

READ ME

„The genetics behind the evolution of a community of four bacterial species“

Bachelor thesis at the Biology Department, Science Faculty, University of Bern

Written by Sabrina Michaela Rasch

Supervisor: Dr A. Kapopoulou, Institute of Ecology and Evolution,
Bioinformatician Theoretical Ecology and Evolution Lab

Prof. Dr C. Bank, Institute of Ecology and Evolution, Head of the Division
Theoretical Ecology and Evolution

Abstract

Microorganisms are not only a part of our daily lives but are also widely used in industry. For example, the co-cultured community of *Agrobacterium tumefaciens*, *Comamonas testosteroni*, *Microbacterium saperdae* and *Ochrobactrum anthropi* is very efficient and reliable for degrading/detoxifying substances used for cutting or grinding metals, such as metal working fluid (MWF). In this study, I consider the selective pressure on single-nucleotide polymorphisms (SNPs) as I investigate the genetics of this community. For this purpose, the microbes were allowed to grow in different combinations in the MWF for 44 weeks, and the genomes of the community were regularly sequenced for SNPs. 41 SNPs became fixed, meaning they had a frequency above 0.98 in the population. Of these fixed SNPs 17 are intergenic, and the remaining 24 belong to 16 different genes. The function of these 16 genes and their relevance in the degradation/detoxification of MWF are discussed. Furthermore, the SNPs found at the same position as a fixed SNP in another replicate of the experiment were examined. The community where the additional SNPs occurred, whether they were non-synonymous or synonymous, and a line graph showing SNP frequency at each time point of sequencing were investigated. Finally, the selection pressure exerted on the mutations was investigated by examining the ratio of non-synonymous to synonymous mutations.

1 Data

All data that was used for the study.

1.1 Original

Original data as given by my supervisor.

<i>allSNPs_w_gene_position.tsv</i>	Data of all SNPs obtained by the lab of Sara Mitri
<i>alltogether_modified.gen.fasta.gen</i>	Nucleotide sequence of all genes
<i>alltogether_modified.lst.txt</i>	Data of the coordinates of the genes
<i>alltogether_modified.prt</i>	Amino acid sequence of all genes
<i>FixedMutations_over_time.pdf</i>	Plot of number of fixed mutations per microbe
<i>FixedMutation_over_time_Mado.pdf</i>	Plot of frequency of fixed mutations over time
<i>FixedMutations_OverTime.txt</i>	Data of 41 fixed SNPs

1.2 VSC manipulated

Some files had to be manually changed to be able to work with them in R Studio.

<i>allSNPs_w_gene_position_spaced.tsv</i>	data of all SNPs
<i>alltogether_modified.gen.fasta_spaced.gen</i>	DNA sequence of all genes
<i>alltogether_modified.lst_spaced.txt</i>	Protein name with start/end
<i>alltogether_modified_spaced.prt</i>	AA sequence of all genes

2 Papers

All papers that were cited in the report.

3 Results

Output of the “Code.Rmd” as .csv and .png. Additionally, Excel files are provided where the results are summarised.

3.1 1_SNP localisation and protein function

Which genes contain fixed mutations, and what is the function of the protein they code for?

<i>Question_1.xlsx</i>	Summary in a Excel file
<i>Result_AA_sorted.csv</i>	Amino acid sequence to genes with a fixed SNP
<i>Result_DNA_sorted.csv</i>	Nucleotide sequence to genes with a fixed SNP
<i>Result_Genes_sorted.csv</i>	Genes with a fixed SNP

3.2 2_Mono- vs co-culture

Do we see the same mutations in the mono- and co-culture replicates?

<i>Plots</i>	Subfolder with the line plots of the frequencies of the SNPs at the same position as a fixed SNP
<i>Question_2.xlsx</i>	Summary in a Excel file
<i>SNPX.csv</i>	All SNPs that are in the same position as a fixed SNP. X is the number of the SNP and can be 1, 3, 4, 5, 6, 7, 8, 11, 12_15, 16, 18, 29, 24, 27.

3.3 3_SNPs under selective pressure

Which genes are under positive/negative selection?

<i>Gene_to_SNPX.csv</i>	All SNPs that are in the same gene as a fixed SNP. X is the number of the SNP and can be 1, 3, 4, 5, 6, 7, 8, 11_12_15_19_24_27, 16, 18, 30, 31, 32, 33, 34, 38_39_40_41
<i>Question_3.xlsx</i>	Summary in a Excel file
<i>Result_dNdS_microcosm.csv</i>	Table with the dN/dS calculation

4 Thesis and Code

The Report and the annotated code in different formats.

<i>Appendix.pdf</i>	The Appendix to the report
<i>Bachelor_thesis.pdf</i>	Bachelor project report
<i>Code.html</i>	Annotated code used for the analysis
<i>Code.Rmd</i>	Annotated code used for the analysis