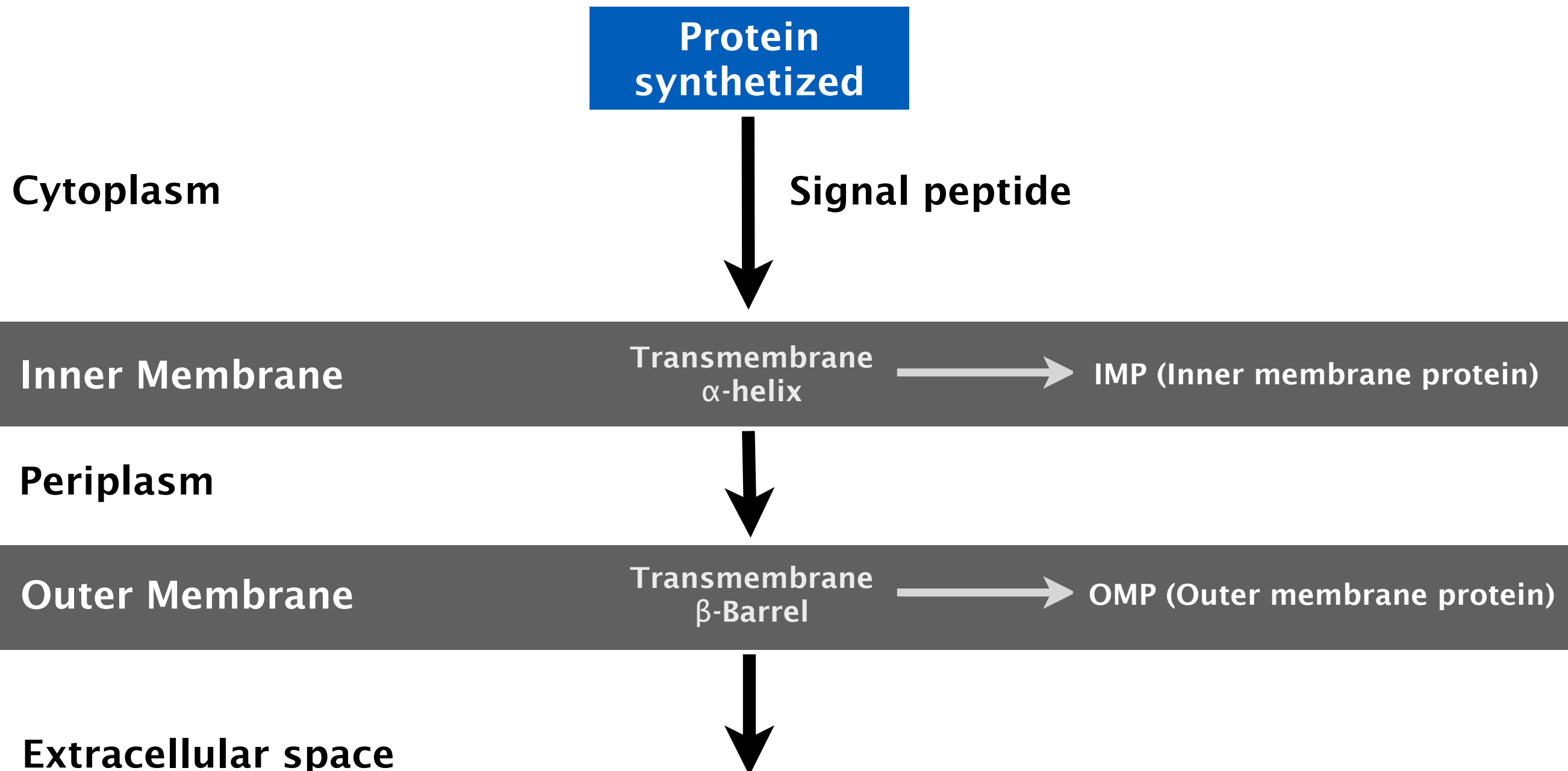
Regulation and interaction



SUBCELLULAR LOCALIZATION

Prokaryotic Cellular Targeting

Targeting depends on sequence



Plethora of prediction tools (>100)

Difference of implementation

→ Standalone / Web-Tools / Database

Difference of methods

→ HMM, NN, SVM / Motif / Hydrophobicity

Difference of predicted features

→ Specialized Tools / Global Tools
(signal peptide, α -helix TM) (final localization)

