# Testing the Black Queen Hypothesis (BQH)

"The BQH predicts that the loss of a costly, leaky function is selectively favored at the individual level and will proceed until the production of public goods is just sufficient to support the equilibrium community"

- Cost of fitness vs. gain of fitness
  - Loss of functional gene is a potential cost in fitness
  - Fitness benefit is saving energy and resources
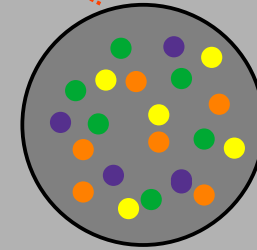
1

Experimental setup
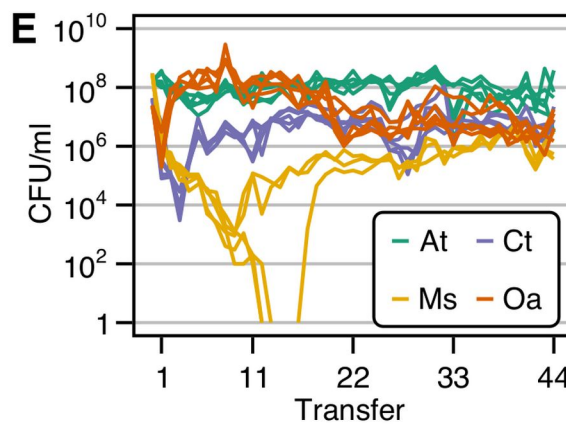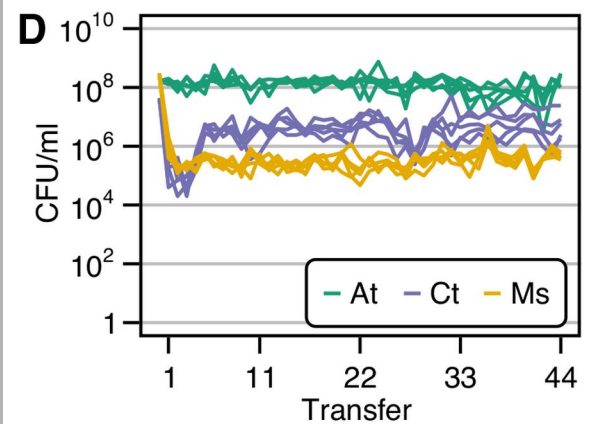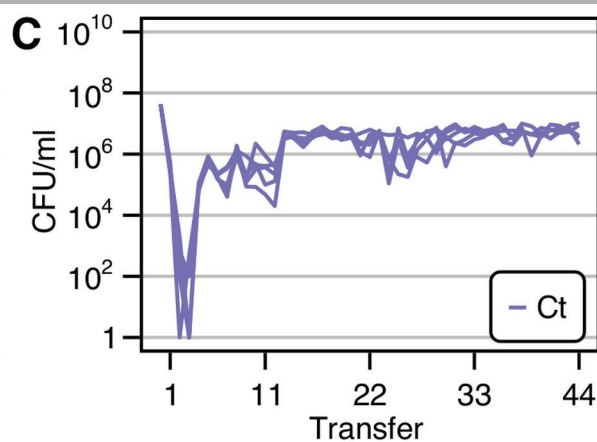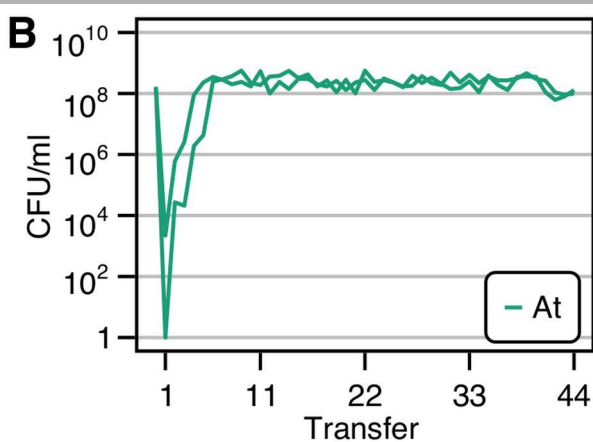
Sampling and DNA sequencing

Reference genome

E

CFU/ml

Legend: At, Ct, Ms, Oa
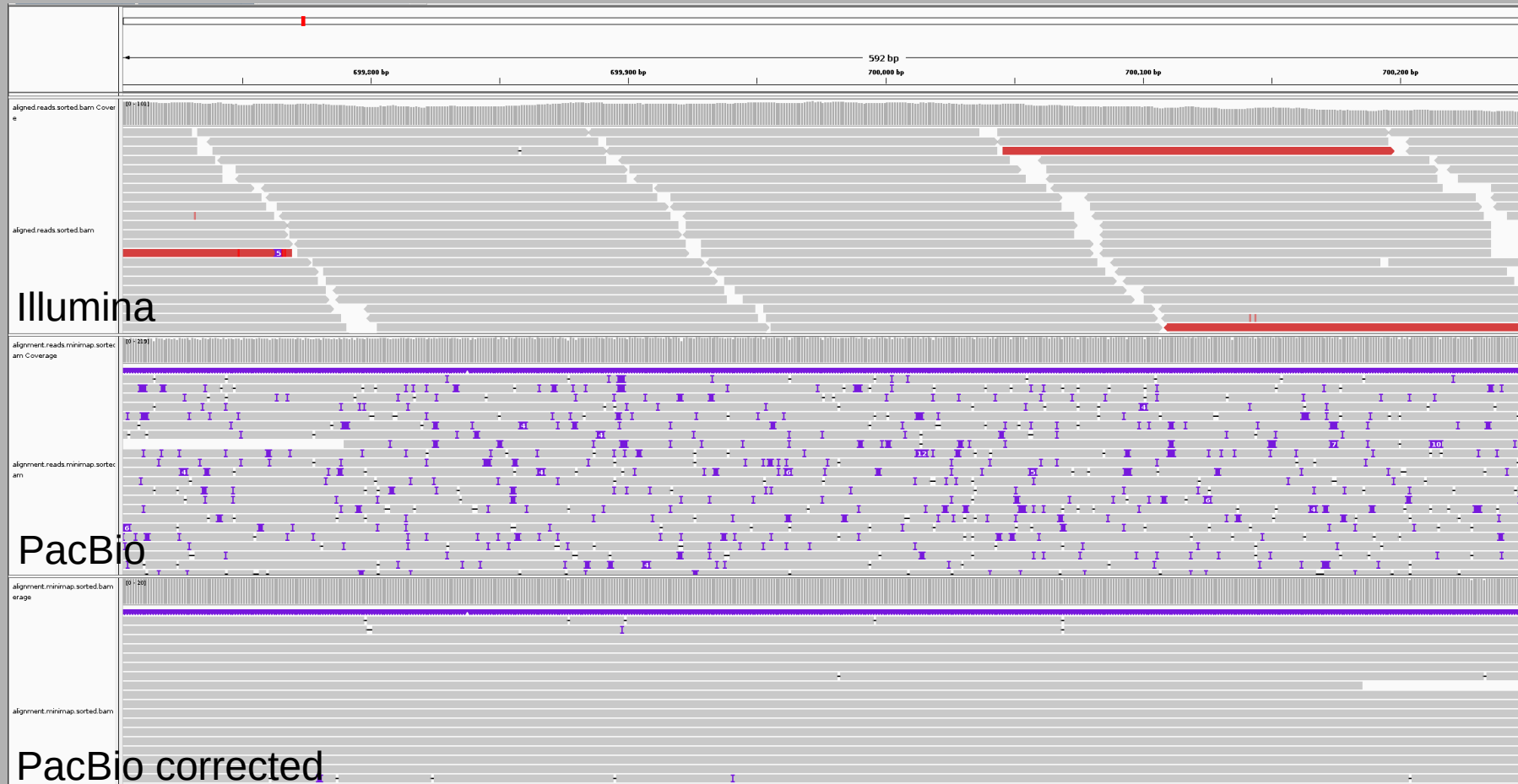
Transfer: 1, 11, 22, 33, 44
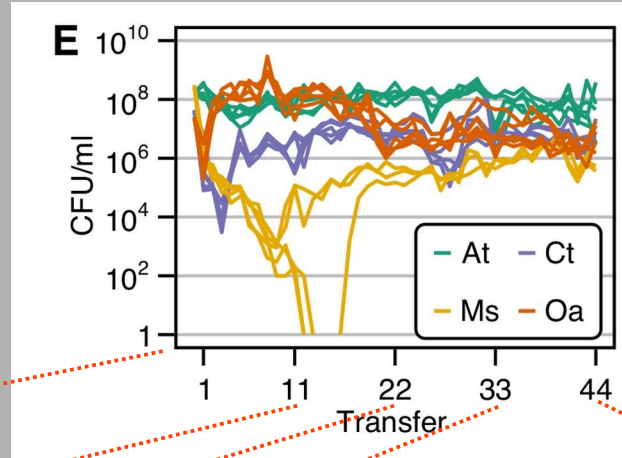
- Illumina population sequencing per sample
- High genomic diversity
- Potentially low coverage
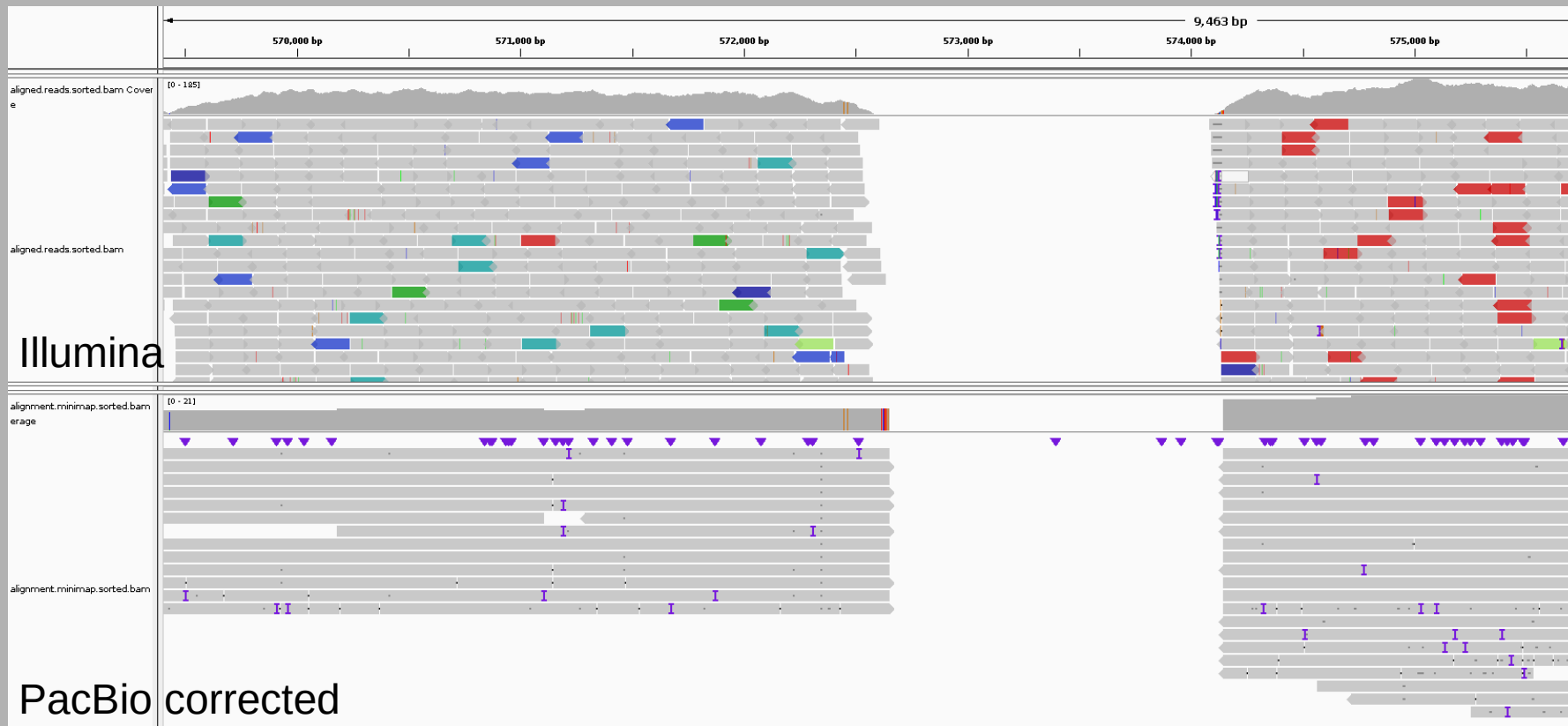
- Colony picking and amplification
- PacBio sequencing
- No genomic diversity
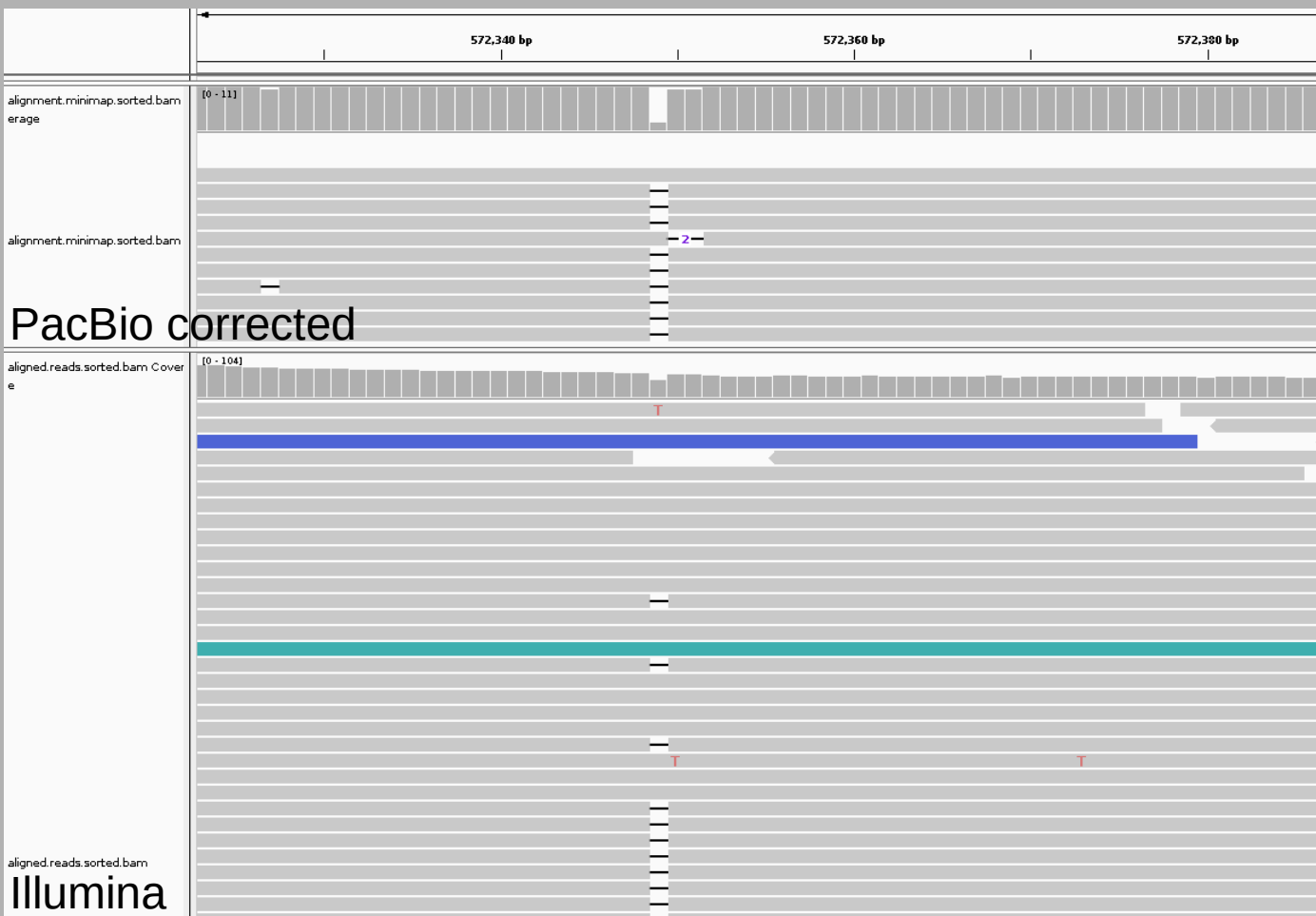- High coverage

