Homeland Security applications with regard to microbial detection in environmental samples are among the many applications of this project. The main characteristic of genomic data is its large size. For example, the National Institutes of Health (NIH) sequence database has a total of over 65 billion nucleotides.

One of the most striking features of genomic DNA sequences is the extent to which repeated substrings occur in the genome. In C. elegans with a genome sequence of 100.2 million nucleotides over 7,000 families of repeat sequences have been identified. Families of repeat sequences account for about one half of the human genome. Repeat sequences come in many different flavors and are responsible for different functions and diseases. Finding repeats has applications in finding defective genes and in forensic DNA fingerprinting. It also allows us to find differences between genomes.

The fault-handling algorithm introduces both self-healing and self-optimizing attributes that specifically target the solution to the issue of improving the performance of the system when another worker node has scarce resources.

The data validation process proved the implementation of the parallel program worked correctly and the data validation process proved that there were no false-positives.