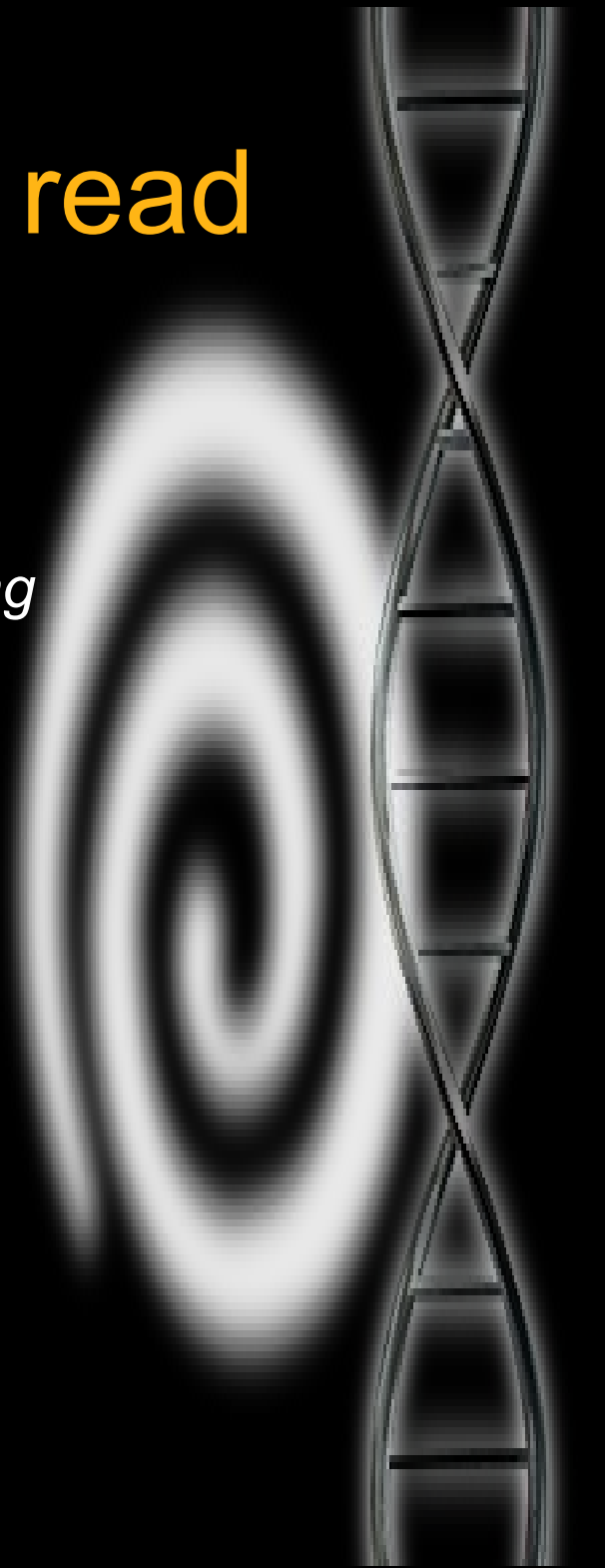
The feasibility of short read sequencing

Based on:
An analysis of the feasibility of short read sequencing
Whiteford et al.
Nucleic Acids Research, 2005, Vol. 33, No. 19

