

Strain tracking for 1 month samples

Chunyu Zhao

01/24/2018

Contents

1	Introduction	1
1.1	Read in the meconium contigs information	1
1.2	Read in sample conversion	1
2	Read in base entropy and substitution rate from sunbeam_neutrino	1
3	Mismatch Rate	2
3.1	first let's read in the 1 month mismatch result	2
3.2	Proteobacteria Escherichia coli	2
3.3	Bacteroidetes Bacteroides vulgatus	4
3.4	Firmicutes Enterococcus faecalis	6

1 Introduction

- For each `SubjectID`, we mapped the corresponding 1mon sample reads back to the `contigs_meconium`.
- So, the contigs info (taxonomic assignment and 139 SCGs alignment) are the one for the meconium samples.

What's new here is: read in the table `sample_conversion_1month.txt`, and update the count/substitution rate.

1.1 Read in the meconium contigs information

1.2 Read in sample conversion

2 Read in base entropy and substitution rate from sunbeam_neutrino

The following are some cases:

- 1. Some `contigs_meconium`'s `aligned_SCG` (e.g. E coli): zero reads from 1mon mapper. What should we do about these genes?
- 2. There are also some samples's `contigs_meconium` (e.g. s272.STL.V02_1mo): `1mon_reads` mapped to the same contigs, but not the same region (aka gene)
- After talking with Kyle: what we are really interested is `mismatch rate`.
- 3. end product: `_hist_mismatch_rates.rda`

