Extensive characterisation of the gut resistome yields new insights about the microbiome

Amine Ghozlane[1]*, Etienne Ruppé[1]*, Julien Tap[1]*, Nicolas Pons[1], Alexandre G. de Brevern[2], Sean P. Kennedy[1], S. Dusko Ehrlich[1]

[1]Metagenopolis, INRA, Jouy-en-Josas, France, [2]INSERM, Univ Paris Diderot, France. Contact: amine.ghozlane@jouy.inra.fr

Abstract

Background. The gut microbiota harbors an allegedly vast diversity of antibiotic resistance determinants (ARDs) yet their census (i.e. the resistome) has not been previously determined. Indeed, bioinformatic tools are stymied by the identity gap between known ARDs and those of bacteria from the gut microbiota. Thus, whether subjects can be stratified according to their resistome remains unanswered. Here, we used a new 3-dimensional modeling based approach to accurately identify ARDs. We then stratified MetaHIT subjects with regards to their gut resistome.

Methods. We developed a new method of functional annotation named pairwise comparative modeling (PCM). Homology modeling of candidates with templates (PDB) identified as (i) reference on one hand, and (ii) negative on the other hand are compared. Scores generated by the two modeling paths were compared and the candidates classified into the most appropriate category. When tested with an external functional metagenomic dataset, ARD predictions by were 99.1% (1,380/1,391) true. We then queried the 3.9M MetaHIT gene catalogue for ARDs belonging to 20 classes, conferring resistance to nine major antibiotic families. We attempted to stratify 663 subjects from the MetaHIT cohort according to their ARDs class distributions, and assessed the possible connexions between gut resistome, richness and enterotypes.

Results. Using the PCM, we predicted 6,095 ARDs among which half had an amino-acid identity below 30%. ARDs candidates were assigned to Firmicutes (49%), Bacteroidetes (14%) and Proteobacteria (4%) phyla, while 29% remained unassigned. The distribution of phyla varied according to the ARD family: aminoglycosides-modifying enzymes (AMEs) and class B beta-lactamases (bla) were enriched in Firmicutes while class A bla and Sul were enriched in Bacteroidetes. Of note, we predicted four ARDs in Methanobrevibacter and three in Methanoculleus. A chromosomal localization was suggested for 59.9% of ARDs. Mapping reads frequencies ranged from 0.18% to 0.52% per metagenomes. Six ARD clusters, using distribution patterns of ARDs classes, were detected. We observed that ARDs richness was positively correlated with overall gene richness and that ARDs clusters were associated with enterotypes: *Bacteroides* driven enterotype was associated with two ARD clusters enriched in class D beta-lactamases and tetracycline resistance conferring Tet(X), while Clostridiales driven enterotype was associated with three ARD clusters enriched in AMEs and *Prevotella* driven enterotype with a class B1-bla enriched cluster.

Conclusions. The human gut resistome is associated with gene richness and enterotypes. Our findings open perspectives in deciphering the variable response of the gut microbiota to antibiotics.

Acknowledgement. This research is sponsored by the European Union FP7 projects EvoTAR-282004.

Background

- The intestinal microbiota is a vast reservoir for genes with low identity with known genes, including those conferring antibiotic resistance.
- Current methods based on sequence similarity might not be able to discriminate true antibiotic resistance determinants from homologs with distinct function.
- Functional validation is gold standard yet thousands of putative ARDs should be tested.
- New methods are necessary: we tested a new protocol based on homology modeling on 20 classes of ARDs (antibiotic resistance determinants) family.
- We then stratified MetaHIT subjects with regards to their gut resistome.

Materials and Methods

- Pairwise comparative modeling (PCM) is a method based on a comparison between two modeling paths of one candidate: one path uses a reference template, the other uses negative references templates that do not share the function but some sequence identity with candidate (Figure 1).
- PCM is first applied on reference ARD and negative references to build a model by logistic regression (Figure 2). Then, the modeling and alignment scores of the candidates are submitted to the model and a confidence score (the percentage of times the candidates has been classified as an ARD) is obtained.