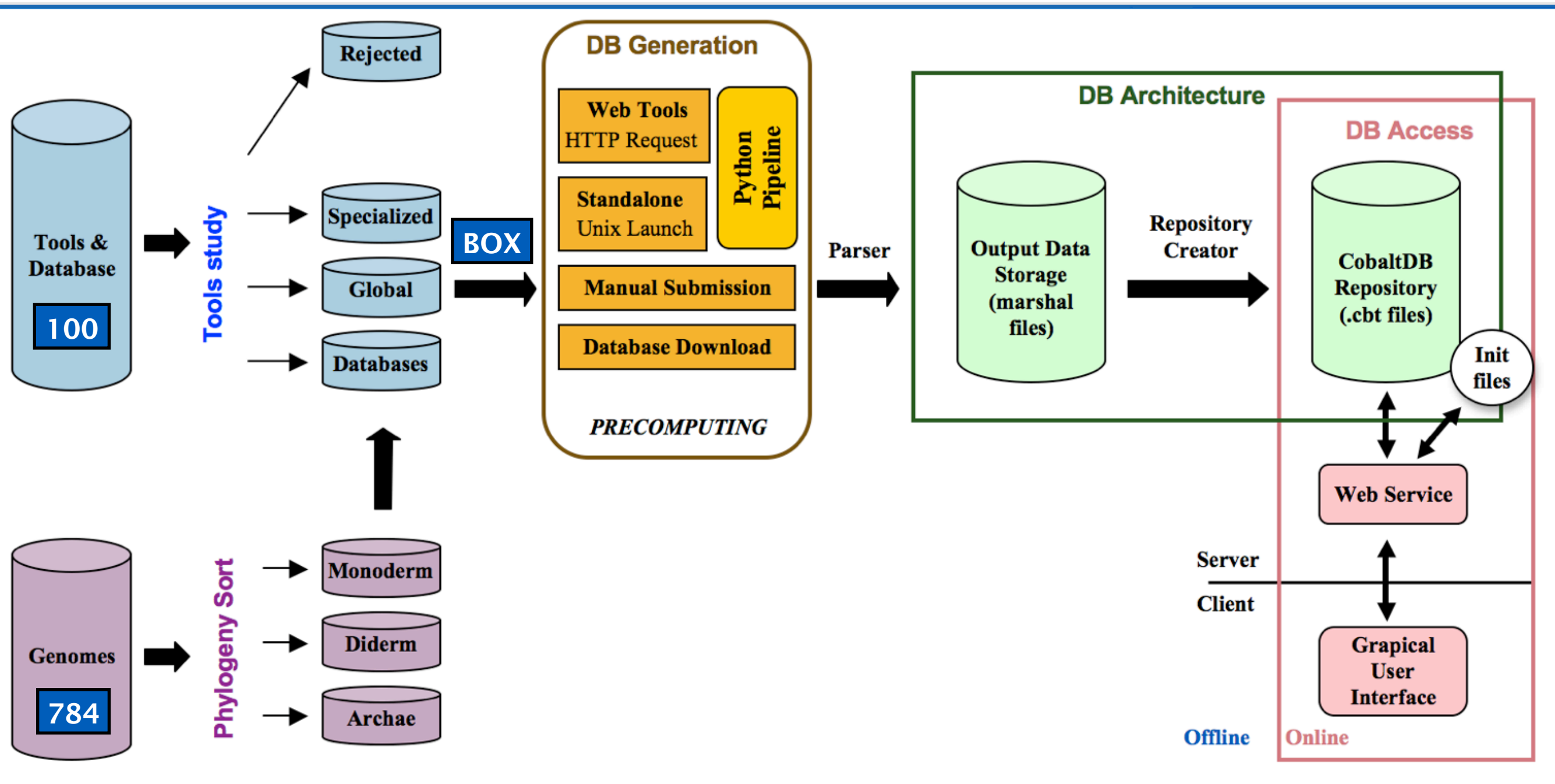
Sensibility & Specificity



→ Phylum / Protein type

CoBaltDB

GROUP / ORGANIZE / DISPLAY
localization informations

CoBaltDB Workflow



 **CoBalt** Version 1.0 

Search Cellular Localisation of

☒ *all proteins of a given Genome*

☐ *some protein(s) defined by their Locus Tag(s)*

Search

- Acaryochloris marina MBIC11017
- Acholeplasma laidlawii PG-8A
- Acidiphilium cryptum JF-5
- Acidithiobacillus ferrooxidans ATCC 53993
- Acidobacteria bacterium Ellin345
- Acidothermus cellulolyticus 11B
- Acidovorax avenae subsp. citrulli AAC00-1
- Acidovorax sp. JS42
- Acinetobacter baumannii AB0057
- Acinetobacter baumannii AB307-0294
- Acinetobacter baumannii ACICU
- Acinetobacter baumannii ATCC 17978
- Acinetobacter baumannii AYE
- Acinetobacter baumannii SDF
- Acinetobacter sp. ADP1
- Actinobacillus pleuropneumoniae L20
- Actinobacillus pleuropneumoniae serovar 3 str. JL03
- Actinobacillus pleuropneumoniae serovar 7 str. AP76
- Actinobacillus succinogenes 130Z
- Agrobacterium hydovora subsp. hydovora ATCC 7066

CoBalt Version 1.0

Input Specialized Tools (Feature Predictions) Meta Tools (Localization Predictions) Additionnal Tools (Posts)

Genome Escherichia coli str. K-12 substr. DH10B Replicon All Replicons

Search Next >

Proteins whose cellular localization is experimentally validated. Click on headers to sort columns.

