Uncovering the Dynamics of CRISPR Array Evolution with a Spacer Order Respecting Maximum Likelihood Approach

Axel Fehrenbach*, Alexander Mitrofanov[1], Omer Alkhnbashi[1], Rolf Backofen[1], Franz Baumdicker*

*University of Tübingen, Cluster of Excellence Controlling Microbes to Fight Infections & Cluster of Excellence Machine Learning for Science
[1]University of Freiburg, Department of Computer Science

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Overview

- CRISPR arrays provide a chronological history of acquired spacers.
- We introduce **SpacerPlacer**, a ancestral reconstruction tool that employs evolutionary models and enforces spacer insertion order to reconstruct ancestral arrays.
- With **SpacerPlacer** we investigate, if spacers are deleted in blocks, and compare estimated model parameters between CRISPR-Cas types, genus and more.

Placing Spacers with SpacerPlacer: Ancestral Reconstruction Algorithm
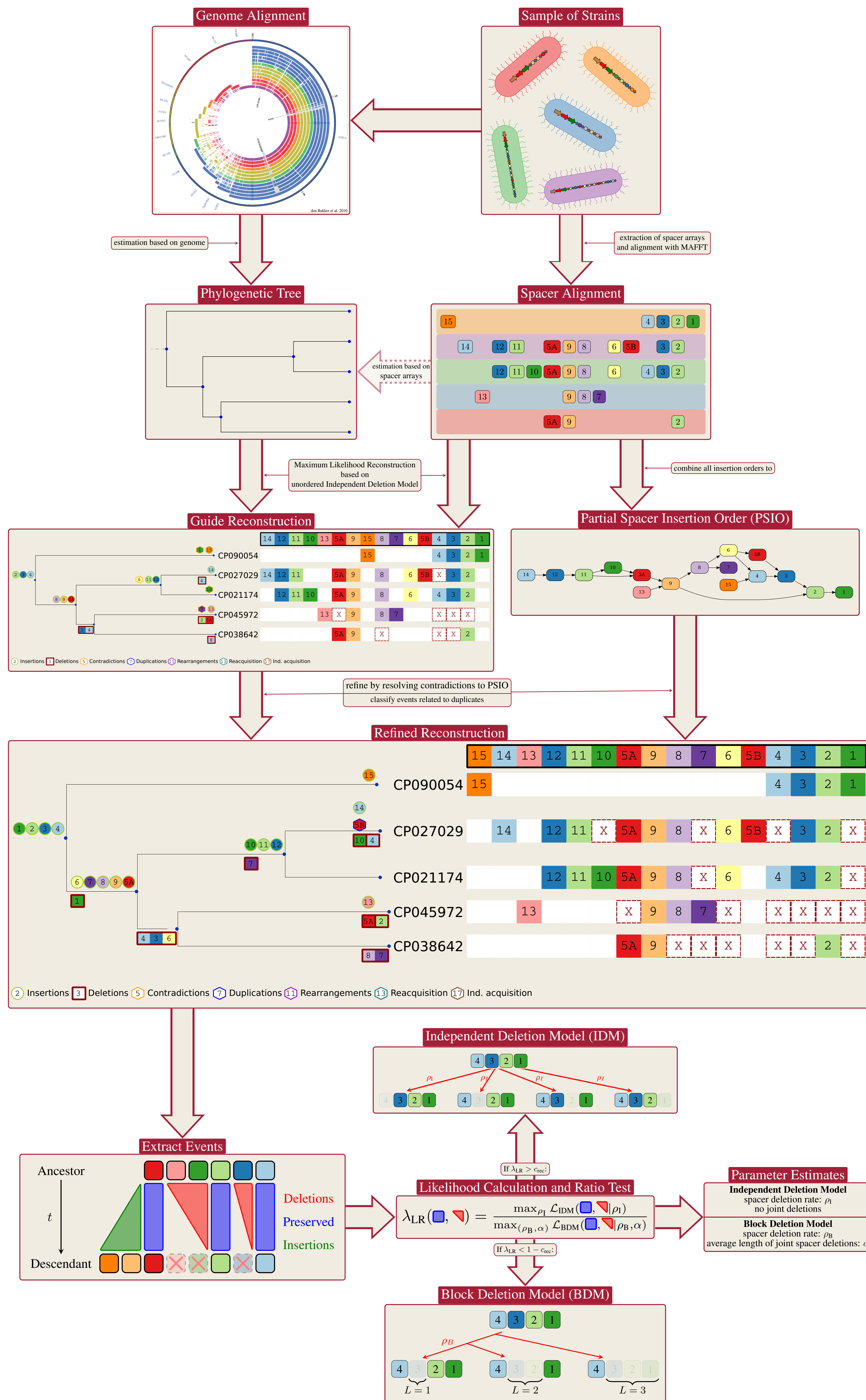
Data