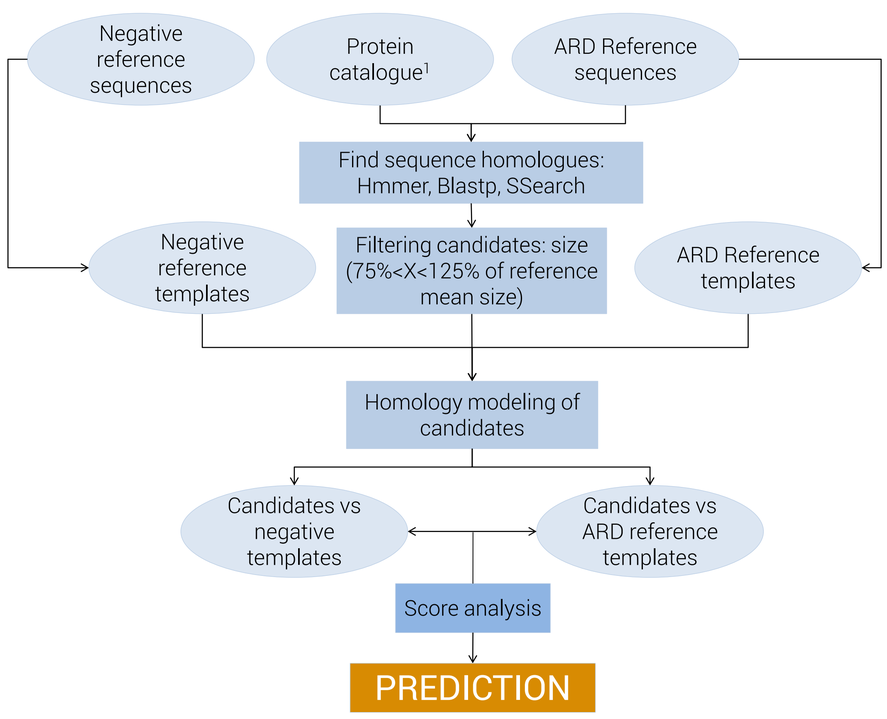

Figure 1: Concept of pairwise comparative modeling.
[1]Nielsen, H. B. and Almeida, M. et al. Nat. Biotechnol. 32, 822–828 (2014).

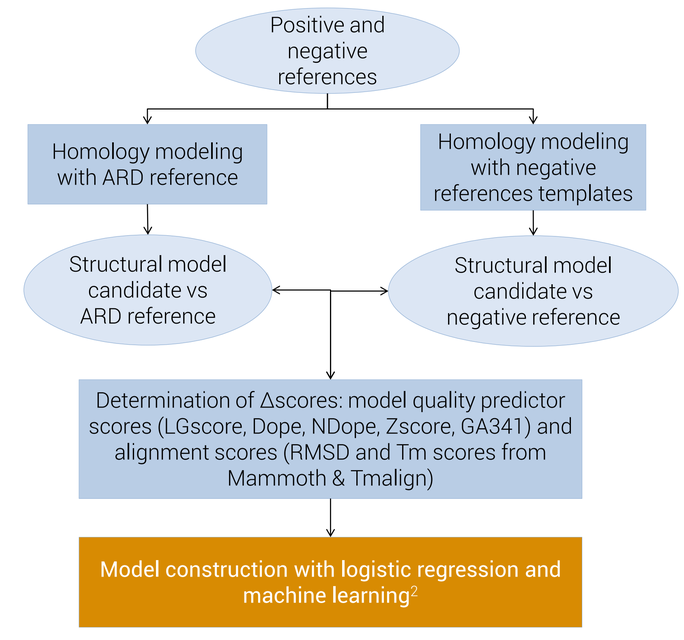

Figure 2: Construction of the logistic regression based on the different modeling and alignment scores obtained with the ARD and negative reference sequences
[2]1000 boostraps, 10 cross-validations

Results/discussion

- 10077 ARD candidates were identified: 6095 were predicted as ARD with >50% confidence (Figure 3). 28% of the predicted ARDs could not be assigned a phylum (Figure 4). PCM showed 99.1% sensitivity (Figure 5). Predicted ARDs richness appeared to be positively correlated with the overall richness among subjects of MetaHIT[1] cohort (Figure 7).

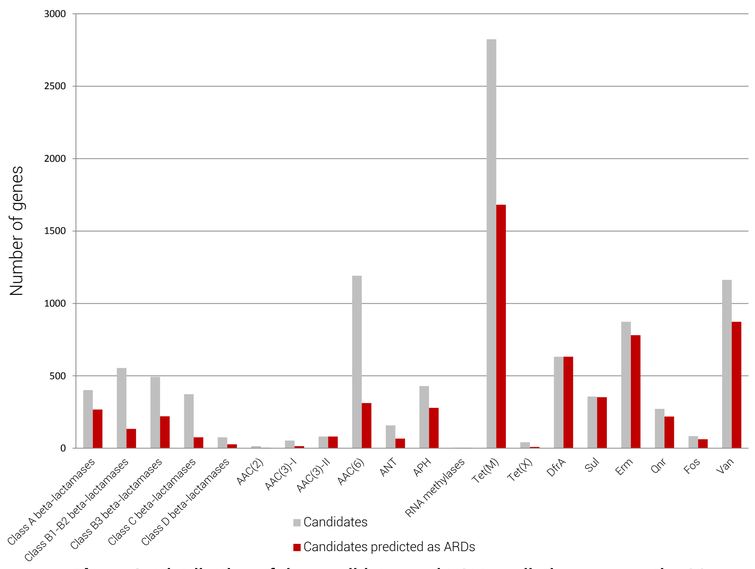

Figure 3: Distribution of the candidates and PCM predictions among the 20 classes of ARDs.

