Pyroprinting Sensitivity Analysis on the GPU

Aldrin Montana, Doug Brandt, Bob Somers, Alex Dekhtyar, Chris Lupo
Chris Kitts, Anya Goodman, Michael Black

Computer Science ~
Software Engineering

Overview

The biology department at Cal Poly, San Luis Obispo developed a fingerprinting method, pyroprinting, for differentiating between strains of bacteria. To investigate the sensitivity of pyroprinting, our research group has conducted an in-silico simulation on *E. coli* DNA sequences from the 16s-23s intergenic transcribed spacer (ITS) region. The simulation parameters consisted of the dispensation sequence (CCTCTACTAGAGCG20(TCGA)TT) and primer (TTGGATCAC). From these parameters, 24 unique alleles were determined from a mix of DNA sequences from NCBI reference *E. coli* genomes, manual sequencing (from Cal Poly), and from plasmids (from Cal Poly).

Goal

We want to study and characterize pyroprinting sensitivity for the purpose of strain differentiation from a theoretical perspective. In particular, we investigate the quantity of simulated pyroprints that are deemed *similar*.

Implementation

Isolate Generation. Generate every combination of seven alleles (unordered, with replacement). Each combination represents a unique theoretical isolate in our study.

Pyroprint Generation. For each isolate, construct its pyroprint:

* Pyrosequence each allele of the isolate
* Combine pyrosequenced alleles to create a single pyroprint

Compute Correlations. Compute the Pearson correlation coefficient for each pair of pyroprints.

Histogram Generation. Separate all computed correlation coefficients into a number of "buckets" based on their values.

Simulation

Figure 1.
A subset of 12 alleles from the 16s-23s ITS region of the reference E.coli DNA cassettes. These alleles are used for the sample pyroprints in Figure 2.

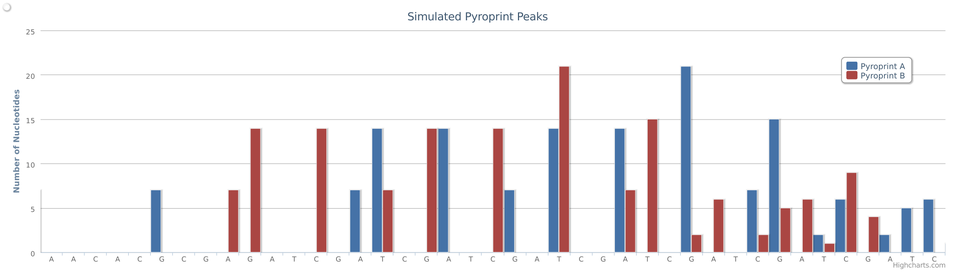
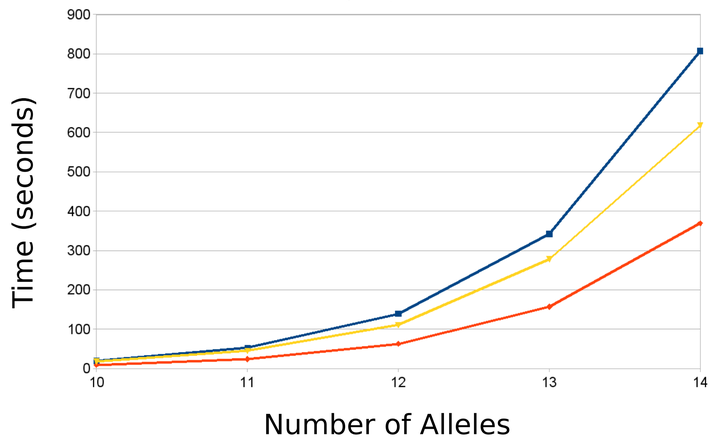


Figure 2.
Two sample pyroprints constructed using the 12 sample alleles in Figure 1. Each pyroprint is a combination of 7 alleles.

Results

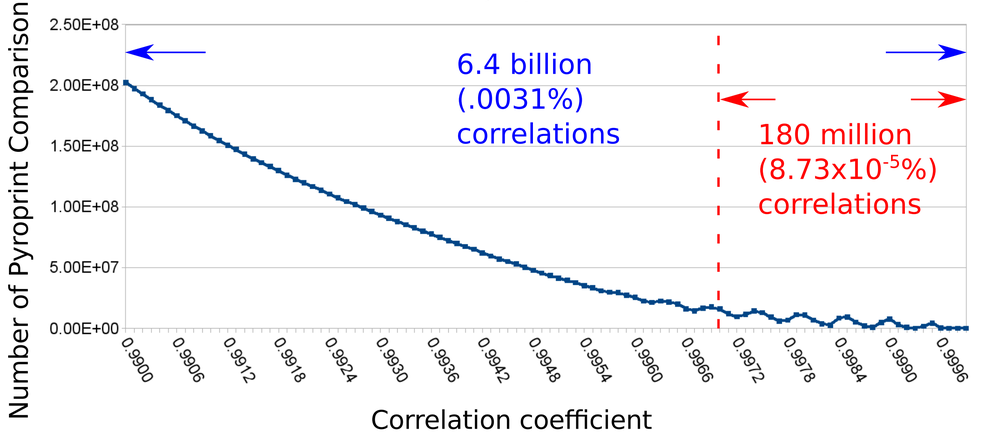
Runtime of Pyroprint Simulation



Dataset size based on Number of Alleles

# of Alleles	# of Isolates	# of comparisons
10	11,440	65,431,080
11	19,448	189,102,628
12	31,824	506,367,576
13	50,388	1,269,450,078
14	77,520	3,004,636,440
24	2,035,800	2,072,239,802,100

Distribution of Pyroprint Correlations



Conclusion

For our simulation, over 99.96% (2.065×10^{12}) of all correlation coefficients fall below 99% similarity. In preliminary *in vitro* studies conducted on pyroprints, multiple pyroprints of the same isolate showed correlations above 99.7%. However we found that only 8.73×10^{-5}% (180 million) of correlations fall into this category. This suggests that pyroprinting is sufficiently sensitive to distinguish bacterial strains.