3

Sequencing and alignment

Illumina

PacBio

PacBio corrected

4

Deletions defined by missing alignment
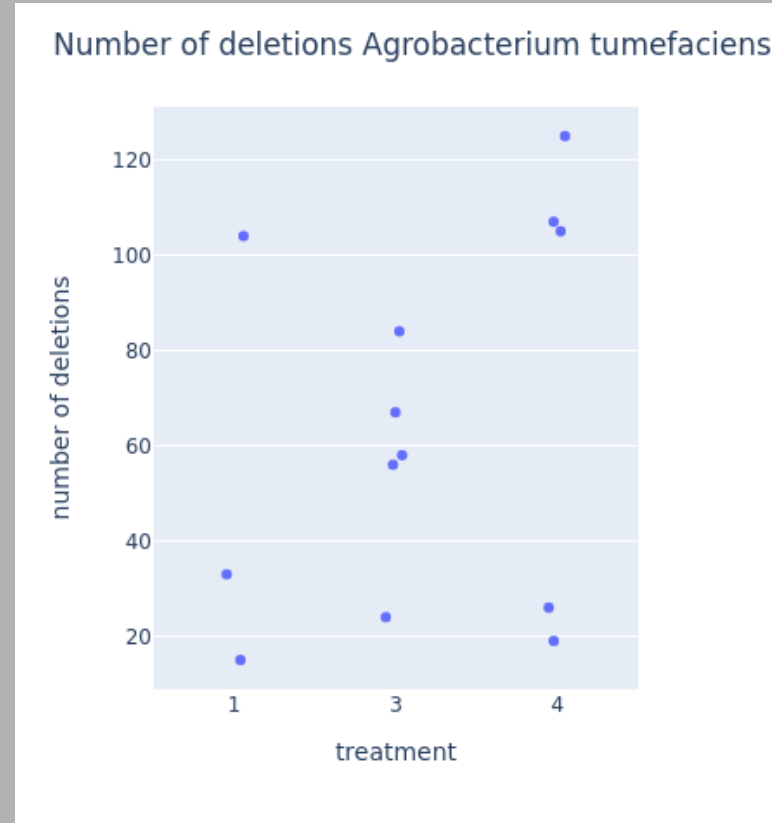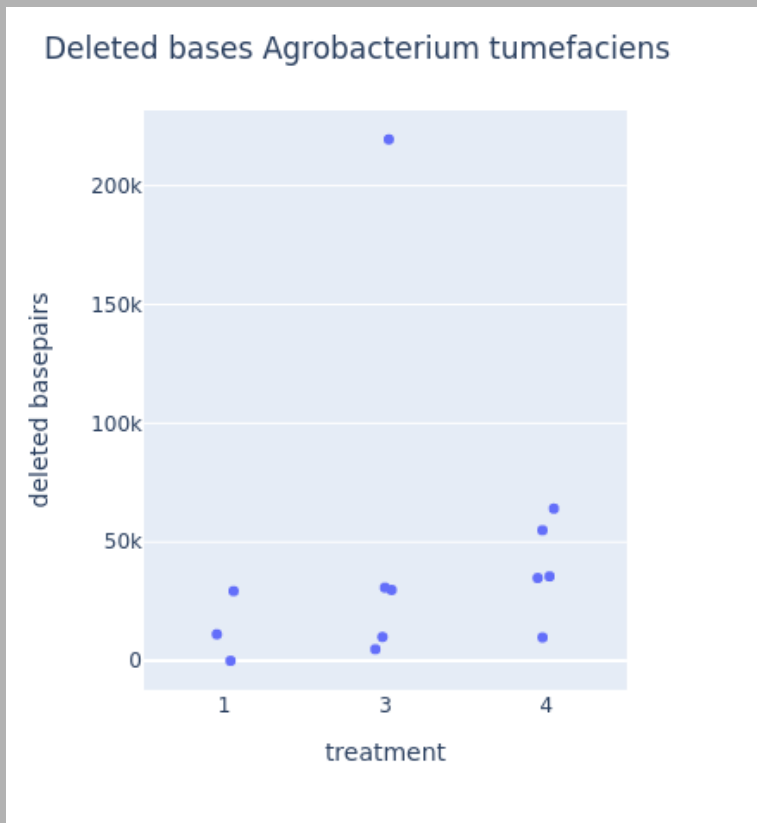

Illumina

PacBio corrected

- Large deletions are clearly visible by abrupt coverage drop

5

Deletions where read spans missing sequence

PacBio corrected

Illumina

6

Deletions in PacBio data

Left: Sum of all deleted basepairs in a sample grouped by treatment. **Right:** Number of deletion events in one sample grouped by treatment.
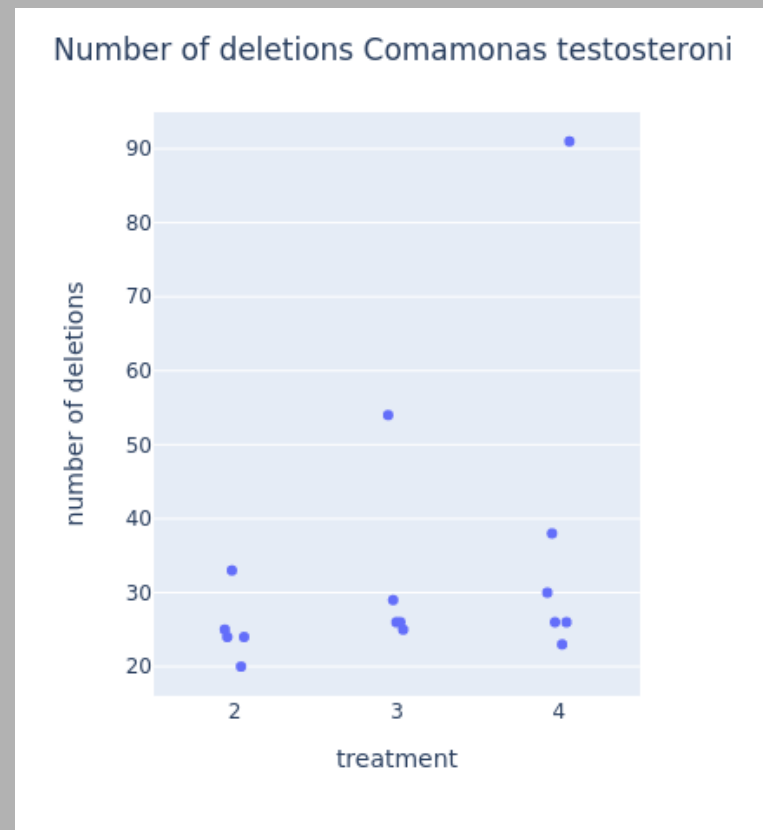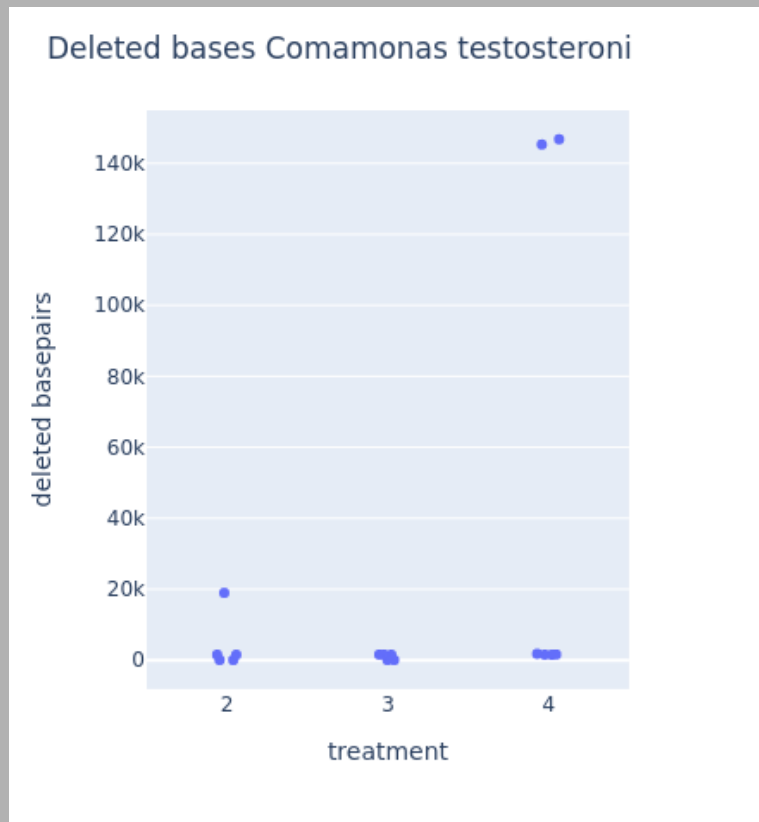
7

Deletions in PacBio data

Left: Sum of all deleted basepairs in a sample grouped by treatment. **Right:** Number of deletion events in one sample grouped by treatment.

Deletions in PacBio data
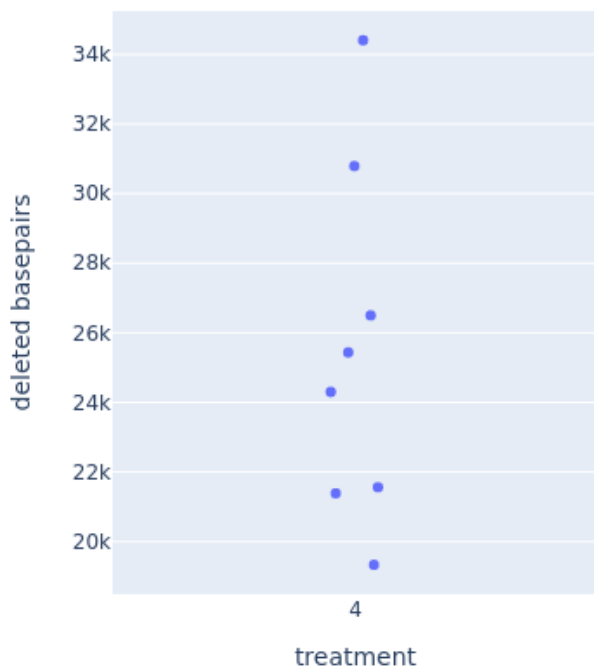

Left: Sum of all deleted basepairs in a sample grouped by treatment. **Right:** Number of deletion events in one sample grouped by treatment.
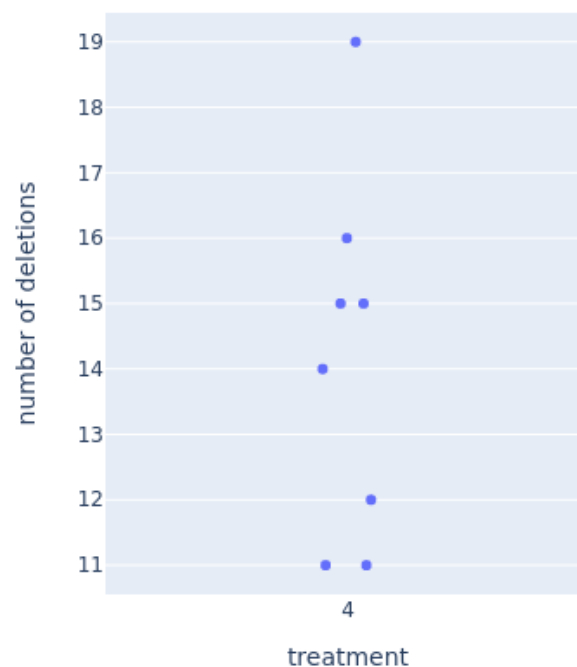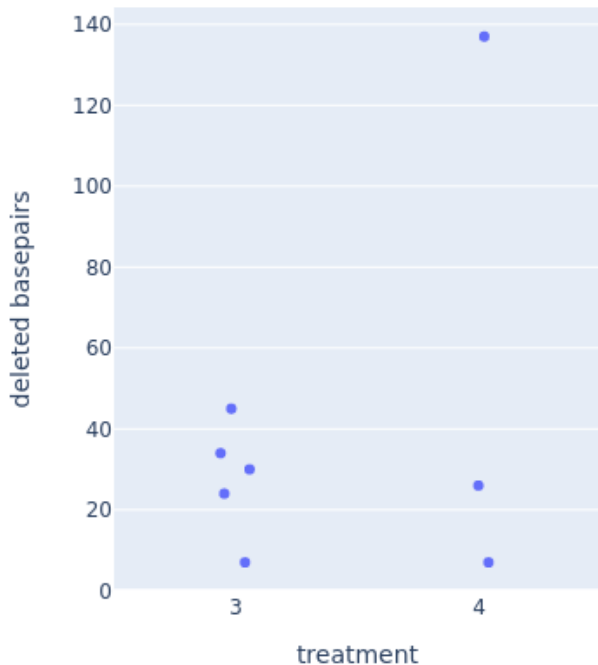
Deletions in PacBio data
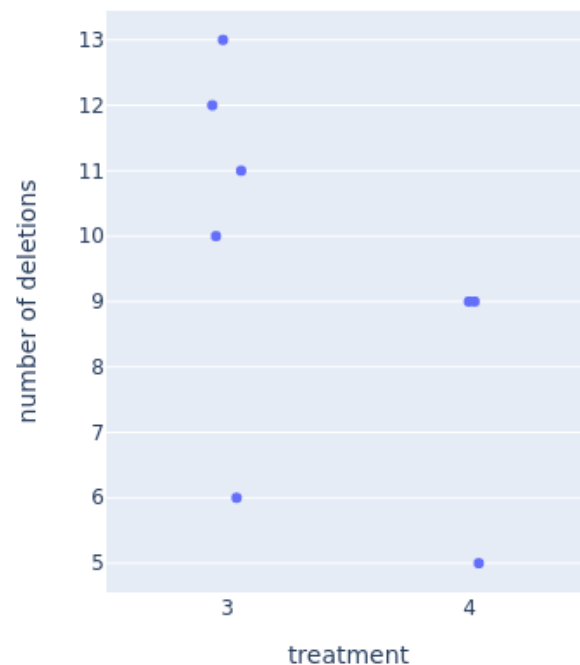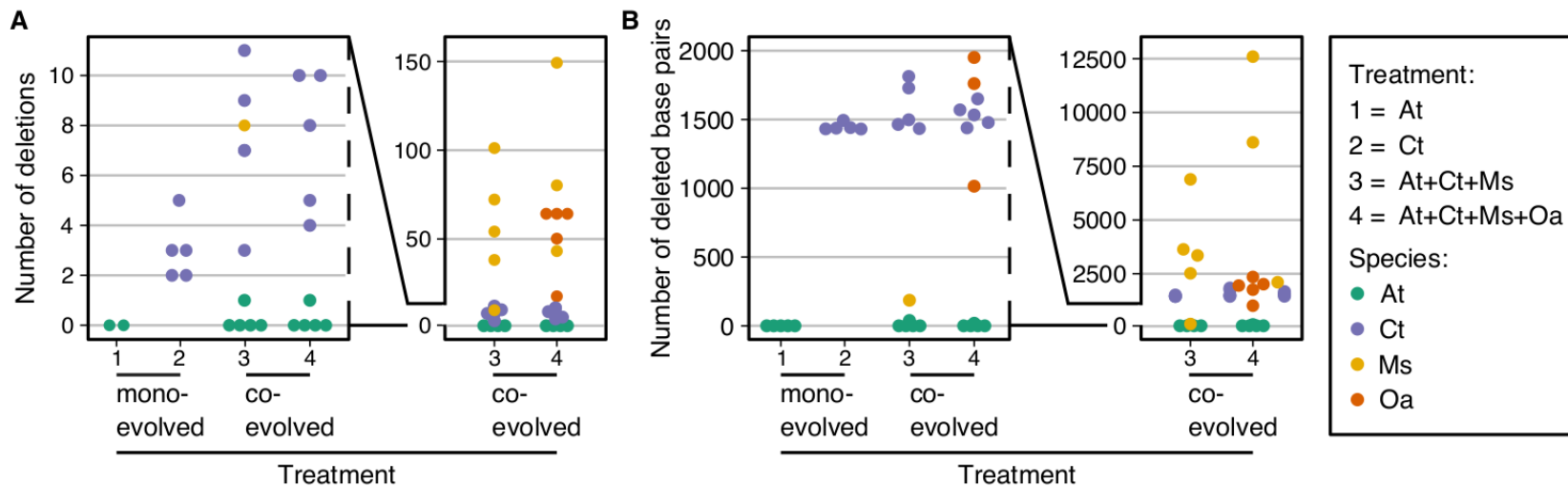

Left: Sum of all deleted basepairs in a sample grouped by treatment. **Right:** Number of deletion events in one sample grouped by treatment.

