



BIOINFORMATICS RESEARCH CENTER (BiRC)  
*Aarhus University*

MASTER THESIS

---

# Inferring population genetics on sediment communities using a metagenomical approach

---

AUTHOR:  
KRISTIAN RIIS HANSEN  
201805452

SUPERVISORS:  
THOMAS BATAILLON  
KASPER URUP KJELDSSEN

*Date:*  
JUNE 15, 2023



# Table of Contents

---

<b>Table of Contents</b>	<b>ii</b>
<b>Abstract</b>	<b>iv</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Genome evolution in sediment bacteria . . . . .	1
1.2 Metagenomic approach . . . . .	2
1.3 Recombination in bacteria . . . . .	4
1.4 Linkage disequilibrium . . . . .	4
1.5 GC-biased gene conversion . . . . .	4
1.6 Polymorphic synonymous sites and Effective population size . . . . .	5
1.7 Selection regimes: Purifying selection & Positive selection . . . . .	5
1.8 Cluster of Orthologous Groups . . . . .	6
1.9 Objectives and hypotheses . . . . .	6
<b>Chapter 2 Methods and data</b>	<b>7</b>
2.1 Data . . . . .	7
2.2 Instrain . . . . .	7
2.3 Splitting contigs into 25kb windows . . . . .	8
2.4 Estimating recombination using Linkage decay plots . . . . .	9
2.5 Estimating selection on regions across MAGs . . . . .	10
2.6 Impact of linked SNVs on recombination estimates . . . . .	11
2.7 Determining gene categories using annotation files . . . . .	12
2.8 Selection and Filtering of MAGs for Analysis . . . . .	12
2.9 Comparing recombination levels across depths and bacterial classification . . . . .	14
2.10 Exploring the Association Between Recombination Rates and GC Content . . . . .	14
2.11 Exploring the relationship between Recombination and Selection regimes . . . . .	14
2.12 Investigating Recombination and Selection Patterns in Gene Categories . . . . .	15
2.13 Selection and recombination trends across COG categories . . . . .	15
2.14 Testing the relationship between polymorphic sites, effective population size and recombination estimates . . . . .	15
<b>Chapter 3 Results</b>	<b>17</b>
3.1 Comparing recombination levels across depths and bacterial classification . . . . .	17
3.2 Evidence for GC-biased gene conversion? . . . . .	19
3.3 Association between pNpS and recombination . . . . .	20
3.4 How are gene functional categories associated with levels of recombination and pNpS? . . . . .	21
3.5 Differential distribution of gene categories in proteobacteria based on selection regimes . . . . .	24
3.6 Comparative analysis of gene categories reveal possible evolutionary trends in sediment bacteria across depths . . . . .	26
3.7 Differential recombination and selection patterns in Proteobacteria across depths . . . . .	28
3.8 Association between effective population size and recombination . . . . .	30

---

3.9	Positive relationship between synonymous polymorphic sites and effective population size . . . . .	31
3.10	Association between recombination rates and polymorphic synonymous sites . . .	32
<b>Chapter 4</b>	<b>Discussion</b>	<b>33</b>
4.1	Mean recombination rates differ between BT-zone and nBT-zone . . . . .	33
4.2	pNpS and recombination shows overall positive relationship . . . . .	34
4.3	gBGC . . . . .	34
4.4	Gene categories do not differ between depths, but instead between MAGs on the genus level . . . . .	34
4.5	Relationship between Ne and recombination . . . . .	35
4.6	Polymorphic synonymous sites increases with effective population size . . . . .	35
<b>Chapter 5</b>	<b>Conclusion</b>	<b>36</b>
<b>Chapter 6</b>	<b>Future Work</b>	<b>37</b>
	<b>Bibliography</b>	<b>38</b>
	<b>Appendix A</b>	<b>41</b>

# Abstract

---

Sediment communities contain abundant and diverse microbial populations which are governed by environmental factors across sediment depth. Previous studies have tried investigating the evolutionary ecology of these sediment communities. However the ecological importance of recombination, selection regimes and the association between these remains unclear in sediment bacteria. Using a metagenomic approach we investigate intra- and extra species patterns of recombination and selection regimes between the bioturbated zone and non-bioturbated zone in sediments located in Aarhus Bay. Differential recombination levels and pN/pS ratios are observed between metagenome assembled genomes (MAGs) grouped by the two zones. The MAGs originating from the bioturbated zone exhibit higher rates of purifying selection and lower rates of recombination, which may be related to their environmental adaptability. Further investigations of recombination related population genetics are laid out, including gene biased GC conversion (gBGC) and the relationship between recombination and the effective population size ( $N_e$ ). As the method for calculating recombination and  $N_e$  are questionable, we suggest a more refined analysis be done. To further deepen our understanding of the ecological and evolutionary dynamics of sediment bacteria, we propose to include genome-wide expression analyses in combination with metabolic analyses.