Poster 259

Pairwise comparative modelling for identification of class A beta-lactamases in the Human intestinal microbiota

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

[1]Metagenopolis, INRA, Jouy-en-Josas, France, [2]INSERM, Univ Paris Diderot, France

Abstract

Background. Metagenomics have revolutionized the apprehension of complex microbial environment. However, the functional annotation based on sequence homology remains challenging. Indeed, the low identity shared by some predicted proteins with known proteins hampers their assignation to a family. Herein, we propose a new 3-dimensional approach, pairwise comparative modelling (PCM) consisting in 3-D homology modelling of a candidate protein using three templates matching the two closest protein families. PCM was applied to the identification of class A beta-lactamases (blaa) in the Human intestinal microbiota (MetaHIT study).

Methods. Potential blaa were searched in a non-redundant protein catalogue (n=3.9M) using a blaa reference dataset (n=676) and the Hmmer, Blastp and SSearch softwares. Candidates were modeled using Modeller with in parallel (i) blaa and (ii) negative references templates. Besides, a set of 60 blaa and 42 negative references were submitted to the same process. Modelling and alignments yielded various scores that were used to build a logistic regression that was subsequently applied to predict blaa among the candidates.

Results. We identified 212 putative blaa candidates which displayed at least 25% identity with any blaa. Only 18 (8.5%) had >95% identity with reference blaa. Most of the candidates shared identity with blaa recovered from functional metagenomic studies (38.7%) or with blaa from anaerobic bacteria (36.8%). PCM predicted 167 blaa while conventional annotation based on eggNOG v3 predicted 194. Indeed, eggNOG v3 predicted 30 blaa that were not identified so by PCM; structural alignment showed that only 3 (10%) had a glutamate close to reference position 166, that is essential to the blaa function. Thus, PCM was likely more specific than eggNOG v3 for blaa prediction.

Conclusions. PCM was consistent with eggNOG for blaa positive identification, but also appeared to be more specific as it invalidated a blaa assignation given by eggNOG in 13.2% cases. PCM is a promising approach in metagenomic investigations.

Background

- The intestinal microbiota is a vast reservoir of genes with low identity with known genes, including those conferring antibiotic resistance.
- Current methods based on sequence homology might not be able to discriminate true antibiotic resistance determinants from homologs with distinct function.
- Functional validation is the gold standard yet thousands of putative ARDs should be tested.
- New methods are necessary: we tested a new protocol based on homology modelling on a widely spread ARD family: class A beta-lactamases (blaa).

Materials and Methods

- Pairwise comparative modelling (PCM) is a based on a comparison between two modelling paths of one candidate: one path uses a reference template, the other uses a negative reference template that does not share the function but some sequence identity with candidate (Figure 1).
- PCM is first applied on reference set of blaa and negative references to build a model by logistic regression. (Figure 2). Then, the model quality predictor 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 a blaa) is obtained.

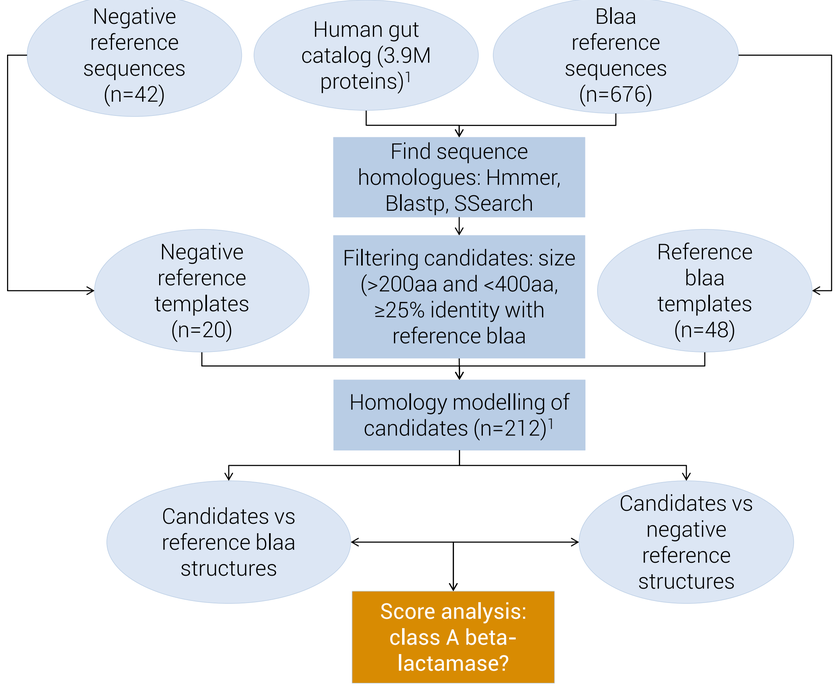


Figure 1: Concept of pairwise comparative modelling applied to class A beta-lactamases (antibiotic resistance determinants). [1]Nielsen, H. B. and Almeida, M. et al. Nat. Biotechnol. 32, 822–828 (2014).

