

# READ ME

Sabrina Rasch

November 2, 2021

## Contents

<b>1</b>	<b>First Data</b>	<b>2</b>
1.1	Original Data . . . . .	2
1.2	Results . . . . .	2
1.3	VSC manipulations . . . . .	2
1.4	Other Files . . . . .	2
<b>2</b>	<b>Infos</b>	<b>3</b>
<b>3</b>	<b>Papers</b>	<b>3</b>

# 1 First Data

Contains Data and R Code that was used to solve the first Steps (see R Notebook "Step-by-step-final-version.Rmd" for more Information on the Steps)

## 1.1 Original Data

- allSNPs\_w\_gene\_position.tsv *data of all SNPs*
- alltogether\_modified.gen.fasta.gen *DNA sequence of all data*
- alltogether\_modified.lst.txt *Protein name with start/end*
- alltogether\_modified.prt *AA sequence of all data*
- FixedMutations\_over\_time.pdf *Plot*
- FixedMutation\_over\_time\_Mado.pdf *plot*
- FixedMutations\_OverTime.txt *41 fixed SNP positions*

## 1.2 Results

Contains .csv files that were output of "Step-by-step-final-version.Rmd", .xlsx where the .csv are in an excel table.

- Function\_of\_Proteins.pdf *Here we have a list of the proteins belonging to the SNPs with more information.*
- Pos\_AA.csv & Pos\_AA.xlsx *The aminoacid sequence to the proteins are listed*
- Pos\_DNA.csv & Pos\_DNA.xlsx *The nucleotide sequence to the proteins are listed*
- Result.xlsx *all tables in one file*
- Result\_sorted.csv & Result\_sorted.xlsx *The proteins belonging to the SNPs are listed*

## 1.3 VSC manipulations

Files where I had to change something manually to add the file into R are stored here.

- allSNPs\_w\_gene\_position\_spaced.tsv *data of all SNPs*
- alltogether\_modified.gen.fasta - Kopie.txt *DNA sequences belonging to the SNPs of AGTU.c01*
- alltogether\_modified.gen.fasta - not nicely.gen *DNA sequences belonging to the SNPs*
- alltogether\_modified.gen.fasta.AGTU.gen *DNA sequences belonging to the SNPs of AGTU.c01 but nicer spaced*
- alltogether\_modified.gen.fasta\_spaced.gen *DNA sequence of all data*
- alltogether\_modified.lst\_spaced.txt *Protein name with start/end*
- alltogether\_modified\_spaced.prt *AA sequence of all data*

## 1.4 Other Files

Files that are stored in the folder not belonging to another folder.

- By Hand.xlsx *Counted the nucleotides by hand to find the SNP reference from the gen.fasta data.*

- Confusion.xlsx *Excel table of not nicely search for the reference nucleotide from the gen.fasta data for the SNPs.*
- Step-by-step.Rmd *Code where I have everything in there (all trials done manually or with only little data).*
- Step-by-step-final-version.Rmd *This code can be used to solve the first Steps.*

## 2 Infos

Contains documents that have an informational purpose.

- Introduction paper.pdf *The first paper Mado gave me, that explains a lot about my data and why I do this project.*
- Meetings.docx *Notes from meetings with Mado*
- Presentation in Labmeeting 23.09. (Mado).pptx *PPP from Mado about her talk in the Labmeeting*

## 3 Papers

Contains papers that could be useful.

- N.Surname\_Year\_Title.pdf
- P. Piccardi\_2019\_Toxicity drives facilitation between 4 bacterial species.pdf