As the sequencing technology improved to be shot-gun and deep coverage sequencing method, the old methods of k-mer counting tools (overlap layout consensus paradigm) became too slow, memory wasteful and out of time. Here, Jellyfish choose to develop a new algorithm within k-mers method (even though the dynamic programming can also be applied genome assembling), to which they combine multithreaded and lock-free hash table for counting. By multithreaded method, each thread in the process can run on a separate processor at the same time, realizing parallel execution. And by lock-free hash table, there is no waiting even during table resize. The above refinement of algorithms and data structure optimize the program for faster running and less memory cost.

As for other computational consideration, Jellyfish is well to be performed on super computer platform. Such as, introduce of MapReduce programing paradigm and implement of “compare and swap (CAS)” in lock-free data structure for k-mer’s counting. They are all specially designed for parallel execution.

On the other hands, the statistic of k-mer count can also be applied to estimating genome size. That is if an ample fraction of k-mer appears “x” times, we can surmise the sequencing coverage to be around “x” times and derive the genome size from “x” and the total length of the reads. Furthermore, analysis of k-mers distribution has potential for investigating new biological insights, and also for rapidly estimating the overall characteristics of a genome (such as genome size, heterozygosity rate and repeat content from unprocessed short reads).