Article

Overview of Dyadobacter fermentans DSM 18053 proteomics

S. Starikov[1]

[1]Department of Bioengineering and Bioinformatics, Moscow State University

ABSTRACT

Many important problems in biology require superb skills in Excel usage. The importance of these skills correlates highly with the size of databases you need to conduct an analysis on, making this sorcery absolutely essential for those interested in bioinformatical analysis. Here we present an example of what can be achieved by an average sleepy, hungry, exhausted, easily distracted, absolutely indifferent student in an hour of work in Excel.

1 INTRODUCTION

We have conducted research on *Dyadobacter fermentans* DSM 18053 because we were doomed to (I suspect that the distribution of the organisms given to students to analyze depended solely upon the lexicographical order of our names). The study covered the protein length distribution and the numbers of genes encoding proteins and RNAs, providing thorough insight into great complexity of bacterial genome. Just kidding, the results just show us some statistical distribution, giving us a vague idea of what length of protein is the most common and how many products can a rather compressed genome encode. I will use word 'bacteria' as referring to *Dyadobacter fermentans* DSM 18053 throughout the article.

2 METHODS

The whole lot of work has been done with the invaluable help of Excel 2007. Of the endless variety of functions and methods provided by this wonderful soft, I used a range that could have been easily fit into an average 1st grader's head in a mere couple of minutes – MAX, MIN, Sorting, Histogram and Compressed table. Though I'm bound to say that my superiority over this kid is obvious – he would have needed a supplementary course in internet browsing to get the data I analyzed – [1]. BUWAHAHAHA.

3 RESULTS

To make our minor study seem meaningful and remarkable, we divided the section into subsections.

3.1 Protein length distribution

After long hours of sophisticated analysis of this bacterial proteomics and through dogged determination I came to an implication that the commonest length of protein encoded in the DNA is approximately 250 bp (see Figure 1). A better estimate could never

be made because changing the bids would have taken a couple of minutes.
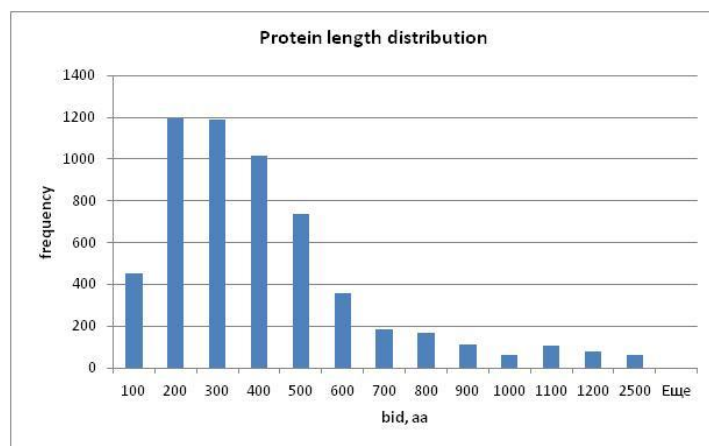

Figure 1. Protein length distribution

3.2 Genome analysis

This subsection contains modest analysis of genes on positive and complementary strands of bacterial DNA. CDSs code for proteins, while RNAs code for RNAs.

Table 1. The numbers of genes encoded on DNA strands

Gene	Encoded in positive strand	Encoded in complementary strand
CDS	2888	2830
RNA	29	22

4 DISCUSSION

The results are rather predictable. The protein length distribution does not follow any pattern – looks random. There is also a limit to protein length – 2500 amino acids. That can be easily explained by the combination of facts: a huge protein needs a lot of material; bacterial cell doesn't really need something that huge. Gene distribution over different strands looks random too – and that is also understandable – transcription works equally for both strands, which means there is no base for discrimination between strands.

This footnote exists just because I have no idea how to delete it.

---

5   SUPPLEMENTARY MATERIALS

Excel file

ACKNOWLEDGEMENTS

REFERENCES

[1] ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/Dyadobacter_fermentans_DSM_18053_uid59049/ - link to NCBI database