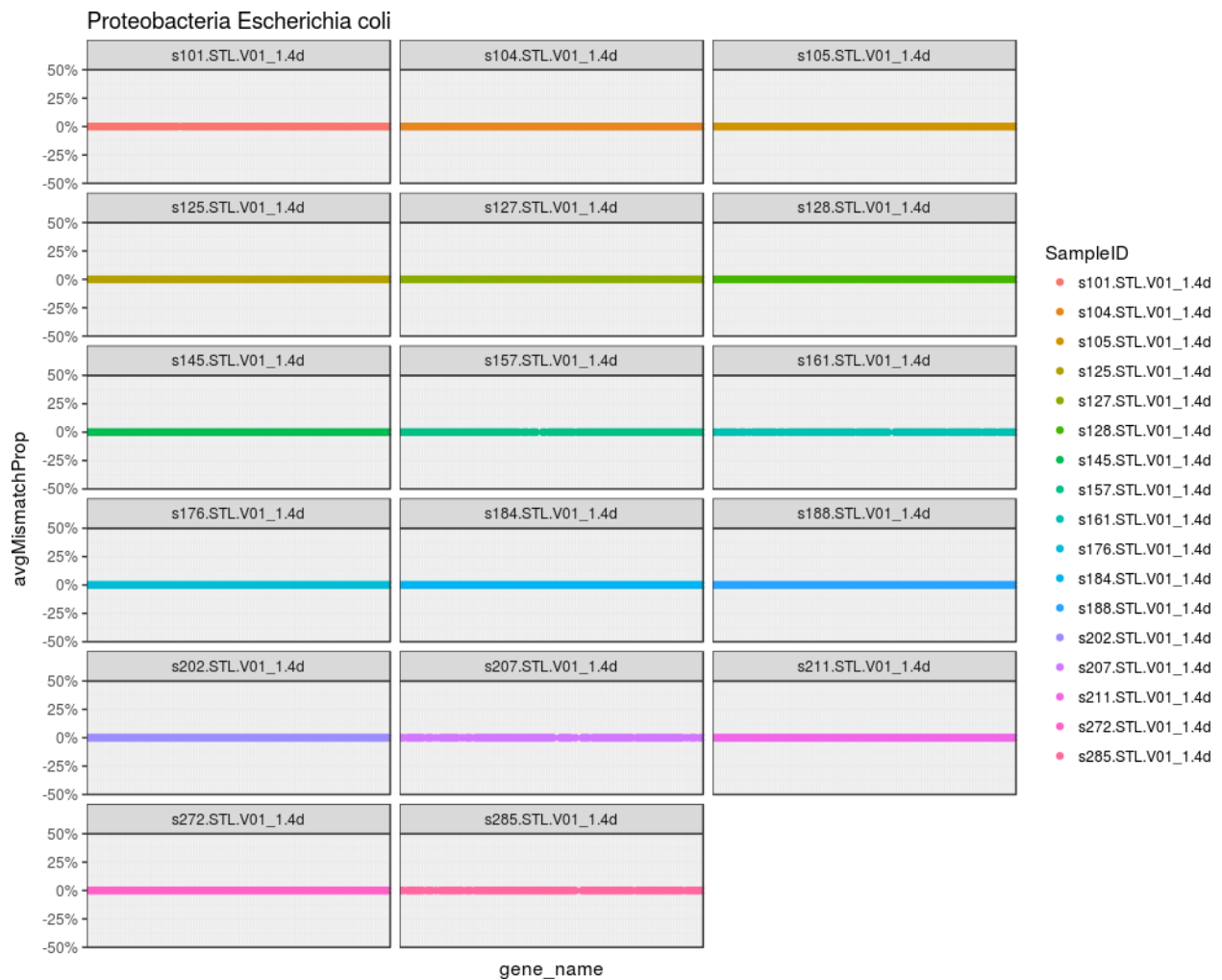
3 Mismatch Rate

3.1 first let's read in the 1 month mismatch result

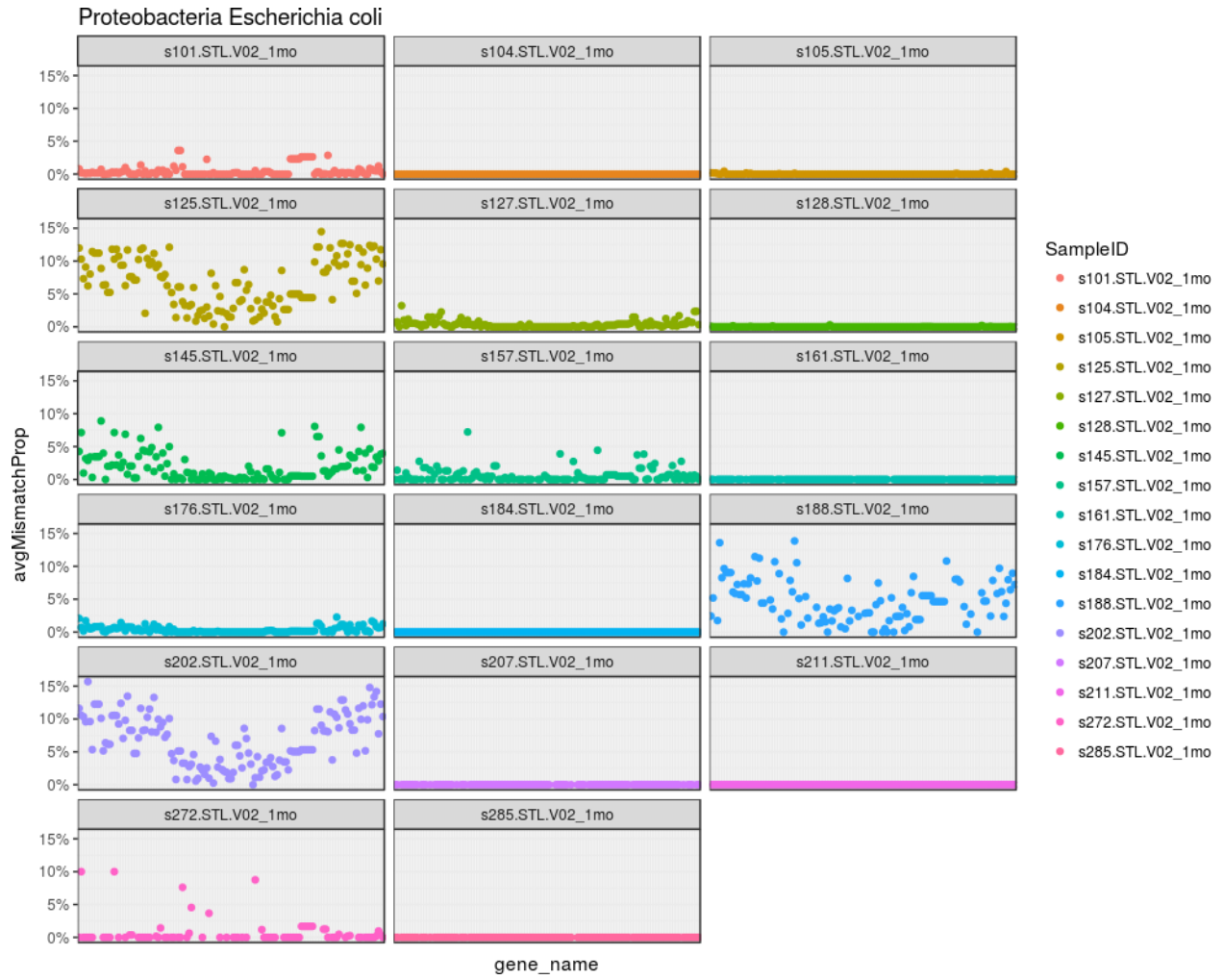
- one sample: multiple contigs aligned to different locations to the same gene => take proportion and aggregate into mean; however, once we aggregate, we won't be able to trace back the alignment length and etc.