Locus Tag	Protein Id	Gene	Description	Exper...	Replicon	LipoBox	TatBox	SecBox	SPdb	Swissp...	SignalP...	HelixBox	TOP_DB	TMPDB	BarrelB...	TOP_DB	TMPDB	TMBC...
ECDH10B_0010	YP_001728993.1	yaaH	inner membrane protei...	<input type="checkbox"/>	NC_010473 (chro...								Alpha h...			Alpha h...		
ECDH10B_0011	YP_001728994.1	yaaW	hypothetical protein	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0012	YP_001728995.1	htgA	hypothetical protein	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0013	YP_001728996.1	yaaI	hypothetical protein	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0014	YP_001728997.1	dnaK	chaperone Hsp70, co-c...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0015	YP_001728998.1	dnaJ	chaperone Hsp40, co-c...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0016	YP_001728999.1	insL-1	IS186/IS421 transposase	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0017	YP_001729000.1	mokC	regulatory protein for...	<input type="checkbox"/>	NC_010473 (ch...	0	0	57				100			20			
ECDH10B_0018	YP_001729001.1	hokC	toxic membrane protei...	<input checked="" type="checkbox"/>	NC_010473 (chro...								Inner M...			Inner M...		
ECDH10B_0020	YP_001729002.1	nhaA	sodium-proton antipor...	<input checked="" type="checkbox"/>	NC_010473 (chro...								Alpha h...	Inner M...		Alpha h...	Inner M...	
ECDH10B_0021	YP_001729003.1	nhaR	DNA-binding transcrip...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0022	YP_001729004.1	insB-1	IS1 transposase InsAB'	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0023	YP_001729005.1	insA-1	IS1 repressor protein l...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0024	YP_001729006.1	rpsT	30S ribosomal subunit...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0025	YP_001729007.1	yaaY	hypothetical protein	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0026	YP_001729008.1	ribF	bifunctional riboflavin...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0027	YP_001729009.1	ileS	isoleucyl-tRNA synthet...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0028	YP_001729010.1	lspA	prolipoprotein signal p...	<input checked="" type="checkbox"/>	NC_010473 (chro...								Alpha h...	Inner M...		Alpha h...	Inner M...	
ECDH10B_0029	YP_001729011.1	fkpB	FKBP-type peptidyl-pro...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0030	YP_001729012.1	ispH	1-hydroxy-2-methyl-2...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0031	YP_001729013.1	rihC	ribonucleoside hydrola...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0032	YP_001729014.1	dapB	dihydrodipicolinate red...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0033	YP_001729015.1	carA	carbamoyl phosphate s...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0034	YP_001729016.1	carB	carbamoyl-phosphate...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0035	YP_001729017.1	caiF	DNA-binding transcrip...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0036	YP_001729018.1	caiE	acyl transferase	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0037	YP_001729019.1	caiD	crotonobetainyl CoA h...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0038	YP_001729020.1	caiC	crotonobetaine CoA lig...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0039	YP_001729021.1	caiB	crotonobetainyl CoA:ca...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0040	YP_001729022.1	caiA	crotonobetaine reducta...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0041	YP_001729023.1	caiT	transporter	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0042	YP_001729024.1	fixA	electron transfer flavop...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0043	YP_001729025.1	fixB	electron transfer flavop...	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0044	YP_001729026.1	fixC	oxidoreductase	<input type="checkbox"/>	NC_010473 (chro...													
ECDH10B_0045	YP_001729027.1	fixY	4Fe-4S ferredoxin-h...	<input type="checkbox"/>	NC_010473 (chro...													

BOX-Score

Get Protein Localization Form... Go to KEGG Go to NCBI Save Feature Table as xls...