- CRISPR array data extracted from *CRISPRdb* [1] (7352 repeats with 32624 arrays of highest evidence).
- Arrays were clustered by species, Cas type, overlap of spacers and spacer diversity into 337 groups with 5192 arrays (2327 are unique).
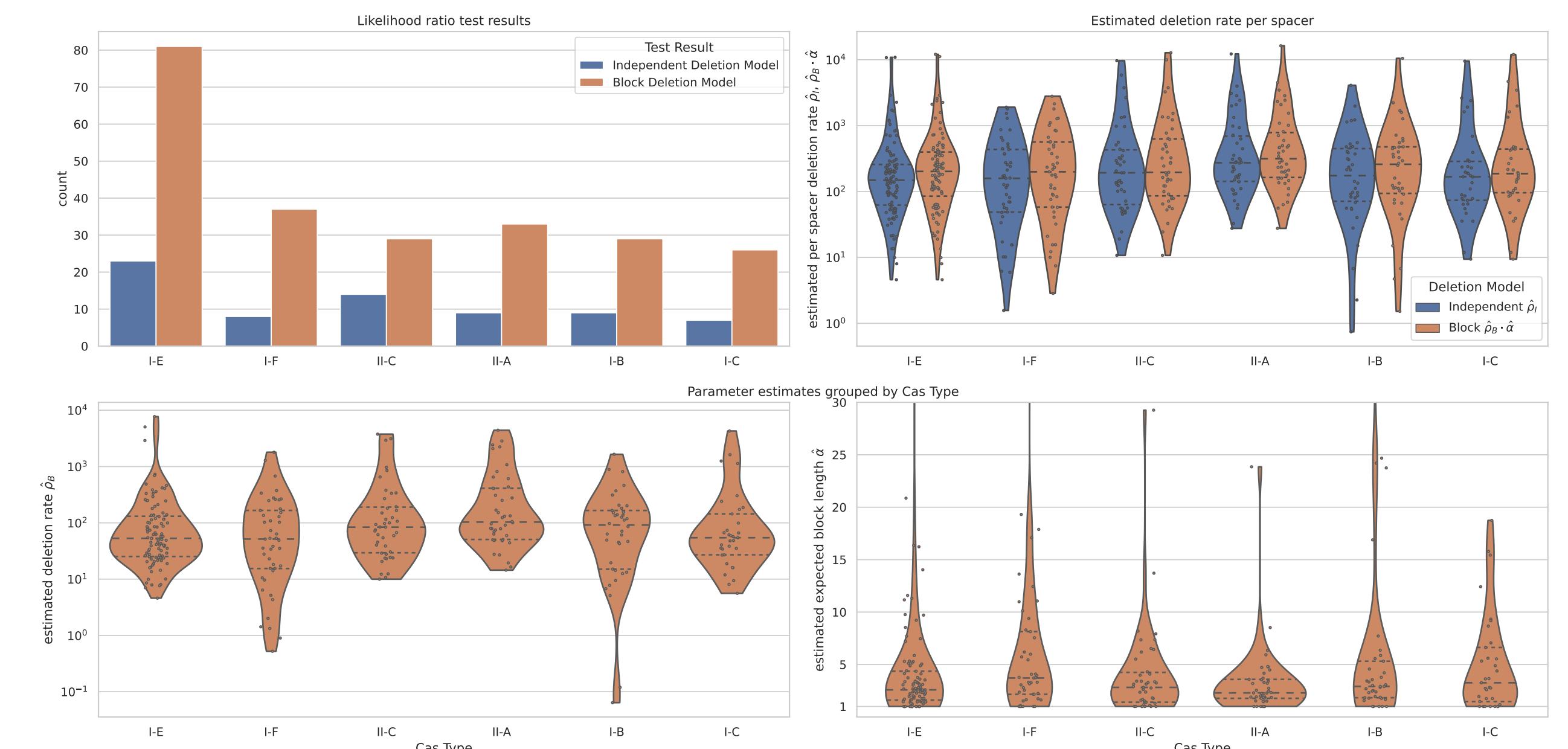
Contact