Illumina analysis

Low Illumina coverage areas

- Many deletions in *MS* were not present in PacBio data
- High coverage fluctuations
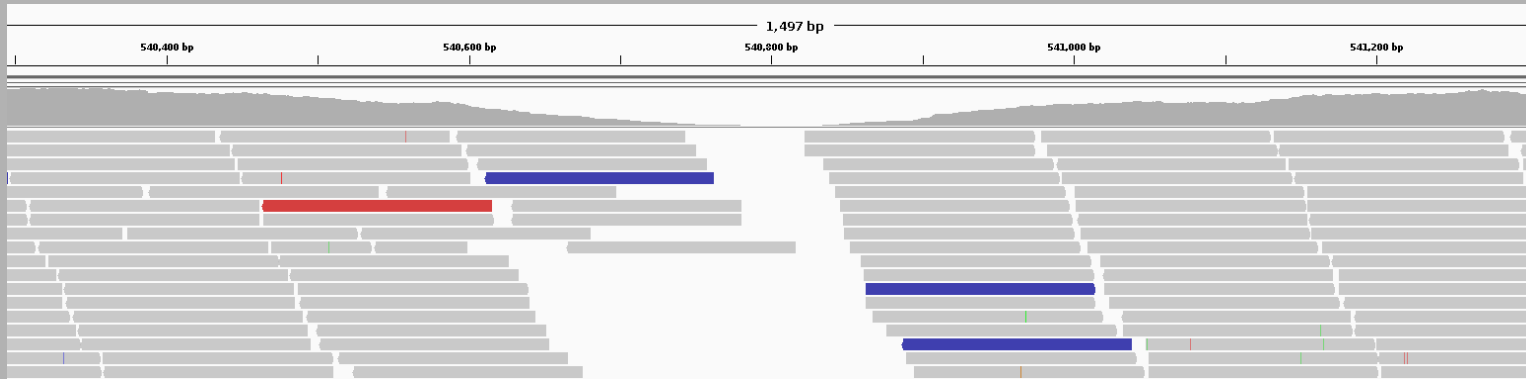
- Zero-coverage areas are often dominated by GC

GCAGCGGGGGCCGCGGCAGCGAGGGCGGCGGGGGTCTGCGACGGCG

CCGGCGGCGCCGCGGGAGGAGCCGGCGGCGTGGTCGGCGGAACCGG

CTCGTACGCGCTCGCCCACCCCGAGGCGCGCGTGACGGCGACCGACC

12

GC bias in Illumina sequencing

- GC bias introduced in PCR step
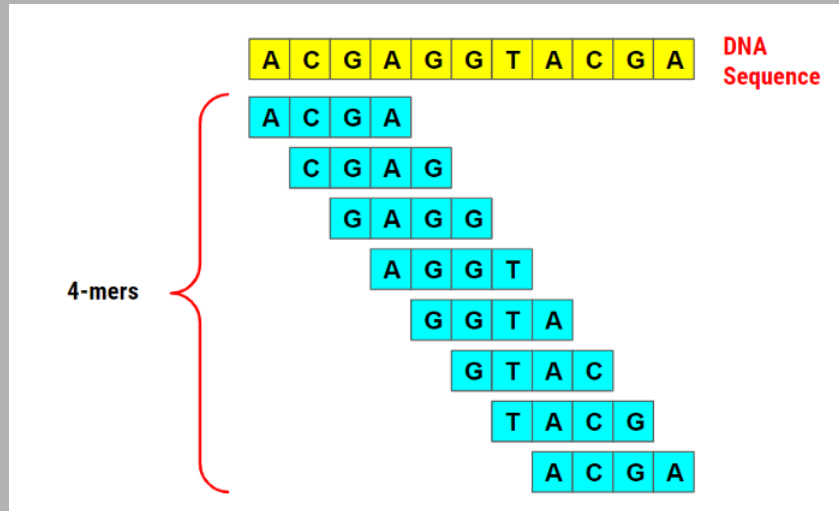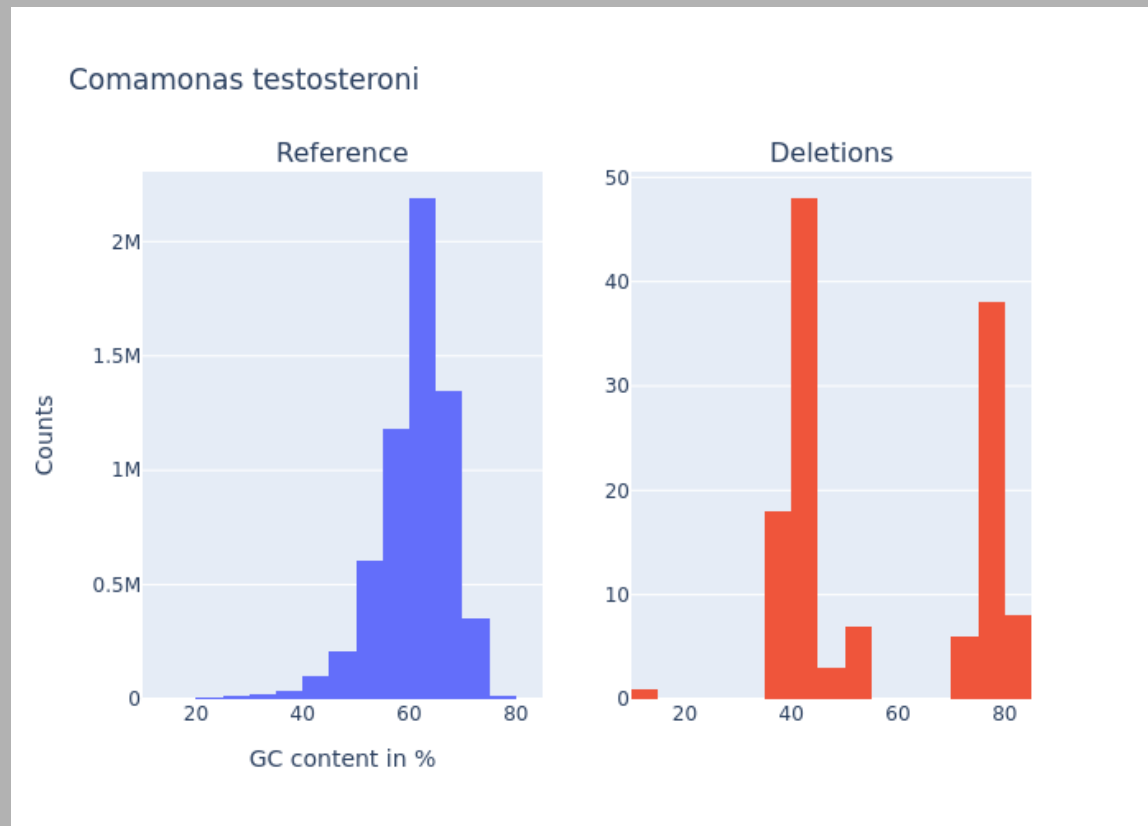- dsDNA with high GC content needs more energy to break up
- Calculation of GC content:

13

GC content of Illumina deletions

Left: GC content distribution of reference genome, window size 150. **Right:** GC content of all deletions of one strain.
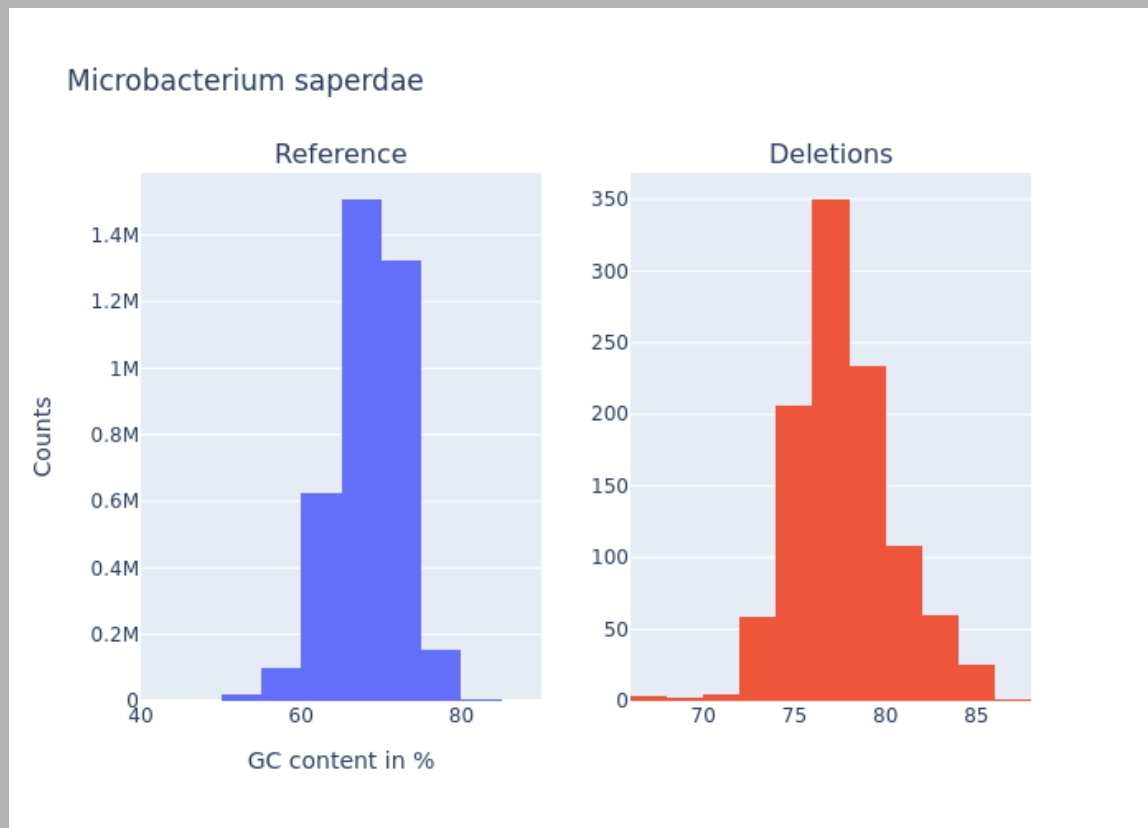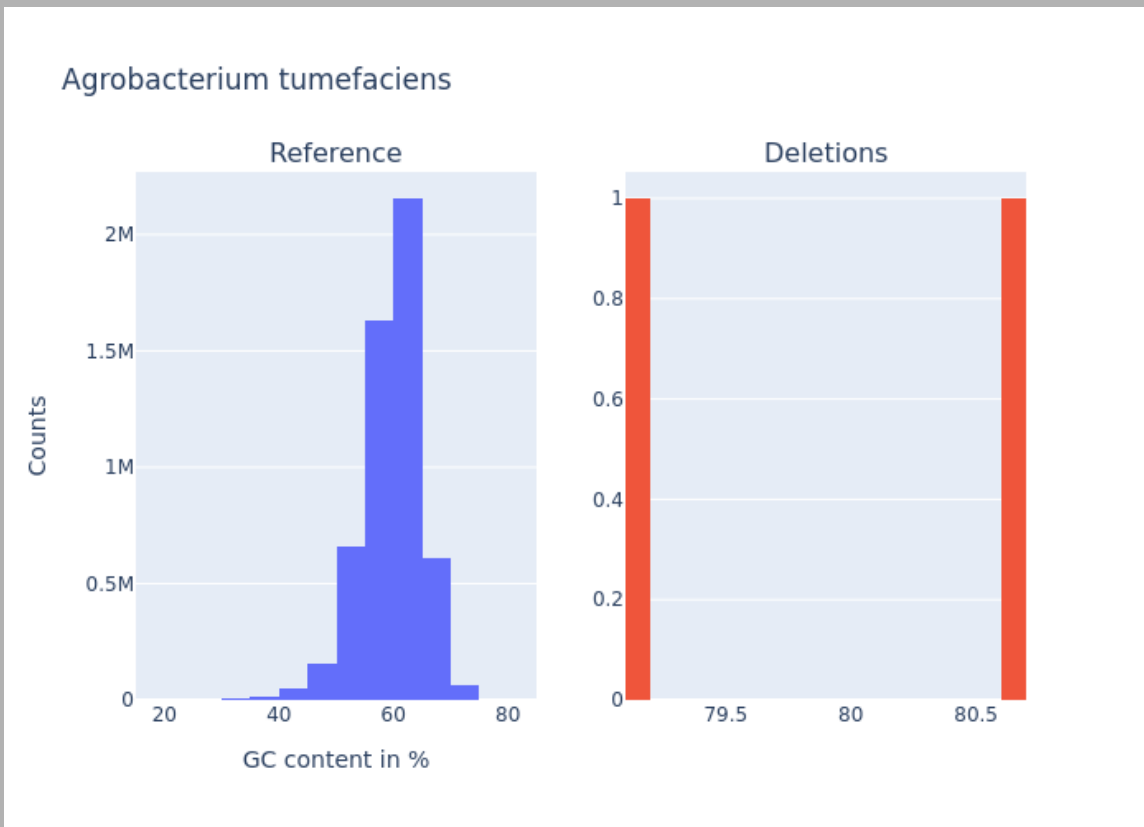Window size position-50:position+length+50

14

GC content of Illumina deletions

Left: GC content distribution of reference genome, window size 150. **Right:** GC content of all deletions of one strain.
Window size position-50:position+length+50

GC content of Illumina deletions

Agrobacterium tumefaciens

Left: GC content distribution of reference genome, window size 150. **Right:** GC content of all deletions of one strain.
Window size position-50:position+length+50