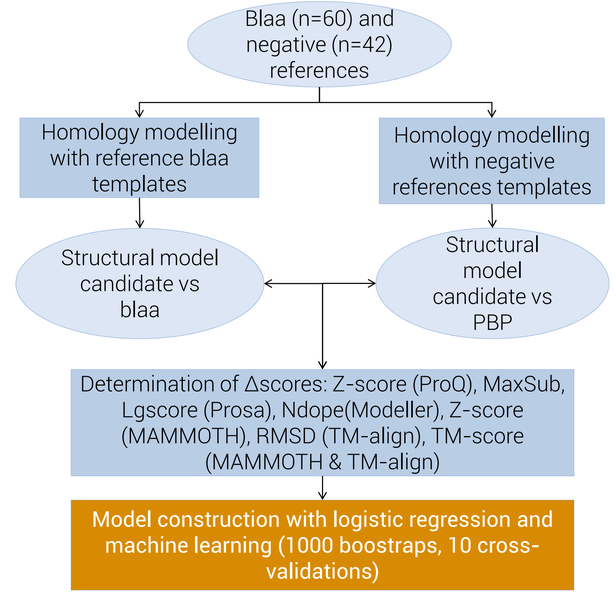


Figure 2: Construction of the logistic regression based on the different model quality predictors and alignment scores obtained with the blaa and negative reference sequences

Results/discussion

212 blaa candidates were identified: 167 were predicted as blaa with >50% confidence (Figures 3 and 4).
Most of the genus (69%) of the bacteria with predicted blaa could not be assigned (Figure 5).

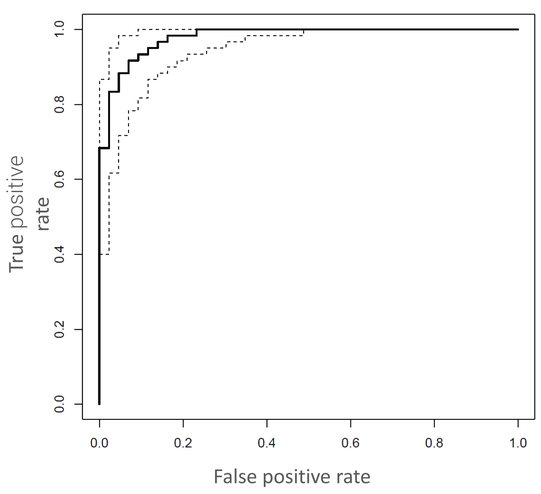


Figure 3: Receiver operating characteristic (ROC) curve build with the PCM results of the class A beta-lactamases and negative references. Area under curve was 0.98.

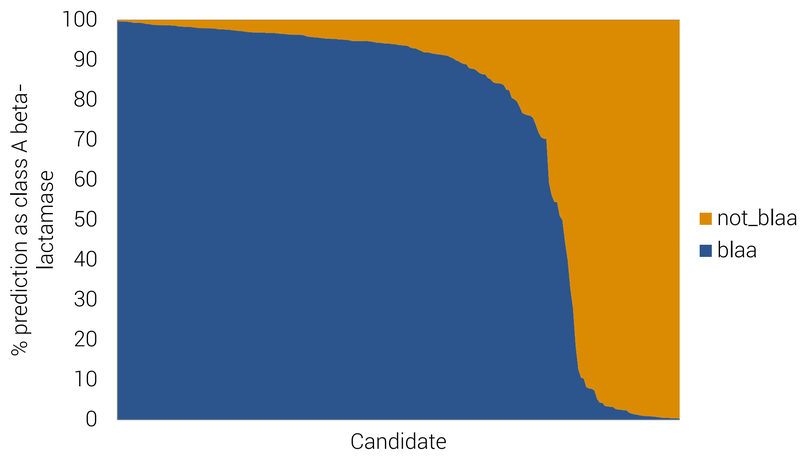


Figure 4: Receiver operating characteristic (ROC) curve build with the PCM results of the class A beta-lactamases and negative references. Area under curve was 0.98.