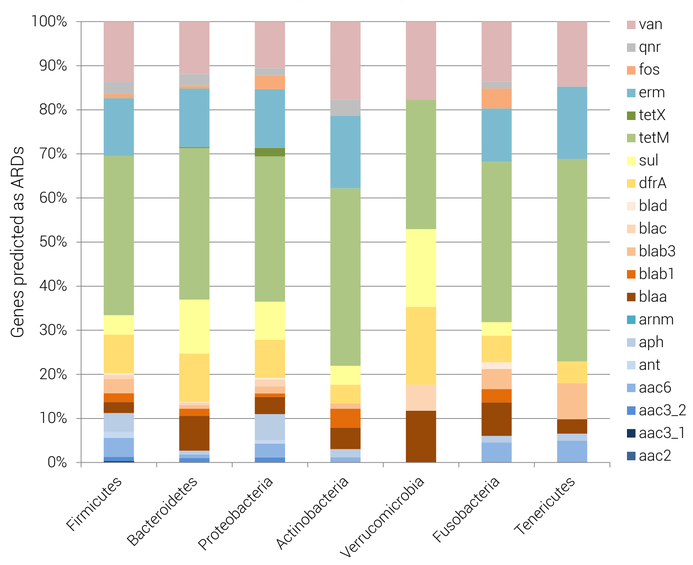

Figure 4: Distribution of the 6095 genes predicted by PCM as ARDs according to the phylum.

Family	Number of ARDs in the set	Number of PCM candidates	Number of true predictions	Resfams[3]	Resfinder[4]	ARG-ANNOT[5]	Blastp 80% identity[6]
Class A beta-lactamases	110	107	106	65	2	2	2
Class B beta-lactamases	141	136	132	101	1	1	4
Class C beta-lactamases	20	20	19	0	0	0	1
Class D beta-lactamases	81	69	67	43	1	1	1
AAC(3)	15	15	15	11	0	0	0
AAC(6)	7	4	2	7	0	0	0
APH	4	4	4	4	0	0	0
ARN methylases	15	15	15	10	0	0	0
Tet(X)	15	12	12	0	0	0	0
DfrA	1015	1008	1002	940	0	0	0

Figure 5: Comparison of the prediction of PCM, Resfams[3], Resfinder[4], ARG-ANNOT[5] and Blastp annotation against eggNOG v3 (http://eggnog.embl.de/version_3.0/) with a 80% identity threshold[6] of an external functional metagenomic dataset[7] .
[3]Gibson, M. K. et al. ISME J. 9,207-216 (2015), [4]Zankari, E. et al. Antimicrob. Agents Chemother. 67,2640-2644 (2012), [5]Gupta, S. K. et al. Antimicrob. Agents Chemother. 58,212-220 (2014), [6]Forslund, K. et al. GenomeRes. 23, 1163-1169 (2013), [7]Forsberg, KJ., et al. Nature, 29; 509 (2014).