16

GC content of Illumina deletions
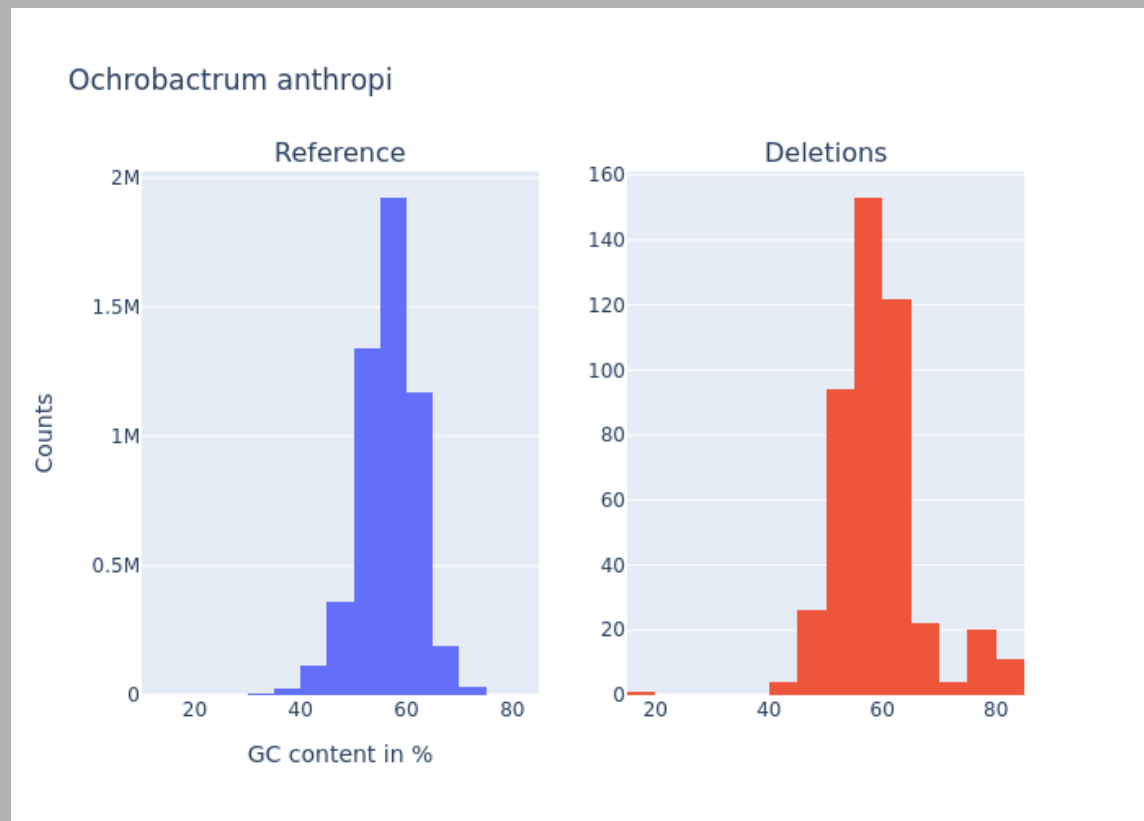

Ochrobactrum anthropi

Left: GC content distribution of reference genome, window size 150. **Right:** GC content of all deletions of one strain.
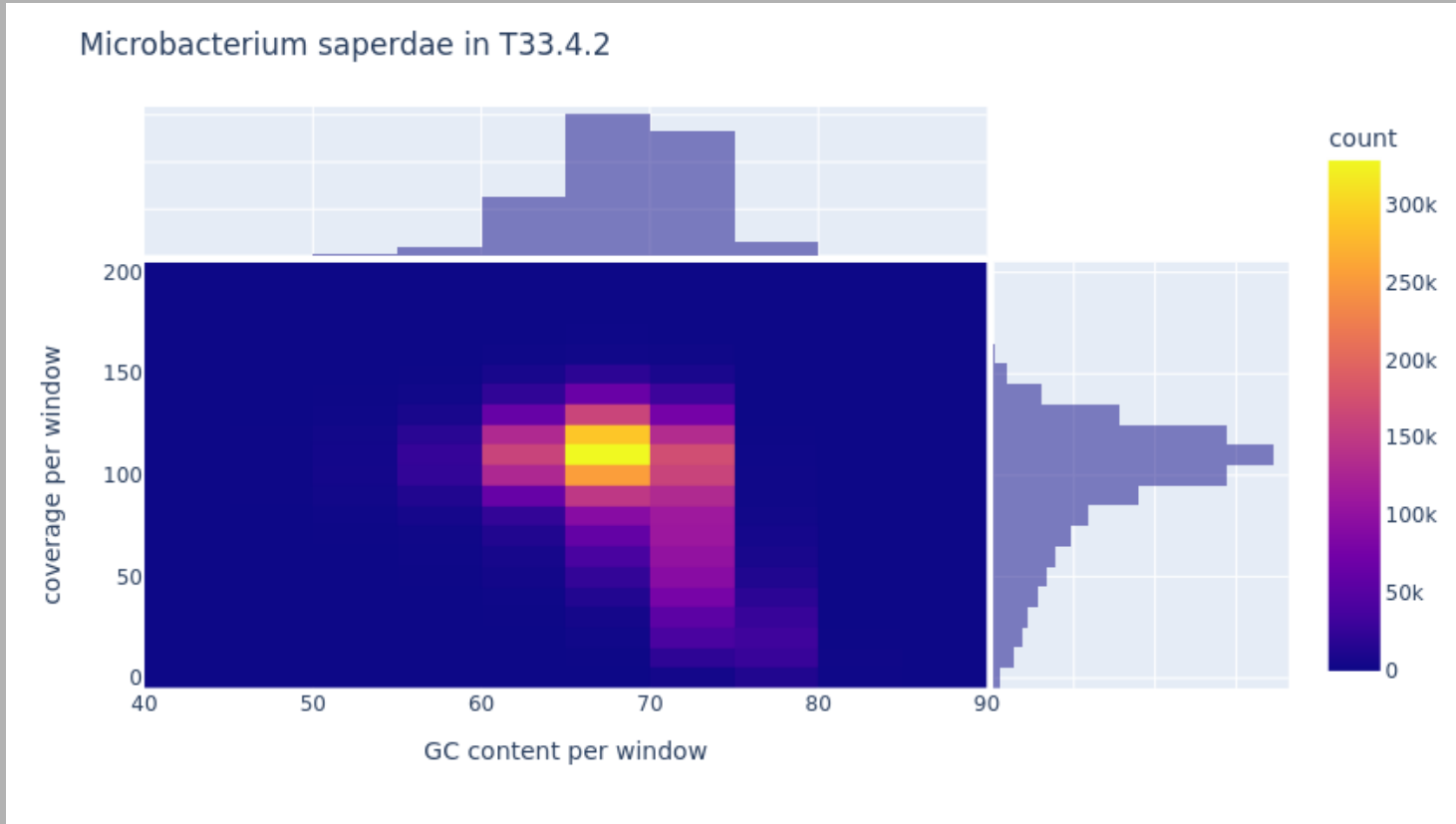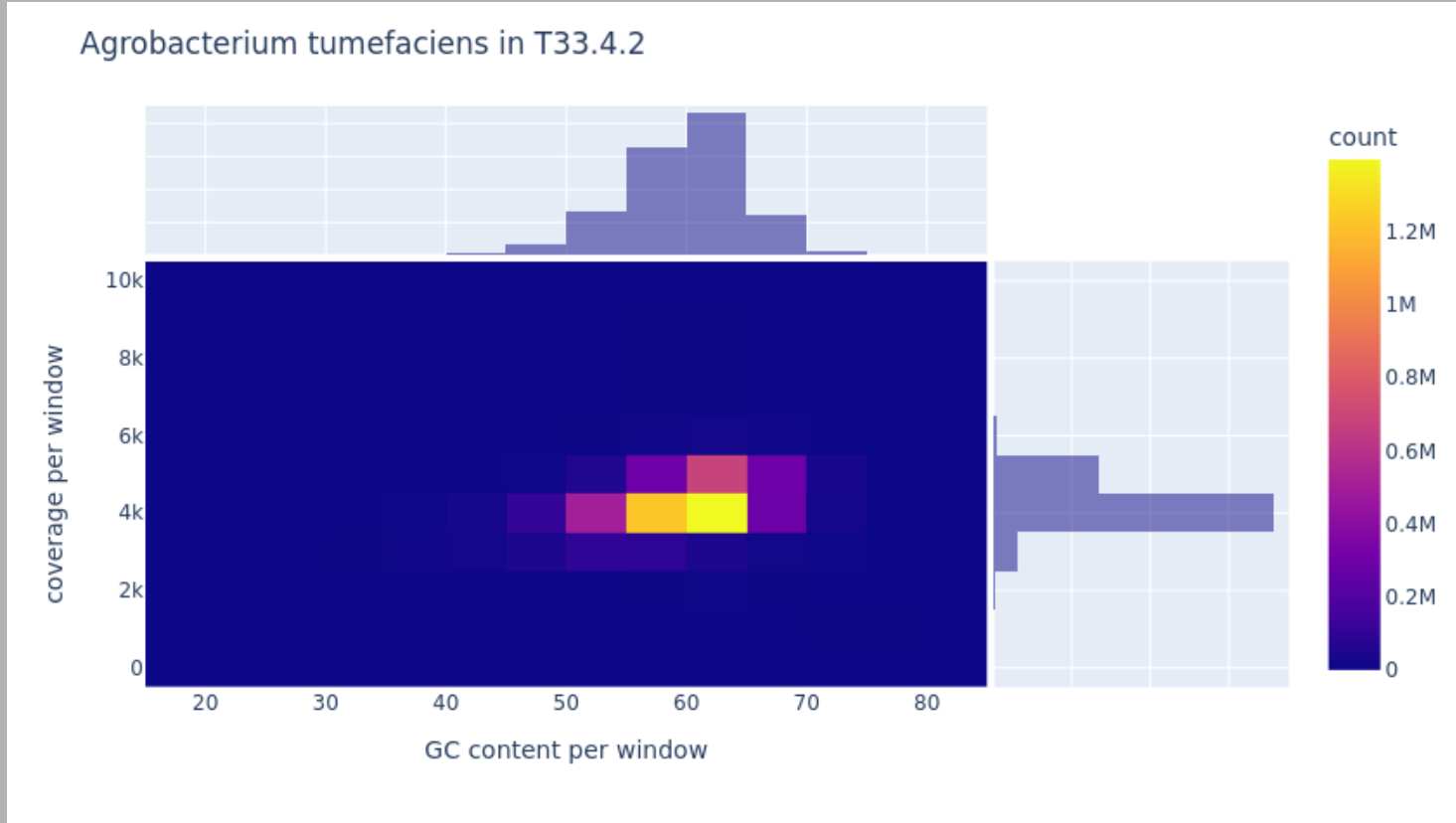Window size position-50:position+length+50

17

GC content vs. Coverage

Microbacterium saperdae in T33.4.2

Reference sequence split to windows. Calculated GC content and coverage
Of every window and displayed as 2d-histogram.

