Wyatt Eng

7/14/2019

1. Describe how the assembly changes with different k-mer values using the assembly

statistics you have collected. How does the contig length distribution change?

When you increase the k-mer length the number of contigs decreases, the N50 increases, and the mean contig length increases. As the k-mer size increases the amount of contigs significantly drops but increases the average length of the contigs meaning that the contig length distribution will favor longer contigs.

2. How does an increased coverage cutoff affect the assembly? What is happening to the

de Bruijin graph when you change the value of this parameter? How does velvet

calculate its value for ‘auto’?

Increasing the coverage cutoff decreases the length of the genome assembly because it is less likely that a contig has high coverage. Having more stringent data means that you are not keeping as many contigs with a high coverage cutoff. Which is why your results would have a lower number of contigs and a higher mean depth of coverage because you are filtering out the contigs that were sequenced less. With a higher coverage cutoff, the de Bruijin graph would be more clear-cut and more of a direct path because of the higher scaffold accuracy. ‘Auto’ can be used if coverage is uniform. ‘Auto’ sets the expected coverage to the calculated length weighted median contig coverage and the coverage cutoff to half of the expected coverage.

3. How does increasing minimum contig length affect your contig length distribution and

N50?

If you are filtering out short contig lengths, the mean contig length will increase, which also leads to a higher N50 value. Increasing the minimum contig length with skew the contig length distribution towards longer contigs.