

Evaluating Assemblies

(find and work with a partner)

N50 is a common evaluation metric for assemblies. For a set of contigs, N50 is the greatest length such that at least half the bases of the assembly are in a contig with length N50 or longer. A higher N50 is “better”. This metric is unfavorable to assemblies formed from many short contigs; it favors assemblies with fewer, longer contigs (which most resemble a single chromosome).

For the sets of contig lengths below, compute N50. *Hint: N50 should always match the length of one of the contigs in the assembly.*

1. {100, 70, 60, 50, 50, 40, 30}
2. {10000, 150, 30, 20}
3. {100, 100, 100, 100, 100, 100}
4. {1000, 250, 250, 250, 250}
5. {1000, 250, 250, 250, 250, 5}
6. What are the *strengths* of N50 as an evaluation metric?
7. What are the *weaknesses* of N50 as an evaluation metric?
8. What might be a better way of evaluating assemblies?