Article

Overview of the proteome of *Campylobacter iguanorium*

Arina Danilina[1,*]

[1]Faculty of Bioengineering and Bioinformatics, MSU.

ABSTRACT

Campylobacter iguanorium strain 1485ET genome has been obtained through the NCBI genome browser (https://www.ncbi.nlm.nih.gov/genome/36237). The feature table of the proteome has been analyzed using Excel. The proteins (expression products) and genes have been categorized by their stated role in the organism. Several illustrations showing the properties of the bacteria's proteome products have been created using Excel as well.

Keywords: *Campylobacter iguanorium*, Proteome.

1 INTRODUCTION

Campylobacter iguanorium is a recently discovered species, isolated from reptiles. The particular strain 1485ET has been isolated from a Bearded Dragon (*Pogona vitticeps*). Its whole genome was sequenced for first time in 2014 by a team from Utrecht University and the U.S. Department of Agriculture. *C. iguanorium* species is genetically related to *C. fetus* and *C. hyointestinalis*. (PubMed article: https://www.ncbi.nlm.nih.gov/pubmed/25146144). However, despite colonizing a shared host, no recent recombination between *C. iguanorium* and *C. fetus* was detected [see genome announcement].

2 METHODS

The sources file has been obtained from the NCBI genome browser. The exact file link: ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/736/415/GCA_000736415.1_ASM73641v1 .Excel has been used for the editing of the file. The product category (e.g. pseudogene, protein coding or RNA coding) was stated in the class column. In the strand column the value "+" stood for the main strand, and "-" for the complementary one.

3 RESULTS AND DISCUSSION

The size of *C. iguanorium* genome is 1,684,608 bp, with 1831 genes (annotated by the team that first sequenced the genome).
This would suggest the ratio of approximately 1087 genes per 1 million bp. Furthermore, as the result of data processing the properties and categorization of the bacteria's proteome products have been visualized.

3.1 The number of proteins of certain length.

The histogram shows the number of proteins based on their length. The length values have been joined into bins with the range of 10 amino-acids. As you can see from Fig. 1., the most frequently occurring protein length lies in the range from 70 to 280 amino-acids. As typical for bacteria, the most common protein length of 190 aa is situated closer to the short end of the specter. This might indicate that the greater portion of the bacteria's proteins could lack a stable three-dimensional structure. However, a fair number of significantly longer proteins is also present. The conformation of the latter is, most likely, stable.

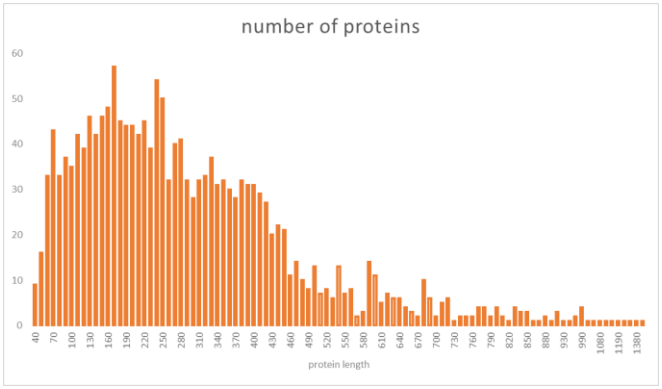

Fig. 1. Relation between the length of proteins (X axis) and their quantity in the proteome (Y axis).

3.2 The number of genes of certain class depending on the strand

The protein coding and pseudogenes on the main and complementary strand had "protein_coding", "pseudogene" and "+" and "-" in the columns class and strand respectively. As you can see form Table 1., the prevailing gene class on both strands is protein coding. A mentionable feature is that there are more pseudogenes and protein coding genes on the main strand, while there are more RNA coding genes on the complementary strand than on the main one. This might be due to the differences in post-transcription processes of the various biopolymer types and the difficulties associated with transcription of a gene from the complementary strand. The latter could also possibly increase the likelihood of transcription errors. This is why, perhaps, protein coding genes are mainly situated on the main strand, while the RNA coding are on the complementary. The reason for those specific classes to be situated this way is, arguably, the fact that proteins are supposed to exist in the

*To whom correspondence should be addressed.

cell for a longer period of time compared to RNAs. It could mean that possible errors are less crucial if made in RNA sequences.

Table 1. The number of pseudogenes, protein and RNA coding genes on both DNA strands.

	protein coding	pseudogene	RNA coding
main strand	939	12	22
complementary strand	818	6	34

3.3 The general number of genes of certain class

The protein, rRNA, tmRNA and tRNA coding genes had “protein_coding”, “rRNA”, “tmRNA” and “tRNA” in the class column respectively. As you can see from Table 2., the majority of the genes are protein coding. The number of genes coding a certain biopolymer clearly correlates with the number of types or varieties of said biopolymer. Table 2. data is consistent with this statement.

Table 2. The number of protein and RNA coding genes in the genome

	protein coding	rRNA	tmRNA	tRNA
quantity	1757	9	1	44

4 CONCLUSION

The results of proteome analysis described in this article could be used for further research into the bacteria *C. iguanorium* and its metabolic pathways. They also might be used in comparative genomics to discover possible genus-specific features of the genome.

SUPPLEMENTARY MATERIALS

- The Excel table (http://kodomo.fbb.msu.ru/~arinadanilina/term1/pr13.xlsx).

ACKNOWLEDGEMENTS

The author thanks all the professors involved for explaining all the material necessary for writing this article.

REFERENCES

NCBI Genome browser (https://www.ncbi.nlm.nih.gov/genome/).
PubMed (https://www.ncbi.nlm.nih.gov/pubmed/).
Genome announcement (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4153485/).
Comparative genomics article (https://www.ncbi.nlm.nih.gov/pubmed/27604878/).