

# Assembly

Learning objectives:

1. Identify the most basic elements underlying the *de novo* assembly process.
2. List potential strategies and complications involved in the software performance of the *de novo* assembly process.

Level of Difficulty: 3/5

Once you've got a filtered set of reads, you may decide you want to piece them together to try to reassemble the genome(s) they came from. This process is kind of like if you were to piece back together a newspaper that had been run through a shredder. What strategies might you use while reassembling? Say you've taped back together as much as you can of the original newspaper, but because the pieces were so small and the shredder ate some of them, the best you can do is several patchy reassembled sections that you can't connect together, and you're left with a bunch of repetitive words and characters like "a", "the", and commas. The sections of assembled pieces are like metagenomic "contigs", and those repetitive leftovers may still be useful down the road in read mapping.

Follow up questions:

- How are nucleotides different from words in a newspaper or song? What context clues in DNA might be used instead of common words, sentence structure, punctuation, and font?
- What other clues in the DNA might guide a metagenomic assembly?
- Why do you suppose some pieces might be missing in metagenomic assembly?
- How might repeated genes affect the assembly process?

**To do:**

A few copies of two common nursery rhyme songs representing different microbial genomes have been cut up into pieces, and some of the pieces may be missing. Try to reassemble as much of the rhymes as you can from the following list of “reads”. Note that a few areas of the rhymes have repeated words, and both rhymes loop back on themselves at the ends, kind of like a circular genome.

-vous? dorm	es matin	nnez les	row row
? Frere J	g dong d	onnez le	row your
am row r	g dong Fr	ous? Dor	rrily lif
atines son	ines ding d	ous? Fre	t gently
cques Fre	ing din	ous? sonn	ut a dre
dream r	ing don	r boat ge	w row r
e is but	m merrily merr	rere Jacq	w your bo
eam ro	merrily mer	rilly merrily	wn the s
eam row r	merrily mer	rmez-vo	y down t
errily merrily	ng ding	rmez-vou	y life is
es dorm	ng ding d	row ro	z les mat