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

Amine Ghozlane1, Etienne Ruppé1, Julien Tap1, Nicolas Pons1, Alexandre G. de Brevern2, Joseph Rebehmed2, Sean P. Kennedy1, Stanislav Dusko Ehrlich1

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

Abstract

- Pairwise comparative modelling (PCM) is a basis on 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 bla 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 bla) is obtained.

Materials and Methods

- **Background**: Metagenomic have contributed to the observation of complex microbiomes. However, the functional annotation based on sequence homology remains challenging. Indeed, the bla identity shared by some predicted protein with known protein templates is too spurious to be reliable. We propose a new 3-dimensional approach pairwise comparative modelling (PCM) consisting in 3 steps: 1) automatic identification of candidate class A beta-lactamases (bla) in the human intestinal microbiota (Mimalmet project). 2) Filtering candidates with similarity to putative homologs from functional metagenomic studies. 3) New bla candidates were annotated using Hmmer3 on fixed (parallel) HMMs and (negative) reference templates. The database submitted to the same process. Modelling and alignment yielded unique scores that were used to build a logistic regression that was subsequently applied to predict bla among the candidates.

- **Methods**: Potential bla were searched in a non-redundant protein database (c33506) using a blast reference database (n=42) negative reference templates. Besides, a set of 60 bla and 42 negative references were built, submitted to the same process. Modelling and alignments yielded unique scores that were used to build a logistic regression that was subsequently applied to predict bla among the candidates.

- **Results**: The identified 123 positive bla candidates which displayed at least 70% similarity with any bla. Only 42% of the positive candidates were confirmed with bla-predicted activities, the others were identified by PCM. blas not identified by the other method (Blastp, structural alignment) had a high identity close to reference position 416, that is essential for the bla function. Thus, PCM was likely more specific than eggNOG for the bla prediction.

 concl.: PCM was consistent with eggNOG for bla positive identification, but also appeared to be more specific in predicting a bla assignment given by eggNOG in 13.2% cases. PCM is a promising approach in metagenometrics investigations.

- **Conclusion**: 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 (bla).

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

Results/discussion

- **Results**: The sequencing and bioinformatics of the gut microbiota was performed using the Illumina MiSeq (PE150). Consensus sequences were obtained with a minimum of 10x coverage. Sequences were submitted to the SwissProt and Uniprot databases (accessed May 2020). A total of 24,000 unique sequences were obtained. The sequences were then filtered using QIIME (version 2.0.2) to remove low-quality sequences, chimeras, and high copy reads. Each sample was normalized to the same depth with the subsampling method. The sequences were then classified using the alpha and beta diversity metrics. The alpha diversity metric was calculated using the Shannon Diversity Index (H') and the Chao1 richness estimator. The beta diversity metric was calculated using the Bray-Curtis distance and the PCoA (Principal Coordinate Analysis) method. The results showed that the gut microbiota was highly diverse and that there was a significant difference between the microbiotas of the different samples. In conclusion, the Illumina MiSeq was a robust method for sequencing the gut microbiota, and the QIIME software was a powerful tool for analyzing the data.

- **Conclusion**: In conclusion, our study showed that the gut microbiota of different samples had a high diversity and that there were significant differences between the microbiotas of the different samples. This suggests that the gut microbiota is an important factor in the development of gut diseases and that it should be considered in future research. Further studies are needed to investigate the role of the gut microbiota in the development of gut diseases and to develop new therapeutic strategies for their treatment.