3.2 Proteobacteria Escherichia coli

3.2.1 Birth

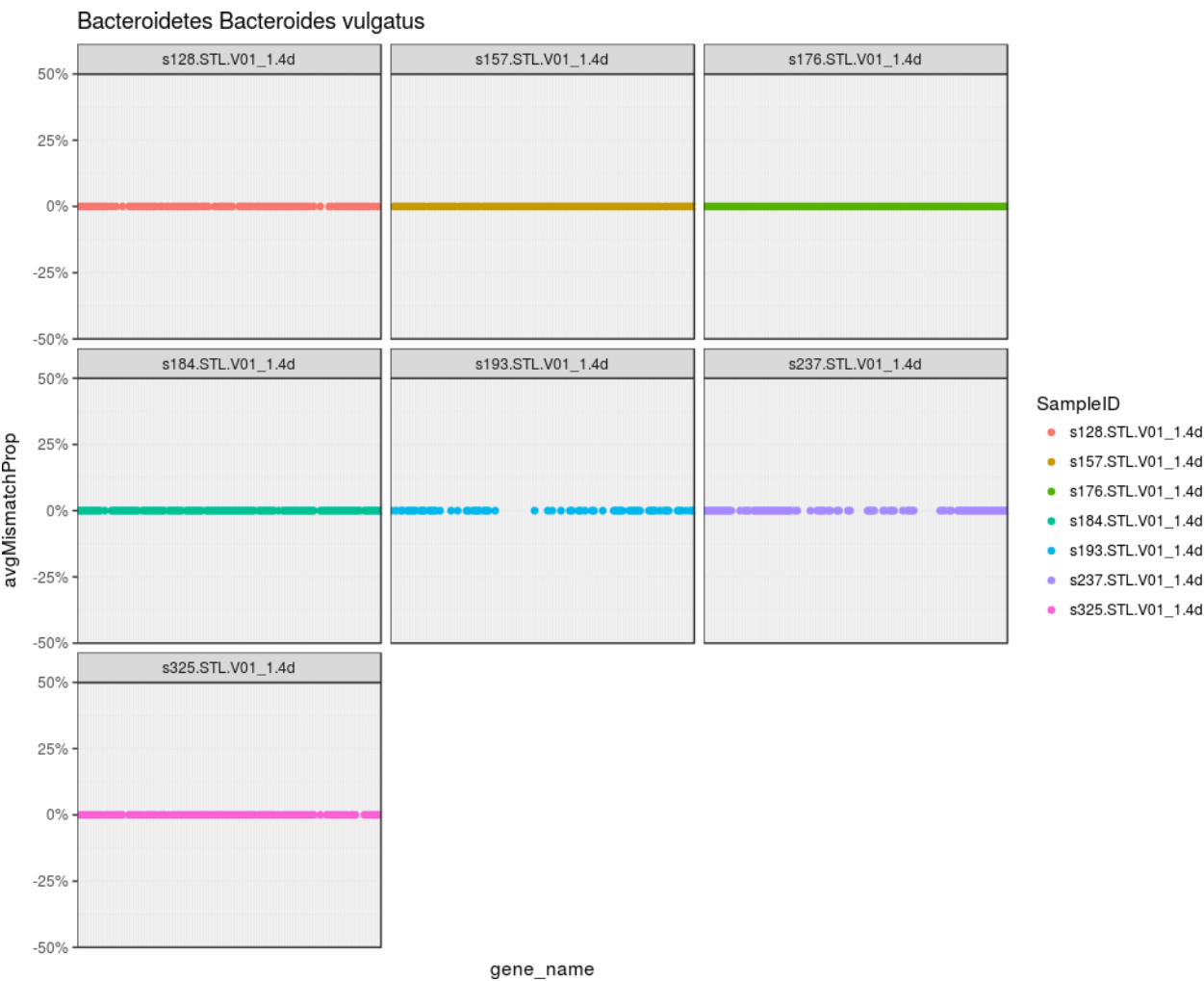


3.2.2 1 month

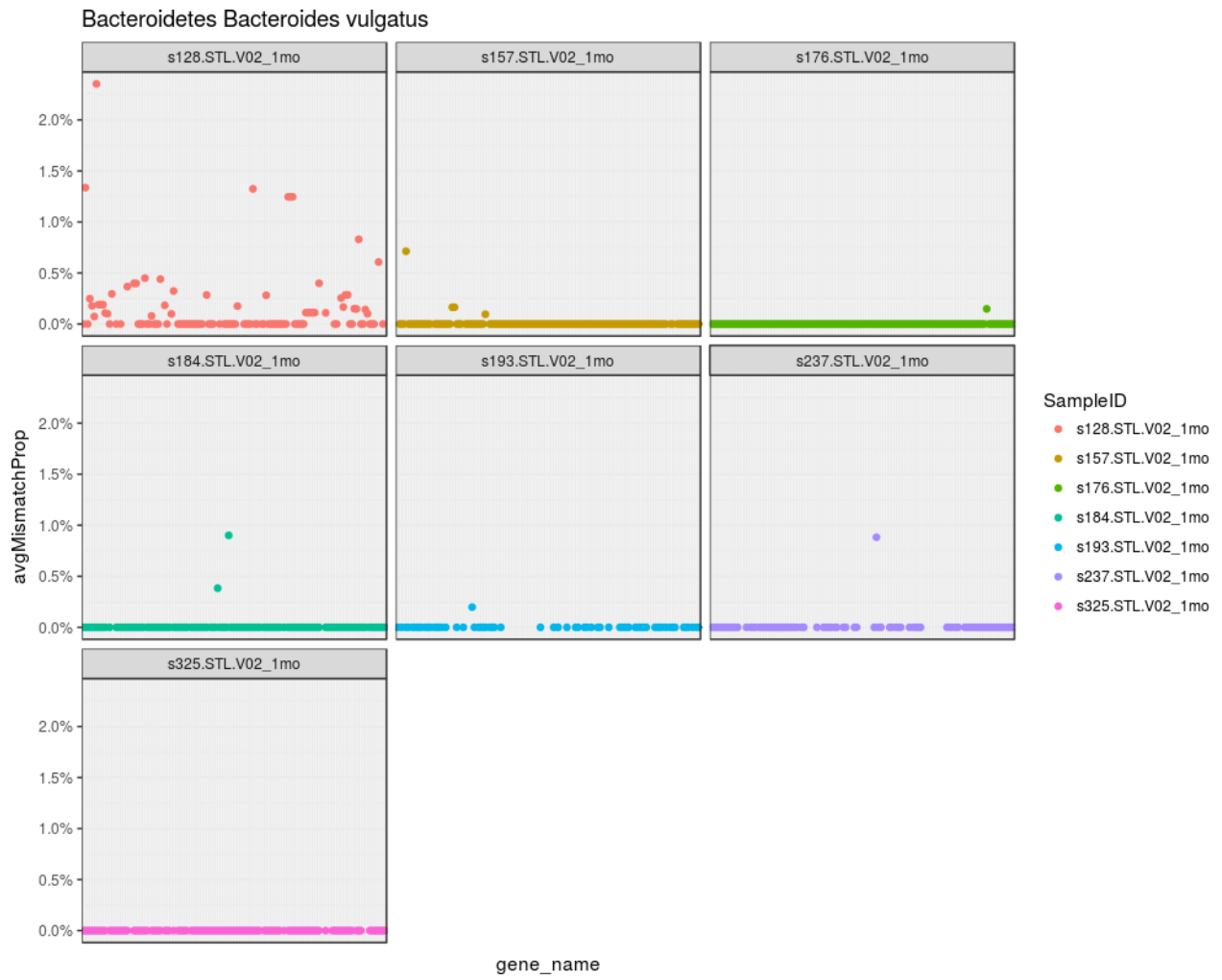


3.3 Bacteroidetes Bacteroides vulgatus

3.3.1 Birth



3.3.2 1 month



3.4 Firmicutes Enterococcus faecalis

3.4.1 Birth



3.4.2 1 month