CoBalt Version 1.0

Input Specialized Tools (Feature Predictions) **Meta Tools (Localization Predictions)** Additionnal Tools (Posts)

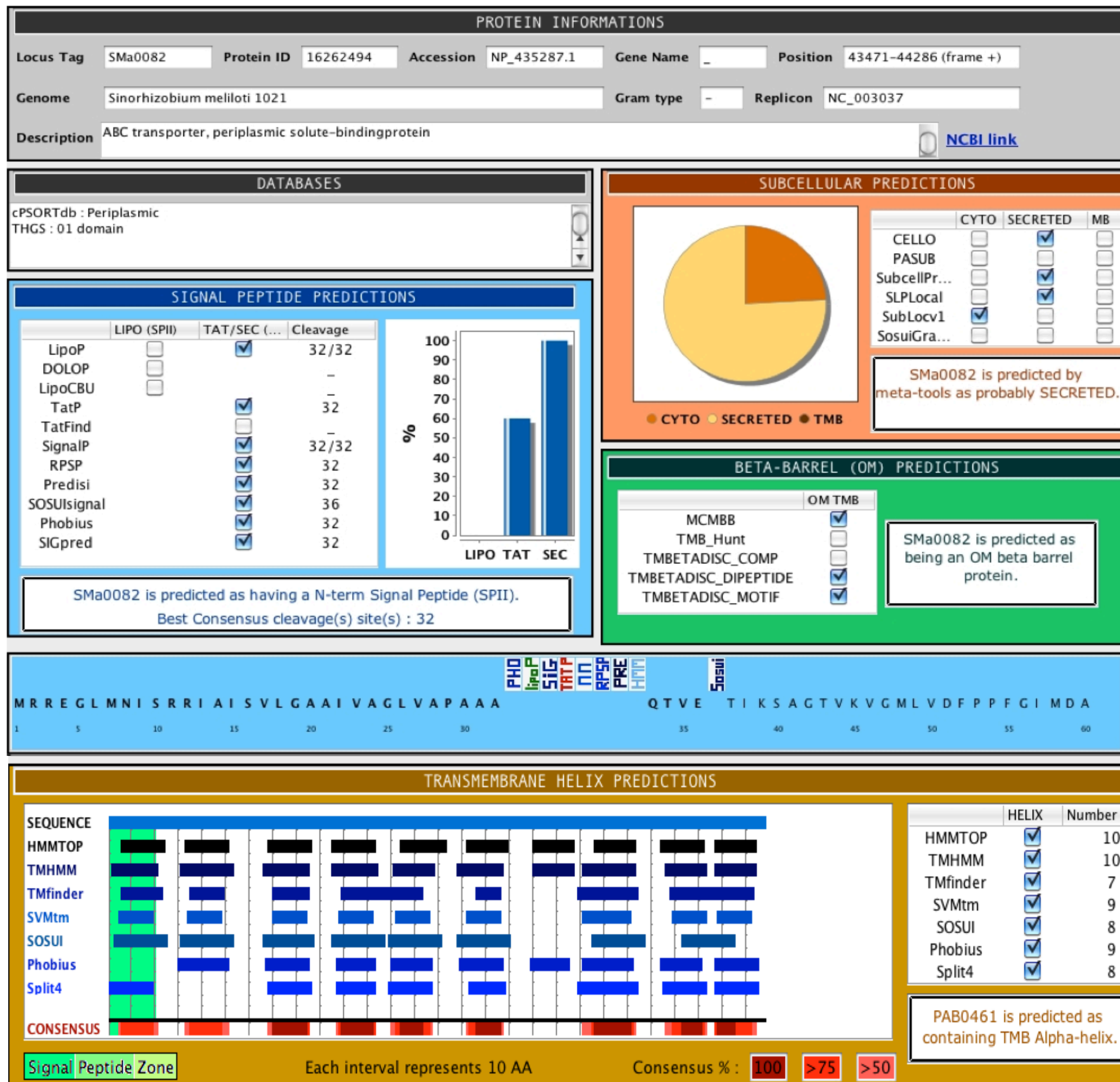
Genome Replicon

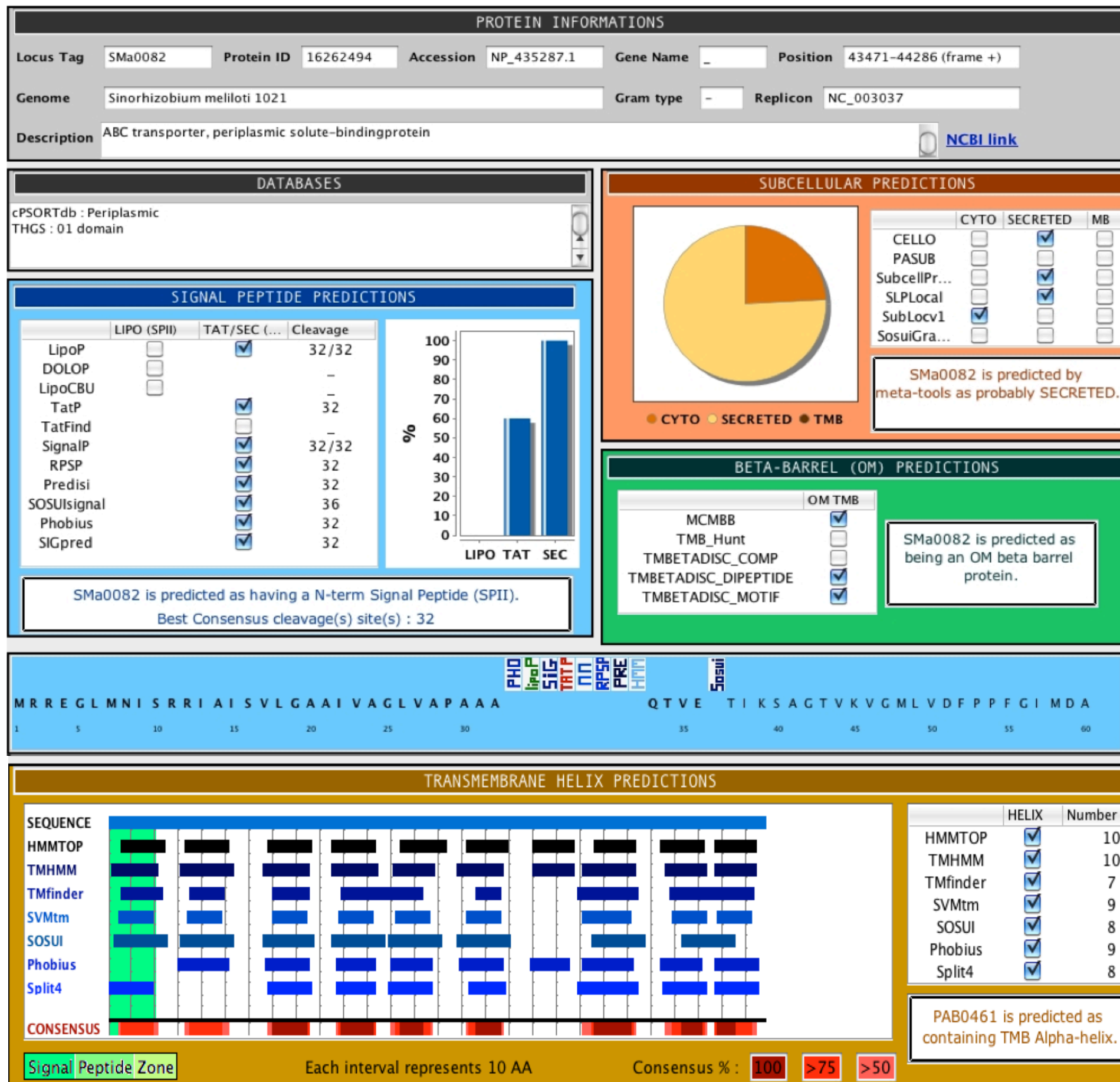
