

# Computation of Allelic Frequencies Using the “Cerberus” Beowulf Cluster

Giuseppe Sena, Grant Fitzgerald, Richard Hayes, Justin Jiang, Julian Kuk, Patrick Ryan, Chester Moses, Lindsay Schulman, and Bruce Jackson

Massachusetts Bay Community College  
Wellesley Hills, MA, 02481



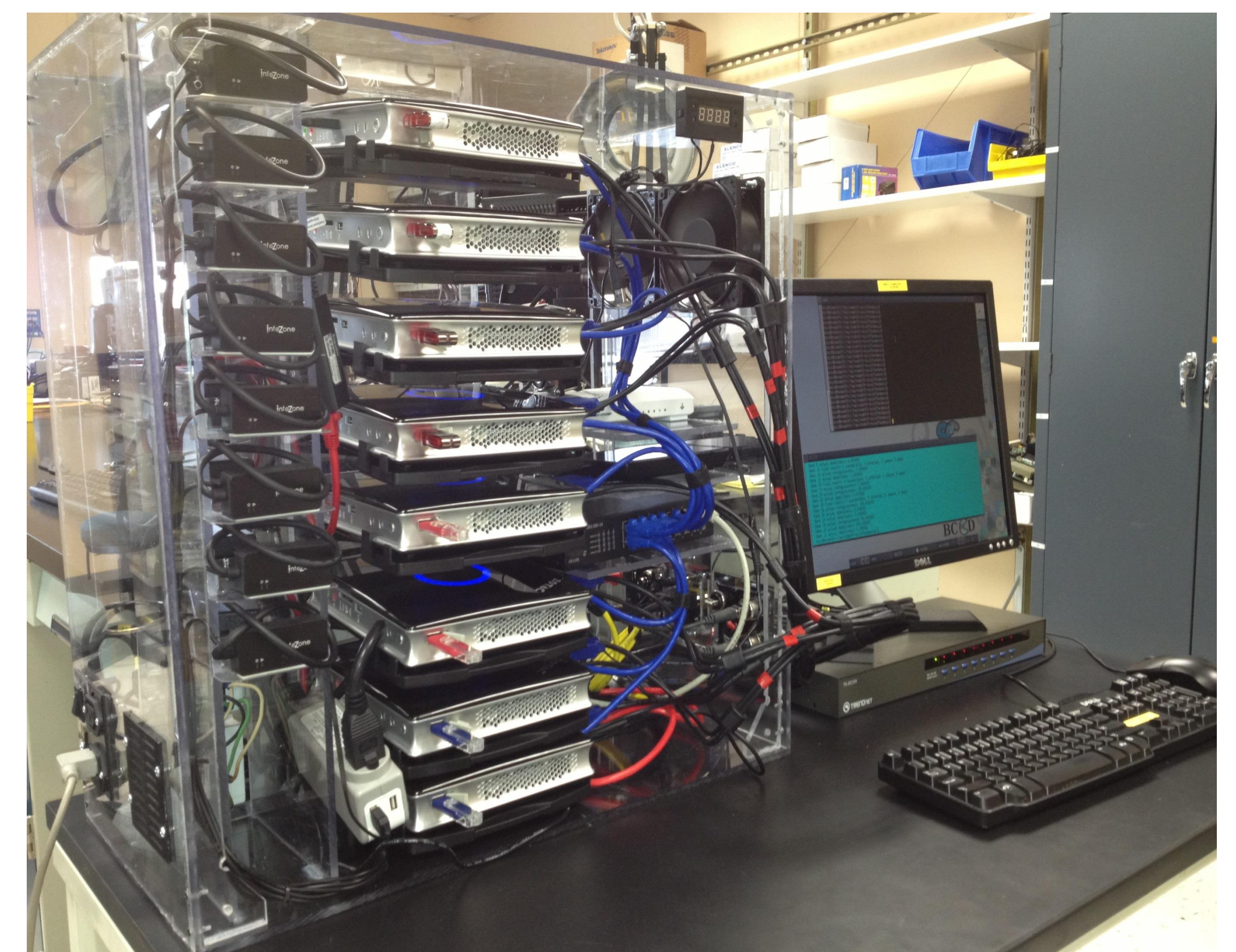
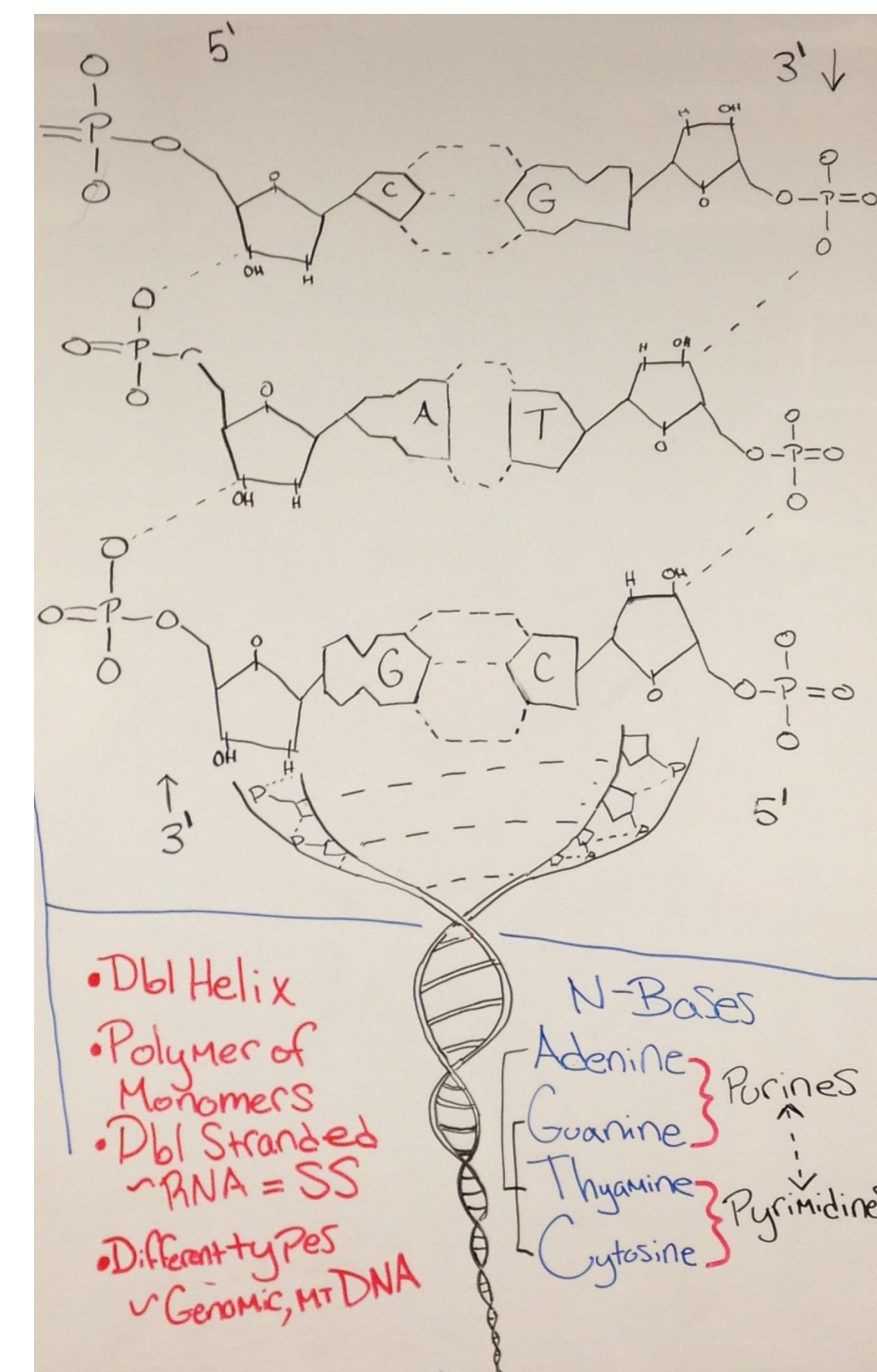
## Forensic DNA Use