axel.fehrenbach@uni-tuebingen.de
franz.baumdicker@uni-tuebingen.de
www.baumdickergroup.de
🐦 fbaumdicker
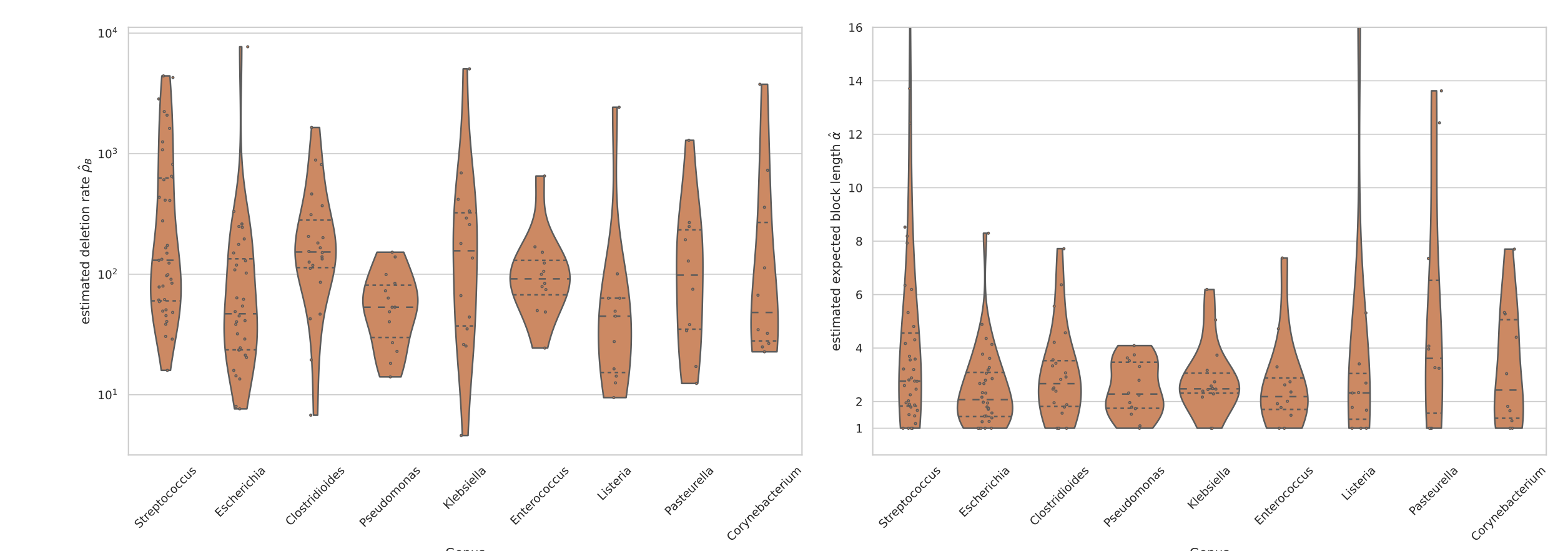https://github.com/fbaumdicker/SpacerPlacer