Search Next >

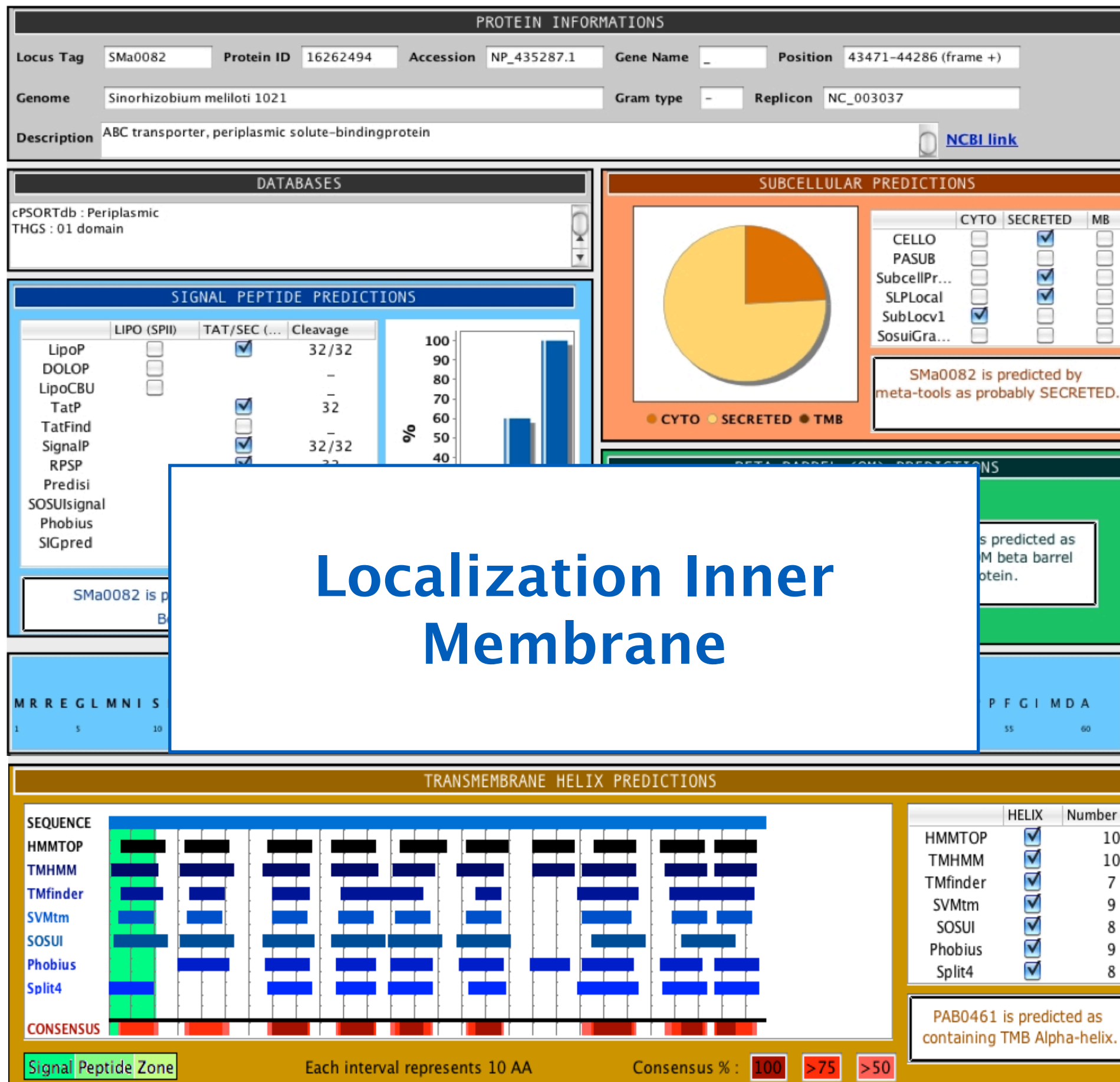
Click on headers to sort columns.

