MetaFoldScan: a pipeline to scan metagenomes and identify structural homologs using HMM

Sandra Dérozier\textsuperscript{1,}, Véronique Martin\textsuperscript{1,}, Jean-Marc Chatel\textsuperscript{1,}, Nalini Rama Rao\textsuperscript{1,}, Valentin Loux\textsuperscript{1,}, Gwenaëlle André-Leroux\textsuperscript{1}\textsuperscript{*}

\textsuperscript{1} MalAGE, INRA, Université Paris-Saclay, 78350, Jouy-en-Josas, France
\textsuperscript{2} MICALIS, INRA, Université Paris-Saclay, 78350, Jouy-en-Josas, France

\textsuperscript{*} equal contribution

**Metafoldscan** -MFS- aims at developing a user-friendly interface to intensively scan metagenomes and identify structural homologs connected to a target protein or to model exhaustively the whole proteins of a genome. To our knowledge, no solution that combines accurate browsing of meta-genomes and reasonable computational time is available. Yet, MFS addresses these points. Currently, it is set up with the core genome of the human gut microbiota that clusters 57 highly prevalent bacteria\textsuperscript{1,2} and with MAM - Microbial Anti-Inflammatory Molecule\textsuperscript{3} and MFD - Mutation Frequency Decline\textsuperscript{4}. Both proteins associate structural challenge with biological and therapeutic relevance.

---

**MetaFoldScan Pipeline**

**Target**

- 3D Fold Bank
- MFD template
- MAM template

**Query**

- Genome X
- Genome Y
- Genome Z

**Filtering of sequence candidates**

**Multiple Sequence Alignment of filtered candidates**

**Prediction and addition of secondary structure elements**

**HMM profile of the target bank\textsuperscript{1}**

**HMM profile of Multiple Sequence Alignment\textsuperscript{5}**

**Profile (candidate) vs profile (bank) comparison\textsuperscript{5}**

**Extraction, Ranking & Analyse of scores**

**Hits prediction**

**Validation**

- Crystal structure in vitro/in vivo testing
- Abundance classification

**Biomarker ?**


---

**MetaFoldScan Galaxy**

**MetaFoldScan Galaxy Galaxie**

**MetaFoldScan Galaxy Galaxie Version 1.0**

**Profile of interest**

- Target

**Query**

**Query**

**Filtering of sequence candidates**

**Extraction & Ranking results**

**Hits prediction**

**Validation**

**Crystal structure in vitro/in vivo testing**

**Abundance classification**

**Biomarker ?**

**MFD template**

**MAM template**

**3D Fold Bank**

**HMM profile of the target bank\textsuperscript{1}**

**HMM profile of Multiple Sequence Alignment\textsuperscript{5}**

**Profile (candidate) vs profile (bank) comparison\textsuperscript{5}**

**Extraction, Ranking & Analyse of scores**

**Hits prediction**

**Validation**

- Crystal structure in vitro/in vivo testing
- Abundance classification

**Biomarker ?**

**Results for Faecalibacterium prausnitzii-SL3_3**

\textsuperscript{1} 1/26/2016 11:20 AM

**Extraction & Ranking results**

**Hits prediction**

**Validation**

- Crystal structure in vitro/in vivo testing
- Abundance classification

**Biomarker ?**

**Challenges tackled & Results**

- Metafoldscan is accessible within the user-friendly Galaxy interface of the Migale platform: [http://migale.jouy.inra.fr/galaxy/](http://migale.jouy.inra.fr/galaxy/).
- Identification of structural homologs of MAM, only in the strain Faecalibacterium prausnitzii-SL3_3, confirming its ecological niche.
- Identification of structural homologs of Mfd, within the gut ecosystem, confirming its ubiquitous prevalence.

**Perspectives**

- Set-up of accurate filters to scale up to the 10 millions genes of the gut microbiota.
- Exploration of a network of proteins putatively highly conserved and involved in both signaling and metabolism in Actinobacteria.
- Validation of the hits and discovery of new enzymes with possibly therapeutic functions.

---

**Bibliography**