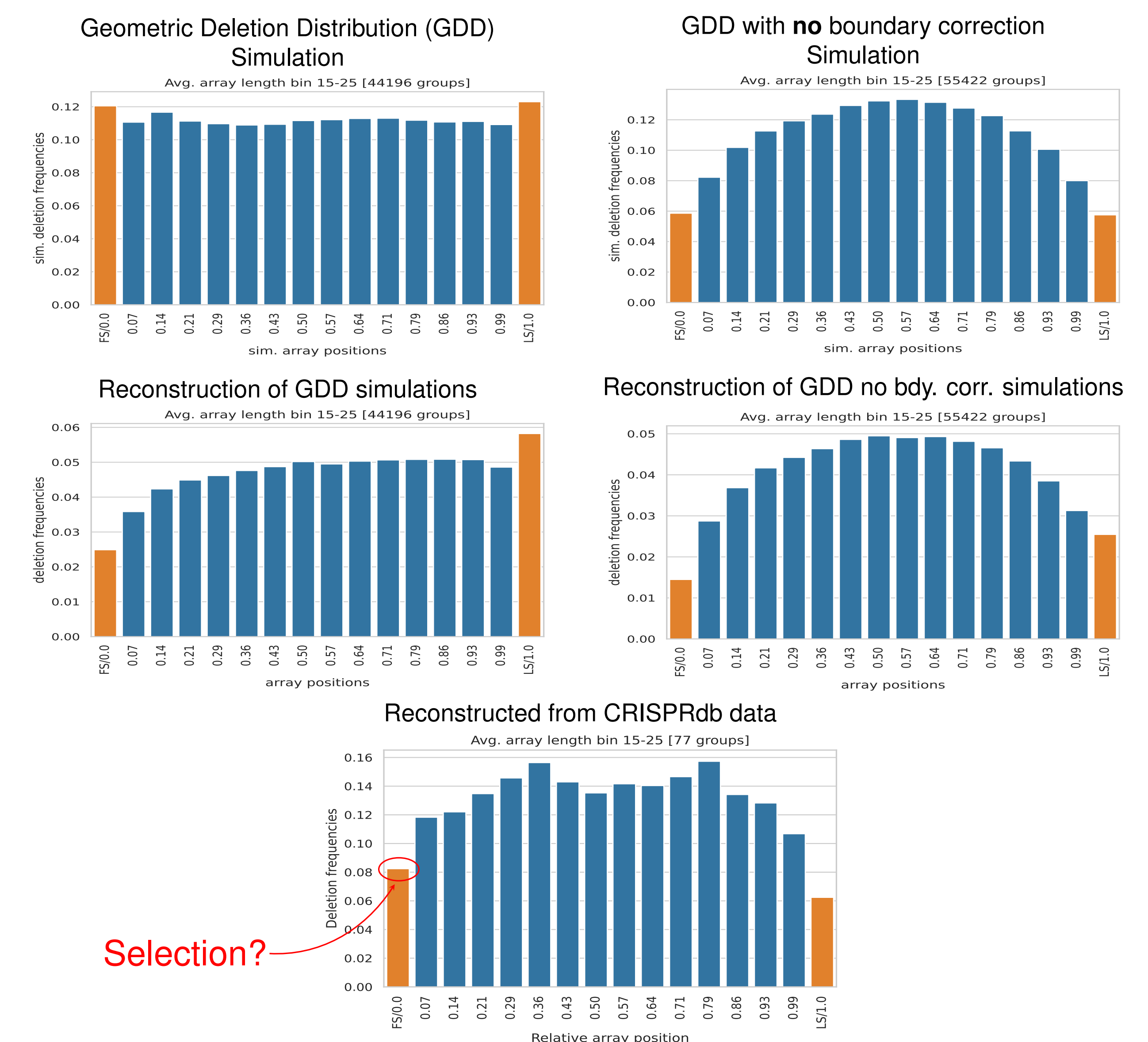## Test Results, Parameter Estimates (Cas Type)

Mean of spacer deletion initiation rate $\hat{\rho}_B \sim 300$ times higher than per site mutation rate.
Mean of average length of block deletions $\hat{\alpha} \approx 4.7$.

Parameter Estimates (Genus)

Deletion Boundary Behavior → Selection?

Selection?

Conclusions and Applications

- Test results strongly suggest that spacers are lost in blocks.
- Marginal differences of deletion rates between Cas types. Larger differences between genera.
- First and last spacer are less likely to be deleted. Indication of selection at array insertion end.
- **SpacerPlacer** serves as stepping stone for further research, e.g.:

CRISPR - evo-inator

CRISPR evolution based inference of array transcription orientation

visit Poster 108 to learn more

Relevant Publications

[1] Christine Pourcel et al. "CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers". In: *Nucleic acids research* 48.D1 (2020), pp. D535–D544.
[2] John Rozewicki et al. "MAFFT-DASH: integrated protein sequence and structural alignment". In: *Nucleic acids research* 47.W1 (2019), W5–W10.