Locus Tag	Protein Id	Gen...	Description	Replicon	CELLO	PASUB	SubcellPred...	SLPLocal	SubLocv1	SosuiGramN	ecce	EchoLOCA...
ECDH108_0001	YP_001728984.1	thrL	thr operon leader peptide	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Extracellular	Cytoplasmic	Unknown	-	Cytoplasmic
ECDH108_0002	YP_001728985.1	thrA	fused aspartokinase I a...	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0003	YP_001728986.1	thrB	homoserine kinase	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0004	YP_001728987.1	thrC	threonine synthase	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0005	YP_001728988.1	yaaX	hypothetical protein	NC_010473 (chro...	Periplasmic	Unknown	Periplasmic	Periplasmic	Periplasmic	Periplasmic	-	Periplasmic
ECDH108_0006	YP_001728989.1	yaaA	hypothetical protein	NC_010473 (chro...	Cytoplasmic	Unknown	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0007	YP_001728990.1	yaaJ	transporter	NC_010473 (chro...	Inner Mem...	Inner Mem...	Extracellular	Cytoplasmic	Cytoplasmic	Inner Mem...	-	Integral Me...
ECDH108_0008	YP_001728991.1	talB	transaldolase B	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0009	YP_001728992.1	mog	molybdochelatase	NC_010473 (chro...	Cytoplasmic	Unknown	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0010	YP_001728993.1	yaaH	inner membrane protei...	NC_010473 (chro...	Inner Mem...	Inner Mem...	Extracellular	Cytoplasmic	Extracellular	Inner Mem...	-	Integral Me...
ECDH108_0011	YP_001728994.1	yaaW	hypothetical protein	NC_010473 (chro...	Cytoplasmic	Unknown	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0012	YP_001728995.1	htgA	hypothetical protein	NC_010473 (chro...	Extracellular	Cytoplasmic	Periplasmic	Extracellular	Periplasmic	Extracellular	-	Cytoplasmic
ECDH108_0013	YP_001728996.1	yaaI	hypothetical protein	NC_010473 (chro...	Periplasmic...	Unknown	Periplasmic	Periplasmic	Periplasmic	Outer Mem...	-	Periplasmic
ECDH108_0014	YP_001728997.1	dnaK	chaperone Hsp70, co-c...	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0015	YP_001728998.1	dnaJ	chaperone Hsp40, co-c...	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0016	YP_001728999.1	insL-1	IS186/IS421 transposase	NC_010473 (chro...	Cytoplasmic	Unknown	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	-
ECDH108_0017	YP_001729000.1	mokC	regulatory protein for...	NC_010473 (ch...	Cytoplasmic	Inner Mem...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Periplasmic
ECDH108_0018	YP_001729001.1	hokC	toxic membrane protei...	NC_010473 (chro...	Cytoplasmic	Inner Mem...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Unknown	PERI	Periplasmic
ECDH108_0020	YP_001729002.1	nhaA	sodium-proton antipor...	NC_010473 (chro...	Inner Mem...	Inner Mem...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Inner Mem...	-	Integral Me...
ECDH108_0021	YP_001729003.1	nhaR	DNA-binding transcrip...	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0022	YP_001729004.1	insB-1	IS1 transposase InsAB'	NC_010473 (chro...	Cytoplasmic	Unknown	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	-
ECDH108_0023	YP_001729005.1	insA-1	IS1 repressor protein I...	NC_010473 (chro...	Extracellular...	Unknown	Periplasmic	Periplasmic	Extracellular	Unknown	-	-
ECDH108_0024	YP_001729006.1	rpsT	30S ribosomal subunit...	NC_010473 (chro...	Periplasmic...	Cytoplasmic	Cytoplasmic	Periplasmic	Periplasmic	Periplasmic	-	Cytoplasmic
ECDH108_0025	YP_001729007.1	yaaY	hypothetical protein	NC_010473 (chro...	Cytoplasmic	Unknown	Cytoplasmic	Cytoplasmic	Extracellular	Unknown	-	Membrane...
ECDH108_0026	YP_001729008.1	ribF	bifunctional riboflavin...	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Periplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0027	YP_001729009.1	ileS	isoleucyl-tRNA synthet...	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0028	YP_001729010.1	lspA	prolipoprotein signal p...	NC_010473 (chro...	Inner Mem...	Inner Mem...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Inner Mem...	-	Integral Me...
ECDH108_0029	YP_001729011.1	fkpB	FKBP-type peptidyl-pro...	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0030	YP_001729012.1	ispH	1-hydroxy-2-methyl-2...	NC_010473 (chro...	Cytoplasmic	Unknown	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0031	YP_001729013.1	rihC	ribonucleoside hydrola...	NC_010473 (chro...	Cytoplasmic...	Unknown	Cytoplasmic	Cytoplasmic	Cytoplasmic	Unknown	-	Cytoplasmic
ECDH108_0032	YP_001729014.1	dapB	dihydrodipicolinate red...	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0033	YP_001729015.1	carA	carbamoyl phosphate s...	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0034	YP_001729016.1	carB	carbamoyl-phosphate...	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Unknown	-	Cytoplasmic
ECDH108_0035	YP_001729017.1	caiF	DNA-binding transcrip...	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
ECDH108_0036	YP_001729018.1	caiE	acyl transferase	NC_010473 (chro...	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic	Inner Mem...	-	Cytoplasmic

Save Meta Table as xls



Signal
peptideProbable
 β -barrelCleavage
at 32Multiple
 α -helix



Signal peptide

Probable β -barrel

Cleavage at 32

Multiple α -helix

Localization Inner Membrane

Localization platform

784 proteomes (2.548.292 proteins) available

More 50 tools and databases

User-friendly interface

Improve localization prediction

Combine diluted and heterogenous localization informations

**Box organisation group multiple tools with different specificity and sensibility
=> the microbiologist construct his own prediction**

CoBaltDB

Complete Bacterial and Archaeal Orfeomes Subcellular Localization Database and Associated Resources

D. Goudenège, S. Avner, C. Lucchetti-Miganeh, F. Barloy-Hubler

 UMR CNRS 6026 - Université de Rennes 1 - B@SIC Team

Introduction

Determining the subcellular localization of proteins is essential for the functional annotation of proteomes. The functions of proteins are strongly related to their localization in cell compartments. Prokaryotic proteins can be cytoplasmic, membranar (inner, outer), periplasmic or extracellular. Different subcellular targeting mechanisms allow non-cytoplasmic proteins to leave the cytoplasm after their synthesis.

In silico prediction of protein localization is a important task for proteome annotation and comparative genomic. However, biologists are confronted with a very large number of computational tools with different methods, localization, specificities and sensitivities. As a result, exploiting these computer resources to predict protein localization accurately involves querying all tools and comparing every prediction output; this is a painstaking task.

