

PiMaker: Vectorized calculation of diversity statistics to measure evolutionary pressure in a variety of populations at scale

Joseph Lalli¹, Kelsey Florek², Shari Barlow³, Amra Uzicainin⁴, Donna Werling¹, Jonathen Temte³, Thomas Friedrich⁵



¹Department of Genetics, University of Wisconsin-Madison, Madison, WI, United States; ²Wisconsin State Lab of Hygiene, Madison, WI, United States; ³Department of Family Medicine and Community Health, University of Wisconsin-Madison, Madison, WI, United States; ⁴Centers for Disease Control and Prevention, Atlanta, GA, United States; ⁵Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI, United States;

Measuring adaptation with PiMaker

Nucleotide Diversity (Najima's π): Average likelihood of two reads being identical at a given site

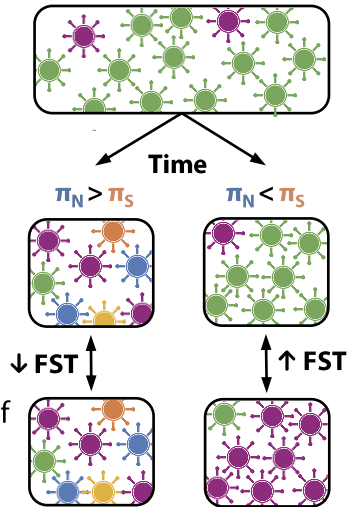
π_N : π at AA changing sites

π_S : π at silent mutations

If $\pi_N > \pi_S$: AA changes are under positive selection

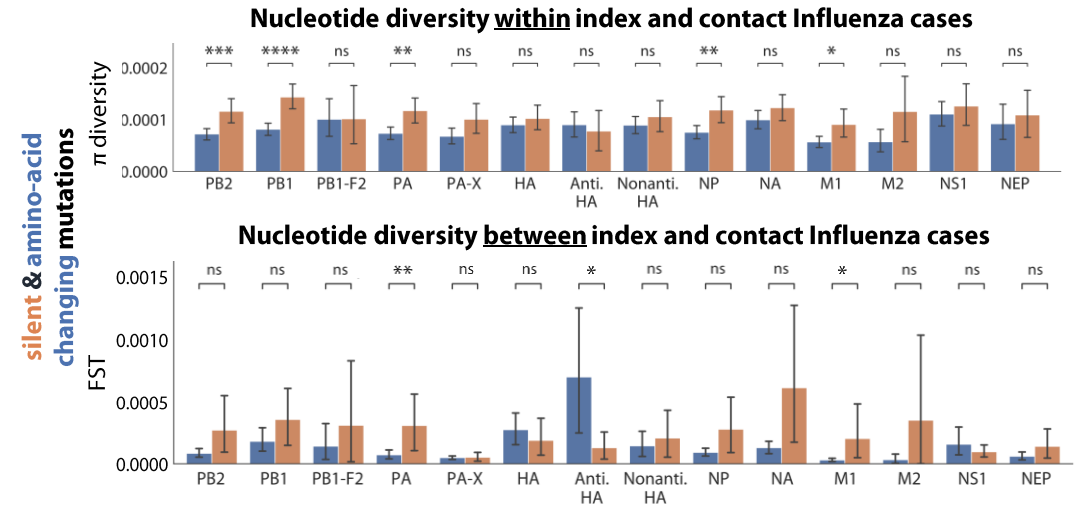
If $\pi_N < \pi_S$: AA changes are under negative selection

FST_N & FST_S : Degree to which π of two populations is segregated vs shared; measures diversification



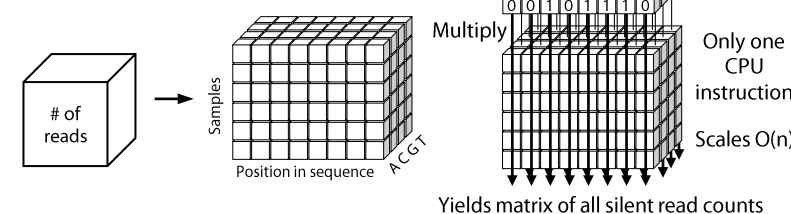
PiMaker reveals natural selection is strongest during transmission of influenza

- Nasal swabs from 72 paired index-contact cases sequenced to a depth of 10,000X
- PiMaker used to identify within and between host diversity
- π_N/π_S analysis shows that influenza mutations are under **weak purifying selection within hosts**
- Novel **antigenic sites** of influenza are under **strong positive selection upon infection of a new host**

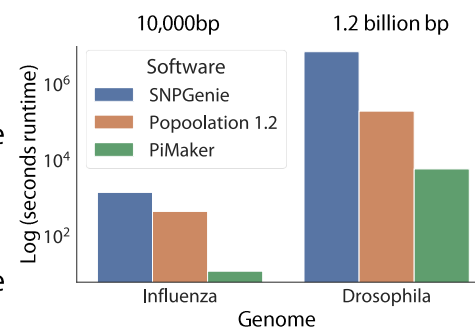


Leveraging vectorization to rapidly analyze large datasets

Organizing raw read counts E.g., to obtain silent mutation data: into matrices allows for **vectorized** computation



- Trades increased memory usage for rapid calculation
- Able to process multiple contigs and multiple samples at once
- Works with VCF files, allowing use of separate basecaller



Further Applications

- **Tumor evolution** hotspot identification
- **Antibiotic resistance** monitoring
- **SARS-CoV-2** evolution

Funding graciously provided by University of Wisconsin MSTP T32 GM140935, NIH grant R01AI125392-02, and CDC grant 5U01CK000542-02.

QR code to PiMaker Github repository:

