1. Compression of DNA coding sequences with Lempel-Ziv: are genes easier to compress than non-genic DNA?
2. Short read mapping on GPGPUs
4. Burrows-Wheeler transform on large DNA strings (this became Bowtie)
5. Automated phylogenetic analysis pipeline
6. Find document similarities using dynamic programming alignment and word-to-word similarity scores.
7. Protein sequence classification using neighbor-joining
8. Compare communication log files to protocol specs to detect bugs.
9. Find deleterious SNPs by using multiple sequence alignments
10. Build annotation pipeline
11. Characterize sequence conservation across species
12. Searching within compressed text - can boyer moore be applied to compressed text?
13. Multiple genome alignment
14. Phylogenetic reconstruction with neighbor-joining
15. Optimal multiple sequence alignment using heuristic pruning of dynamic programming table
16. Alignment of DNA sequences under specific error models (e.g. homopolymer stutter - assume strings of a same letter could be miscounted)
17. Alignment and/or assembly of DNA sequences represented in color space (SOLiD sequencing)
18. Optical map assembly and/or alignment
19. Evaluation of the "correctness" of sequence clustering algorithms
20. MapReduce implementation of (INSERT FAVORITE) algorithm.