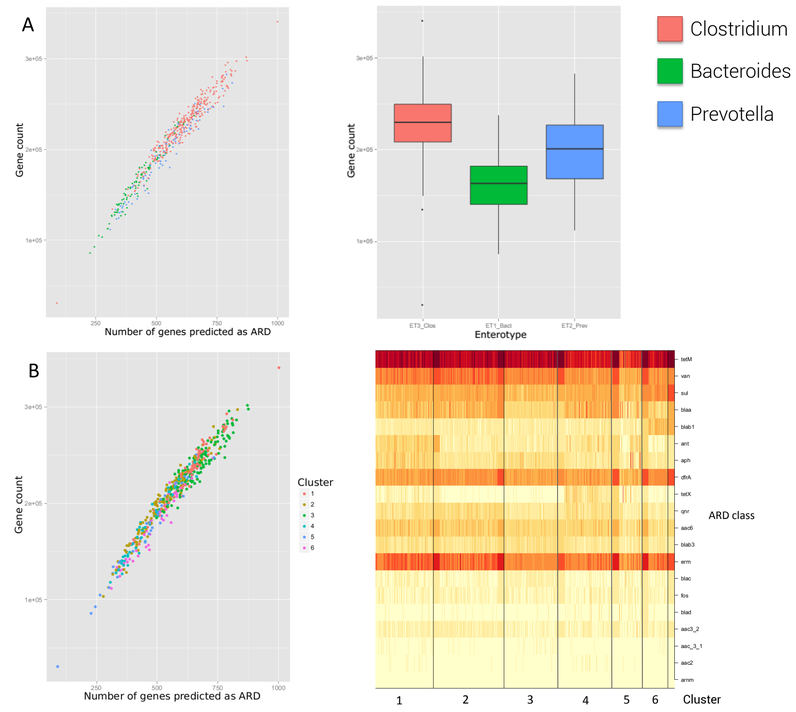

Figure 7: Comparison of the predicted ARDs richness against the overall gene richness of the 663 subjects of the MetaHIT[1] cohort depending on their enterotype classification (A) or their ARD patterns (B). The subject patterns for the ARD classes were calculated with the R package Dirichlet-Multinomial[7]. [7]Holmes, H. PloS ONE 7(2): 1-15 (2012)

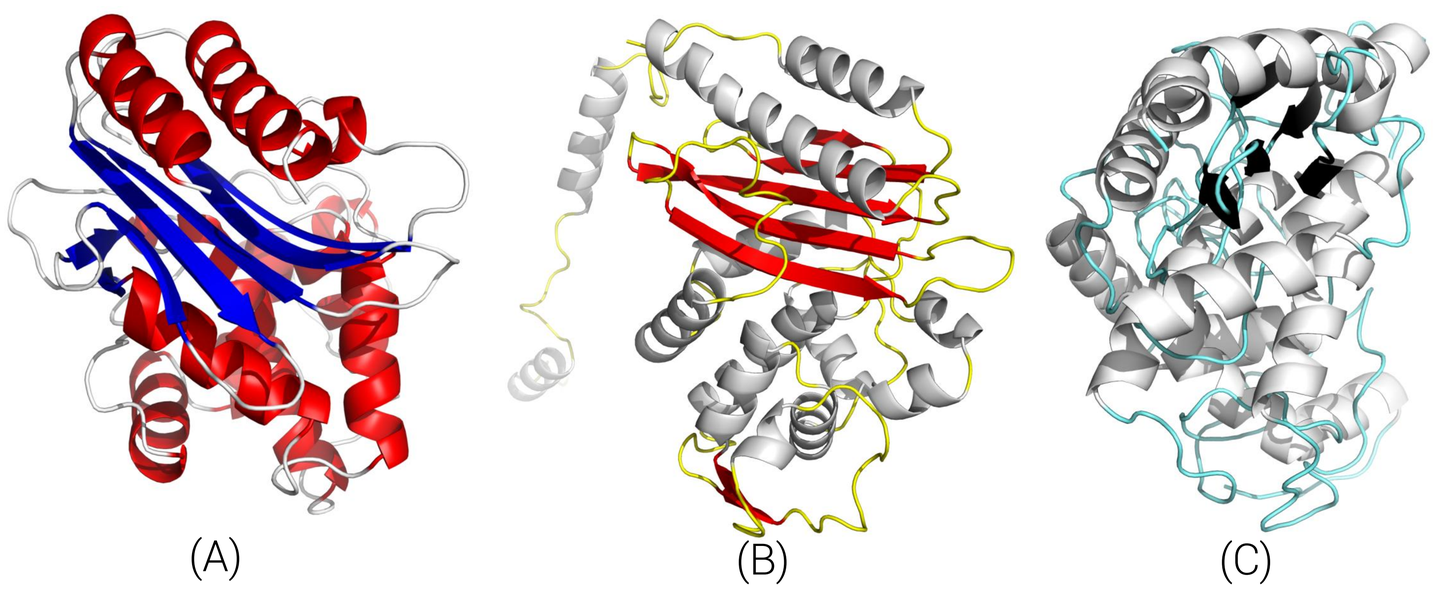

Figure 6: TEM-1, one of the most widely spread class A beta-lactamases (A), candidate from *Akkermansia* (36% identity with the class A beta-lactamases VEB-1 modelled with a reference class A beta-lactamase template (B) and modelled with a negative reference (C). This candidate was predicted as a class A beta-lactamase with 99.9% confidence.

Conclusion

- PCM appears to be efficient in predicting ARDs, even when aminoacid identity is low and to be more sensitive than conventional one-dimensional methods.
- The Human gut resistome was associated with gene richness and enterotypes. Our findings open perspectives in deciphering the variable response of the gut microbiota to antibiotics