CoBaltDB Interface

CoBaltDB presents four tabs that perform specific tasks: the “input” tab allows selecting the organism or a list of locus tags. The “Specialized tools”

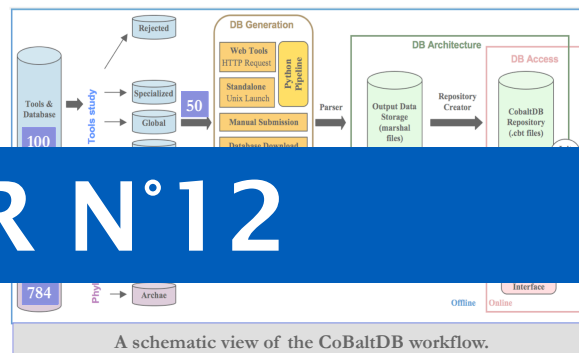
POSTER N°12

global databases. The “additional tools” tab enables queries to be submitted to a set of 50 additional tools. Finally, for each protein, all results were summarized in a synopsis, to allow the investigators to establish their own hypotheses and conclusions.

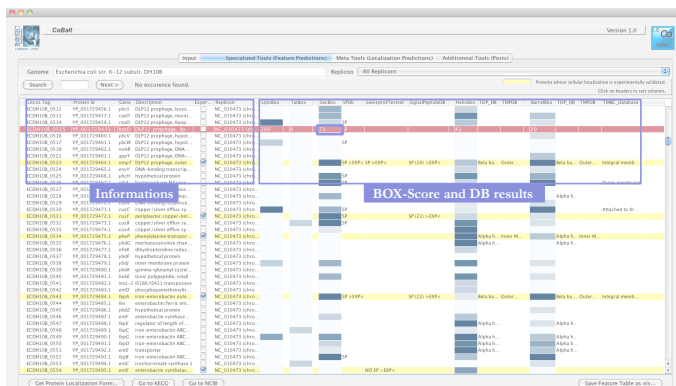
CoBaltDB Architecture

The CoBaltDB database contains three main types of data: i) prediction using 23 feature-based localization tools, ii) prediction using 5 localization meta-tools and iii) data from 20 public database. These data were organized in five “boxes” with regard to the features predicted: signal peptide detection (Lipoprotein, Tat- and Sec- dependent signals); prediction of α -transmembrane segments; and outer membrane β -barrels prediction.

We retrieved and tested 99 available specialized and global tools that use various amino acid features and diverse methods (HMM, NN, SVM...). Some tools are Gram specific, requiring a phylogenic sorting. **Currently, CoBaltDB contains pre-computed results obtained with 48 tools and databases, and additionally provides pre-filled access to 50 publicly available tools.**



A schematic view of the CoBaltDB workflow.



The CoBaltDB Specialized Tools viewer, displays BOX-score and some database results for a given proteome

CoBaltDB is installed at Genouest

<http://www.umn6026.univ-rennes1.fr/english/home/research/basic/software/cobaltan>

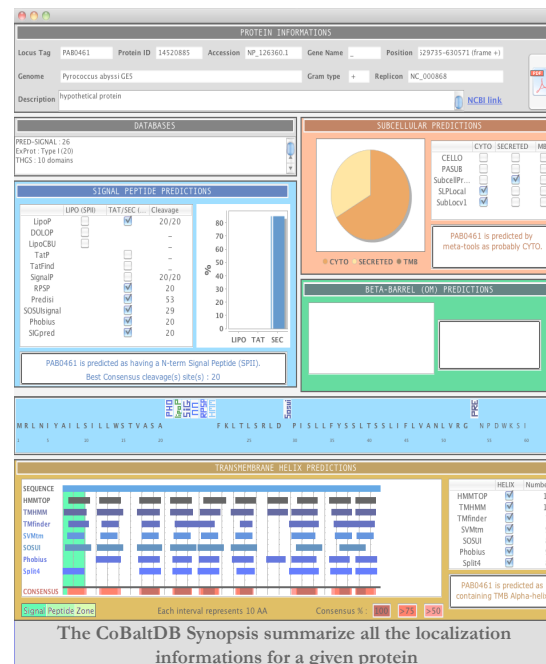
Reference

Goudenège D, Avner S, Lucchetti-Miganeh C, Barloy-Hubler F: CoBaltDB: Complete bacterial and archaeal orfeomes subcellular localization database and associated resources. BMC Microbiol. 2010 Mar 23;10:88.

Conclusion

CoBaltDB is the first friendly interfaced database that compiles a large number of in silico subcellular predictions concerning whole prokaryotic proteomes. Currently, CoBaltDB allows fast access to precomputed localizations for 2,548,292 proteins in 784 proteomes.

The combination and comparative analysis of results of heterogeneous tools improved the computational predictions, and contributed to identifying the limitations of each tool. Therefore, CoBaltDB can serve as a reference resource to facilitate interpretation of results and to provide a benchmark for accurate and effective in silico predictions of the subcellular localization of proteins.



The CoBaltDB Synopsis summarize all the localization informations for a given protein