Evaluation of intergene distances across bacterial species

Laboratory for information systems
Rudjer Boskovic Institute, Croatia

Skunca N., Supek F., Repar J. and Smuc T.

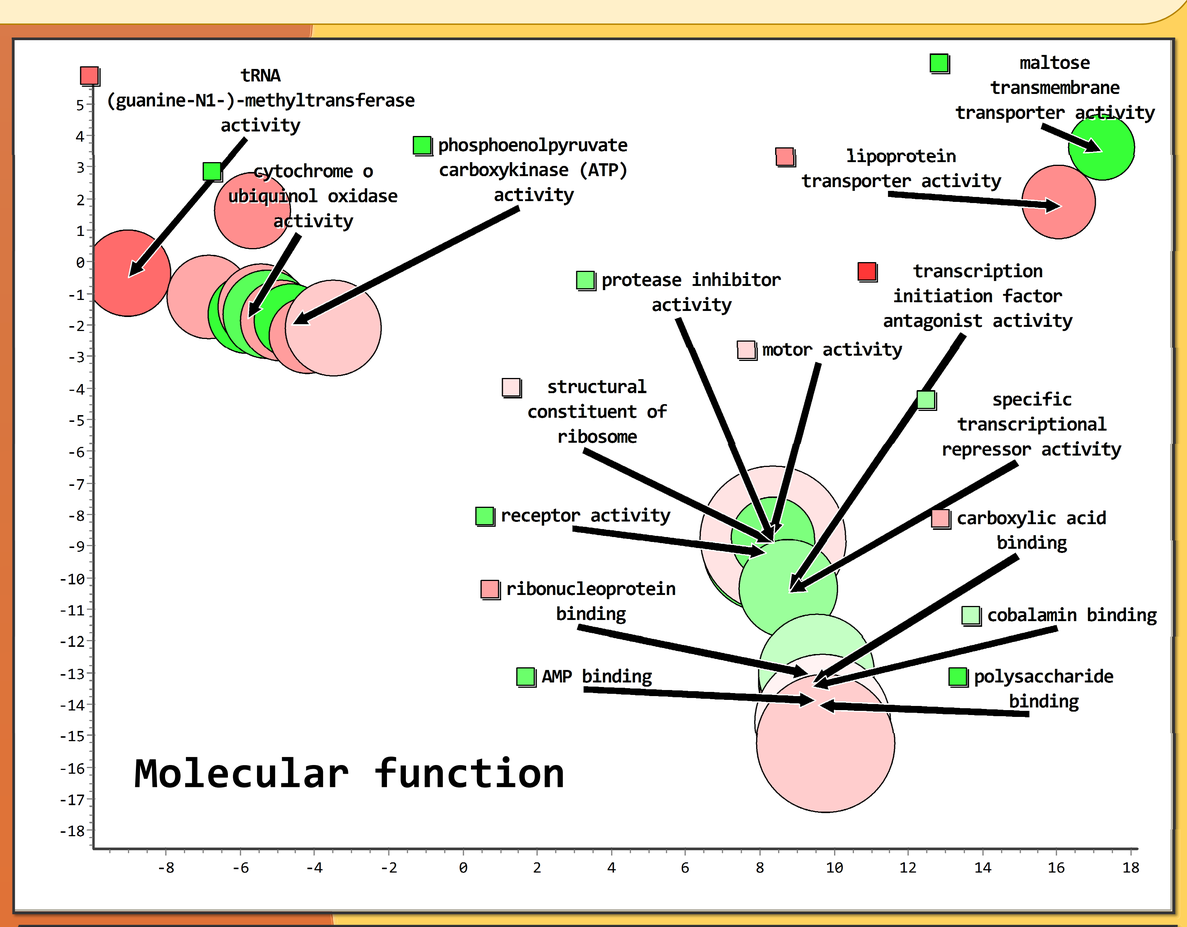
1. Introduction

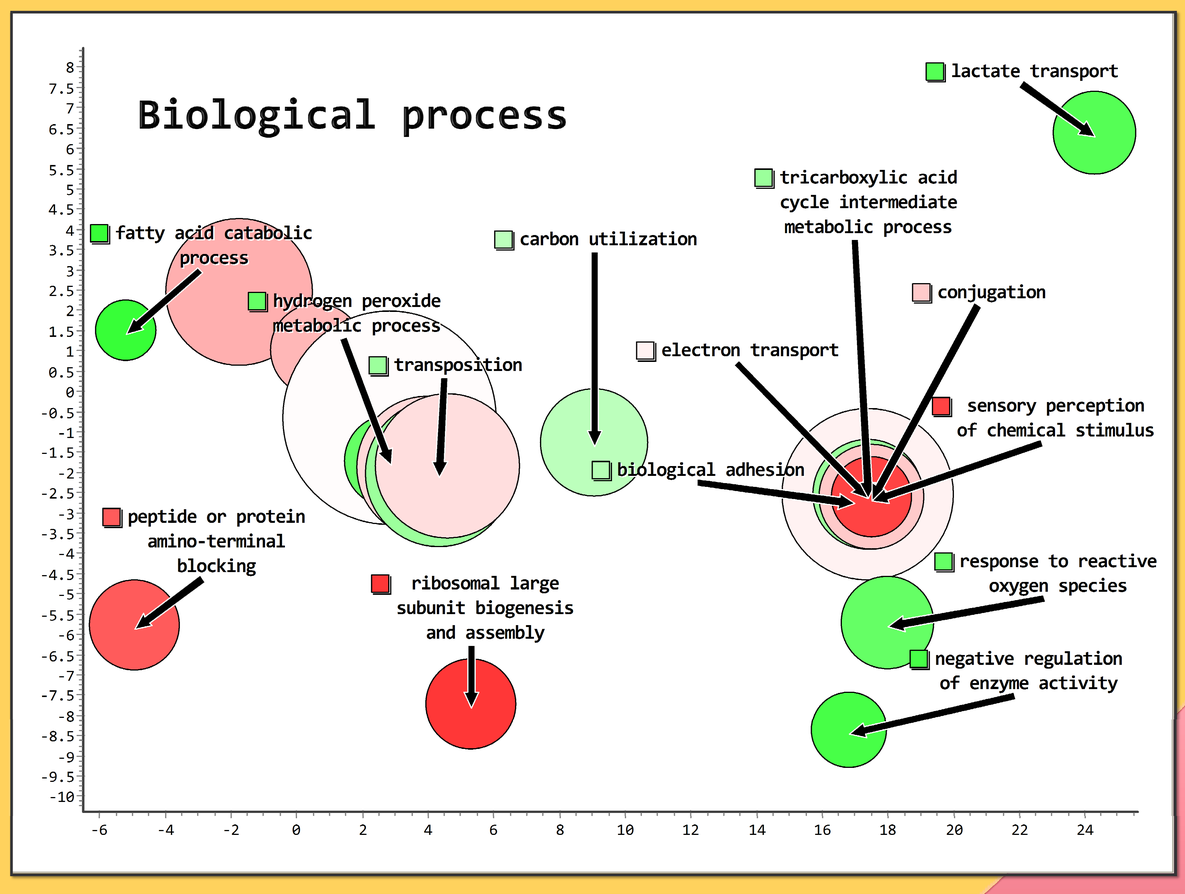
Transcription regulation regions of constitutively expressed genes are one of the points where the prokaryotic cell attempts to reduce its genome, in order to optimize the usage of resources.

We have analysed non-coding regions (NCRs) of prokaryotic genomes in order to connect gene function – in light of Gene Ontology (GO) assignments – with the length of the preceding non-coding region.

2. Materials and methods

- **Division into two NCR length groups:** long (above average - approx. 33% of genes) and short (below average - approx. 67% of genes).
- **Statistical analysis –** using a number of Fisher's exact tests for correlation between two categorical variables – first, short or long NCR; and second, member or non-member of a Gene Ontology category.





3. Results

Our results are visualized in three charts, each representing one ontology. The X and Y axis values are calculated using a semantic similarity metric SimRel. The size of the circle is proportional to the GO group size. The colour of the GO category circle shows enrichment in genes having:

short NCRs — long NCRs

4. Conclusions

The results shown point out that functions one would expect to be under stringent control, and expressed only when the cell is in the conditions requiring the function of these particular proteins, are those having **long** NCRs, and GO categories enriched in **short** NCRs describe basic cell functions.

Acknowledgments

NS gratefully acknowledges the support of the ECCB08 volunteer fellowship in covering the trip and lodging expenses. NS, FS and JR appreciate the funding recived from the Croatian Ministery of Science, Education and Sport in partially funding attendance of the ECCB08 conference.

Literature

- Lynch M. **Streamlining and simplification of microbial genome architecture.** *Annu. Rev. Microbiol.* 60:327–49 (2006)
- The Gene Ontology Consortium. **Gene Ontology: tool for the unification of biology.** *Nature Genet.* (2000) 25: 25-29
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. **GenBank**. *Nucleic Acids Res.* 36(Database issue), D25-30 (2008)
- Manly FJB. **Randomization, Bootstrap and Monte Carlo Methods in Biology**, 3rd ed., 2007
- Schlicker A, Domingues FS, Rahnenführer J, Lengauer T. **A new measure for functional similarity of gene products based on Gene Ontology.** *BMC Bioinformatics* 7:302 (2006)