- Short Tandem Repeats (STR) are specific repeats of a sequence of four or more nucleotides with no breaks in it:

(5' – ATAGATAGATAGATAG – 3')

where 'ATAG' is the repeating sequence

- FBI recognizes thirteen core loci [1]. These thirteen loci are organized in a table form, called an Allele Panel
- More than two alleles may present themselves in any allele of a panel. There should only be one or two STRs per allele (one inherited from their father and one inherited from their mother). This sort of mixture can occur by contamination from the scientist(s) testing the sample or by the presence of DNAs from multiple individuals.



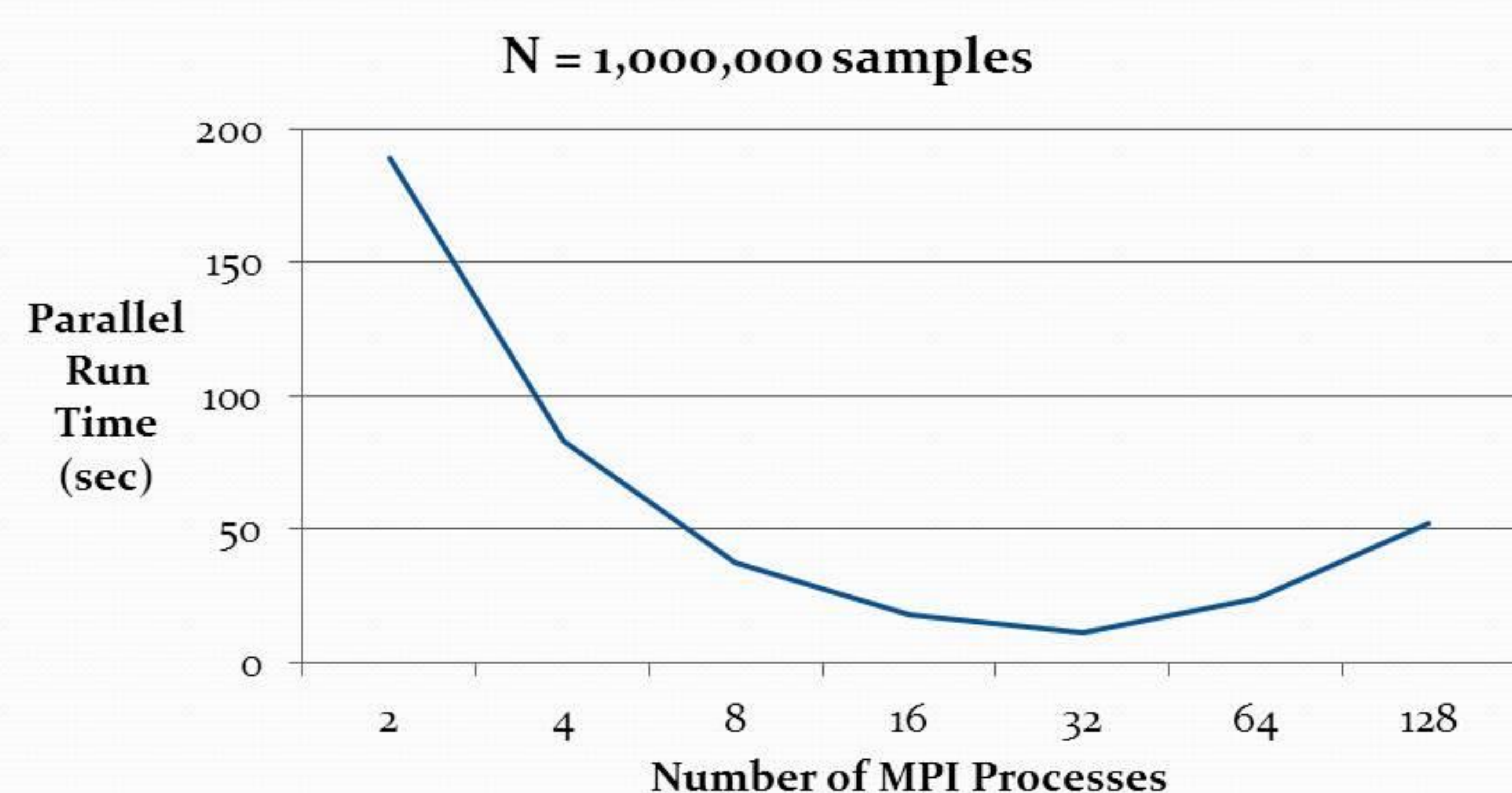
## Cerberus

- “Cerberus” is an eight-node dual-core Linux Beowulf cluster [2]. Funding for Cerberus came from a professional development grant [3]. In this work we used Cerberus to generate all possible allele combinations for a group of genes on DNA samples.
- We developed a distributed application on a cluster using the MPI (Message Passing Interface) library. We analyzed the performance and visualized the output using a simple GUI developed in Python.
- For one specific gene, if you have  $n$  alleles and  $r$  the number of choices (here 2 as each individual has 2 alleles), the number of combinations is:

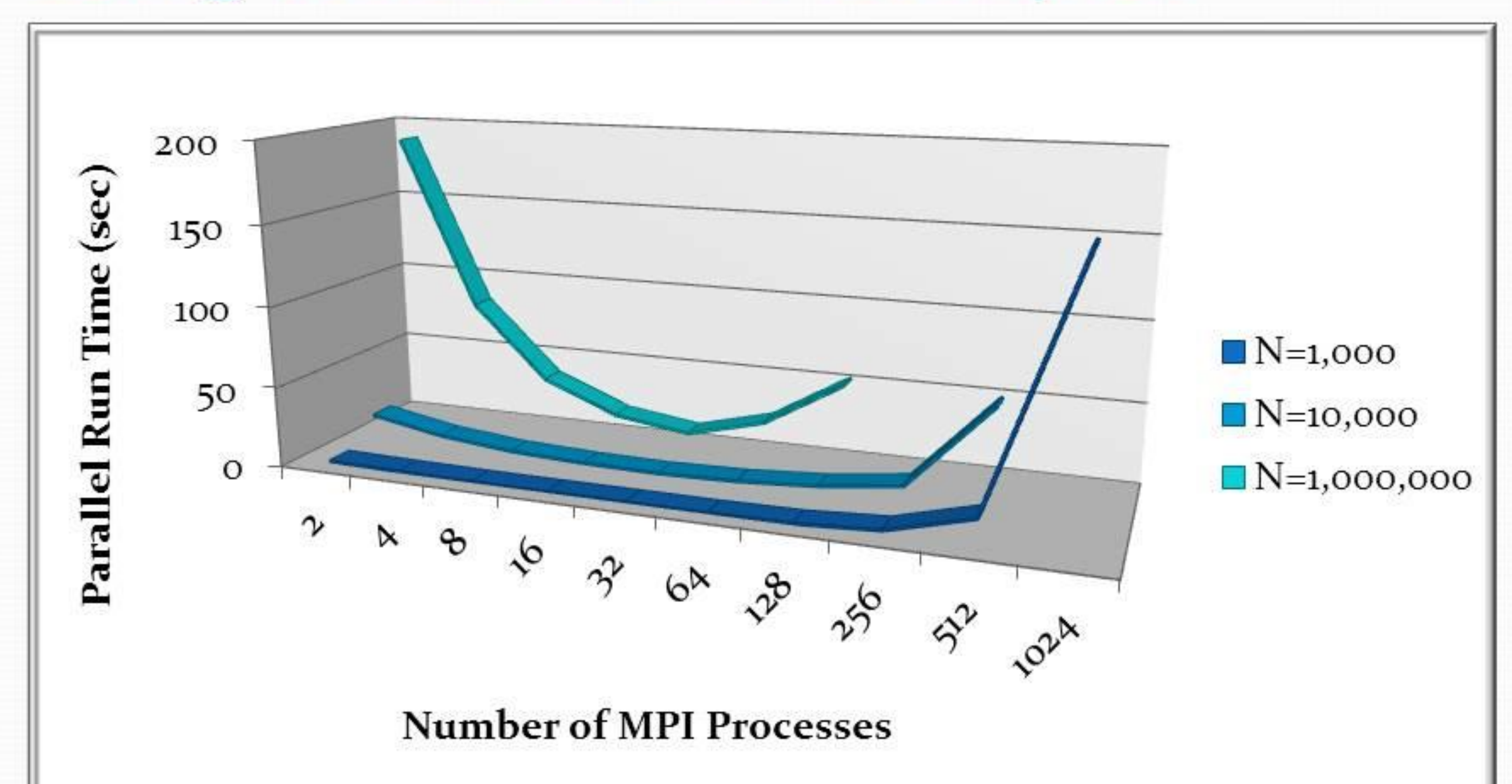
$$\binom{n}{k} = \frac{n!}{r!(n-r)!}$$

$$\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}$$

## Generation of Allele Combinations using Random DNA Samples



## Generation of Allele Combinations using Random DNA Samples



[1] Federal Bureau of Investigation, “Frequently Asked Questions (FAQs) on the CODIS Program and the National DNA Index System,” <http://www.fbi.gov/about-us/lab/biometric-analysis/codis/codis-and-ndis-fact-sheet> [2] Robert W. Lucke. *Building Clustered Linux Systems*. Prentice Hall PTR. 2005. [3] Giuseppe A. Sena. “Developing Labs and Microlabs for Adding Parallelism and Distributed Computing into Computer Science Curricula,” Professional Development Grants, Massachusetts Bay Community College, Computer Science Department. September 2012, unpublished.