GC content vs. Coverage

Agrobacterium tumefaciens in T33.4.2
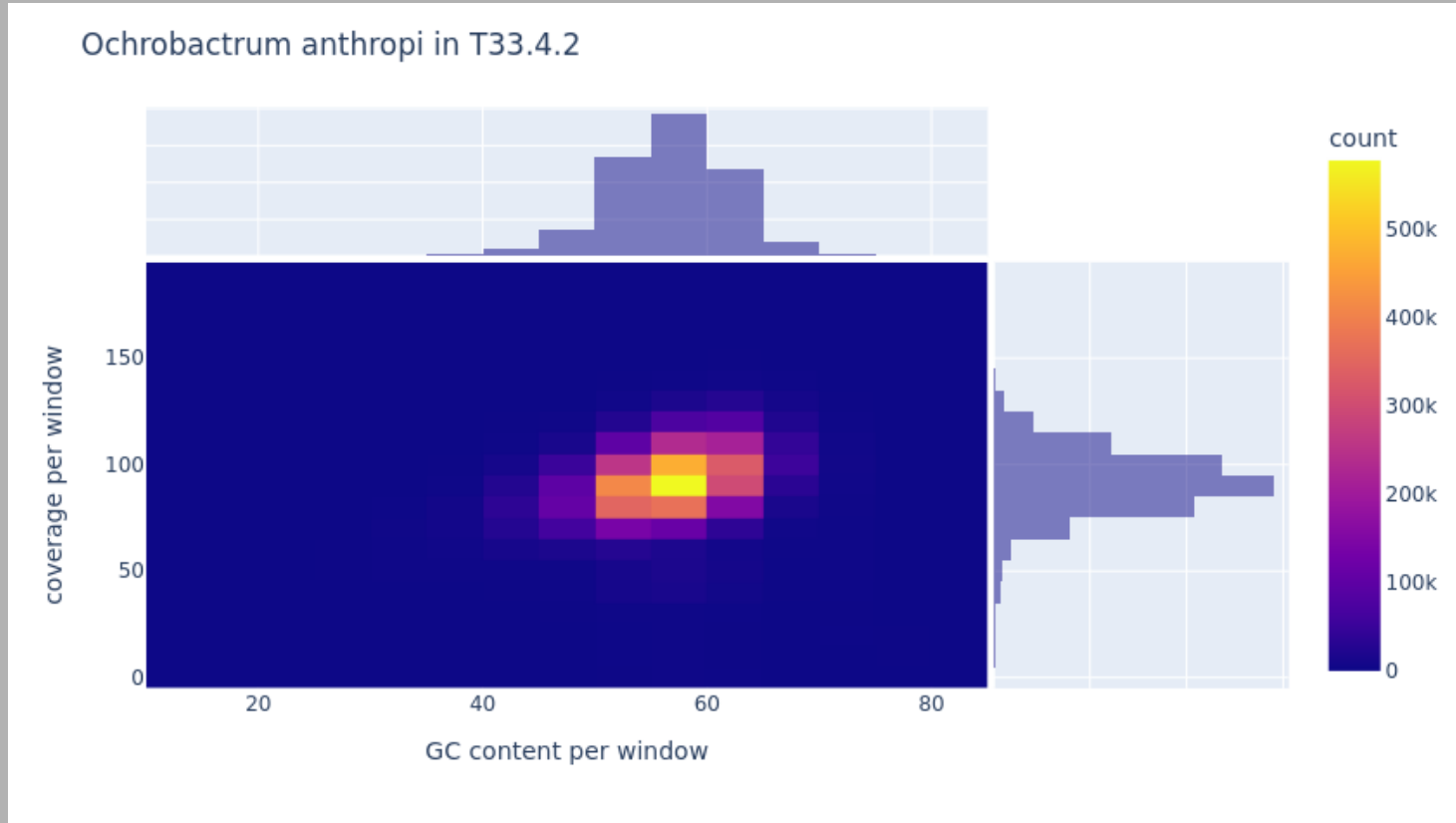
Reference sequence split to windows. Calculated GC content and coverage
Of every window and displayed as 2d-histogram.

GC content vs. Coverage

Ochrobactrum anthropi in T33.4.2
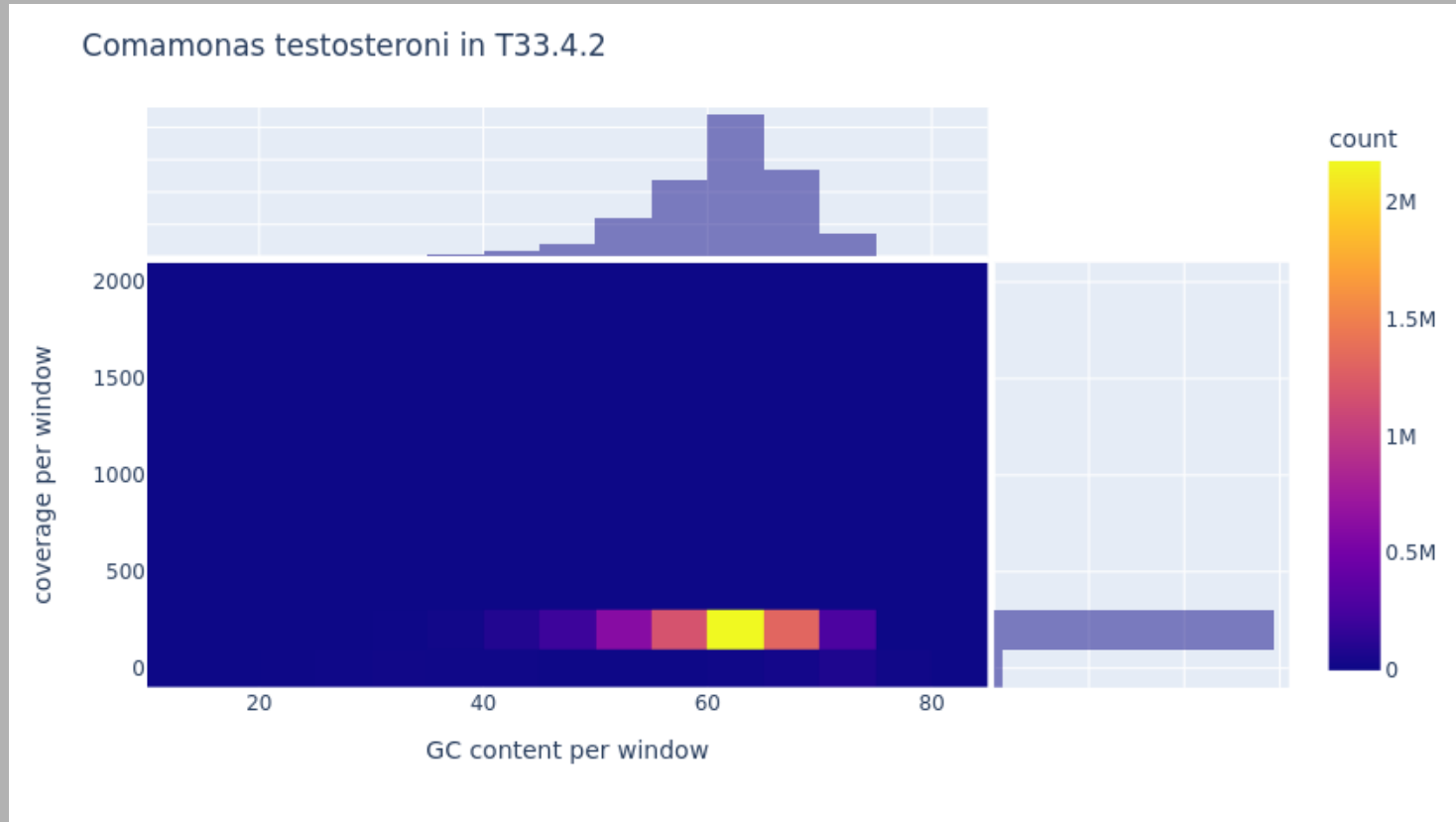
Reference sequence split to windows. Calculated GC content and coverage
Of every window and displayed as 2d-histogram.

GC content vs. Coverage

Comamonas testosteroni in T33.4.2

Reference sequence split to windows. Calculated GC content and coverage
Of every window and displayed as 2d-histogram.