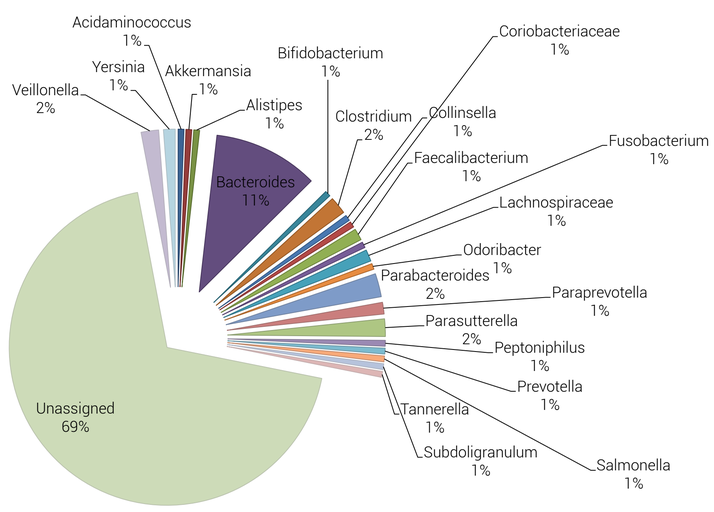
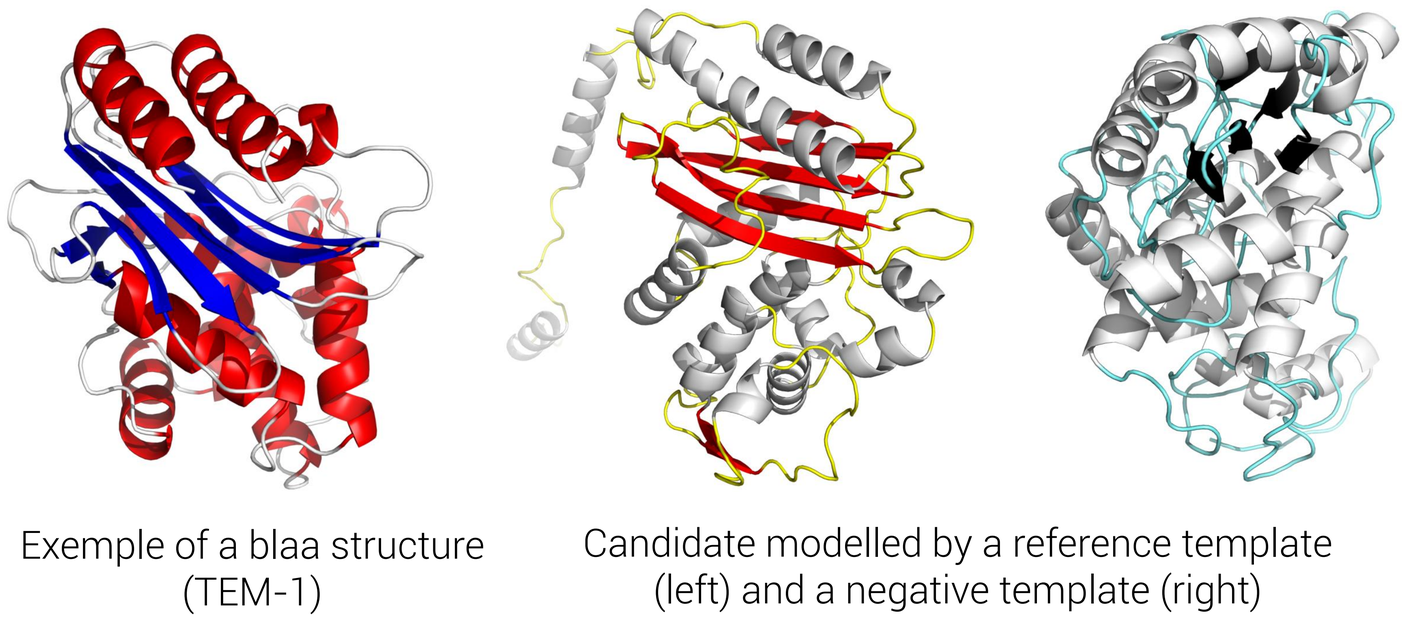


Figure 5: Distribution of genus in which a class A beta-lactamase was predicted.



Exemple of a blaa structure (TEM-1)

Candidate modelled by a reference template (left) and a negative template (right)

Figure 6: Structures of TEM-1, one of the most widely spread class A beta-lactamases, structure of a candidate from *Akkermansia* (36% identity with the class A beta-lactamases VEB-1 modelled with a reference class A beta-lactamase template (left) and modelled with a negative reference (right). This candidate was predicted as a class A beta-lactamase with 86% confidence.

		eggNOG v3 prediction		
		blaa	not blaa	Total
PCM prediction	blaa	164	3	167
	not blaa	30	15	45
	Total	194	18	212

Table 1: Comparison of the prediction obtained by PCM and eggNOG v3 (http://eggnog.embl.de/version_3.0/) which is based on a blast search on proteins clustered in orthologous groups. EggNOG assigned a blaa function to more candidates (194) than PCM (167). Among the 30 discrepancies, 27 were not found to harbour a glutamate in the close environment of the reference position 166, which is essential for class A beta-lactamase function.

		Identity >80% prediction[2]		
		blaa	not blaa	Total
PCM prediction	blaa	25	142	167
	not blaa	0	45	45
	Total	25	187	212

Table 2: Comparison of the prediction obtained by PCM and a 80% identity threshold, such as previously used [Forslund, K. et al. Genome Res. 23, 1163–1169 (2013); Hu, Y. et al. Nat. Commun. 4, 2151 (2013).]. Using a high threshold of identity likely allows to be more specific (no false positive) but might not encompass the diversity of proteins of complex environmentrs such as the intestinal microbiota (many false negative). In this case, as much as 142 candidates might have been misidentified as not blaa.

Conclusion

- PCM appears to be efficient in predicting class A beta-lactamases, even when aminoacid identity is low. For this family, PCM appeared to be more specific than conventional one-dimensional methods.
- PCM shall now be validated with other families and well-characterized datasets.

MetaGenoPolis
Centre de recherche INRA Jouy en Josas
Domaine de Vilvert – Bât. 325
78 352 Jouy-en-Josas CEDEX
France