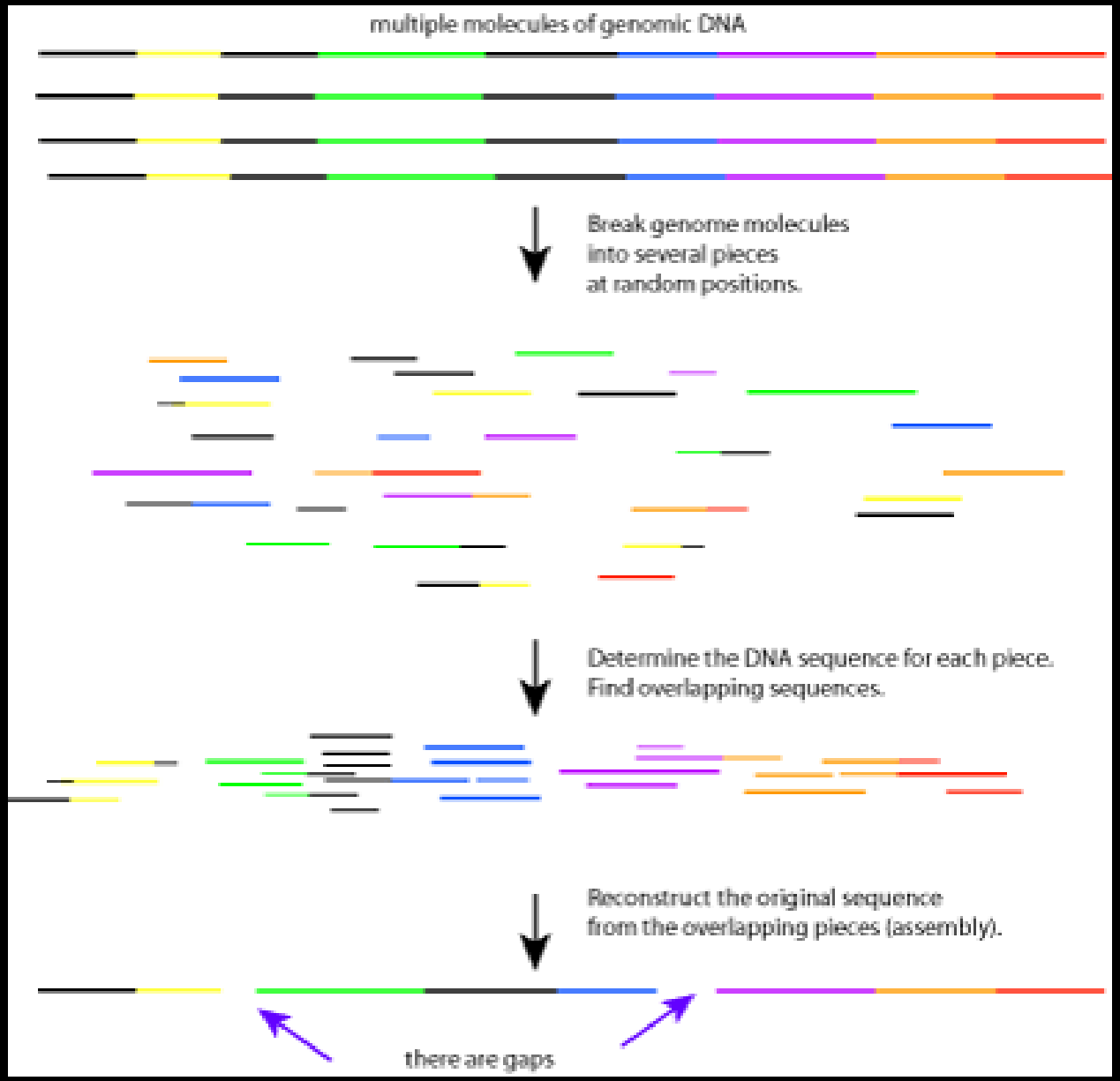
Thomas Unterthiner
0416441



How good is NGS?

- Resequencing?
- De-novo Sequencing?

All the NGS-technology you'll need to understand:



[S. Porter, http://scienceblogs.com/digitalbio/2007/02/sequencing_a_genome_part_vi_ah.php]

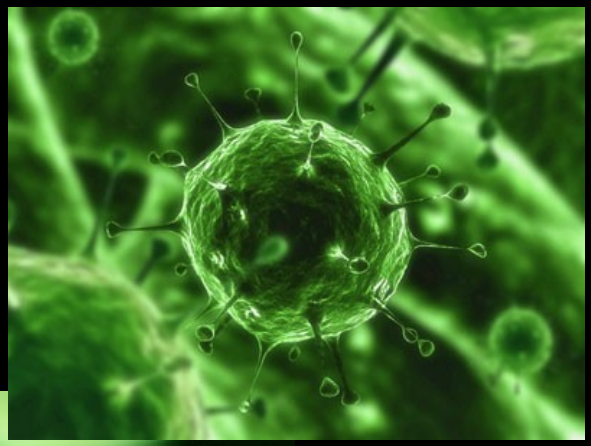
Key Problem of NGS:

The shorter the read, the higher the probability that a read will occur more than once in the sequence

How well does NGS work

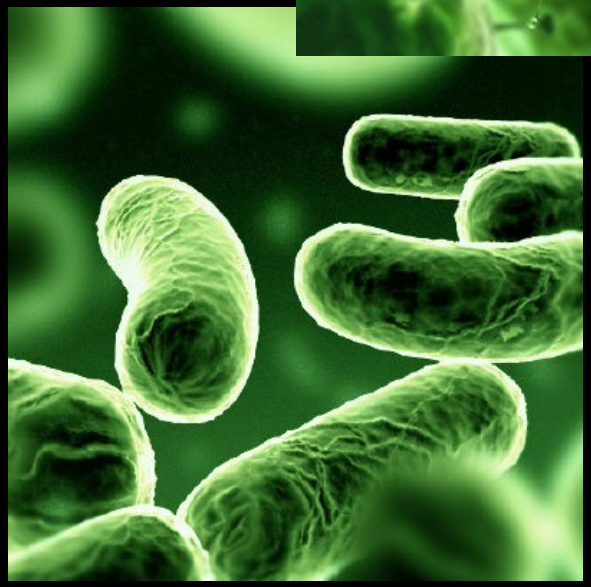
- … for small Viri

  (Genome-length: a few Kb)



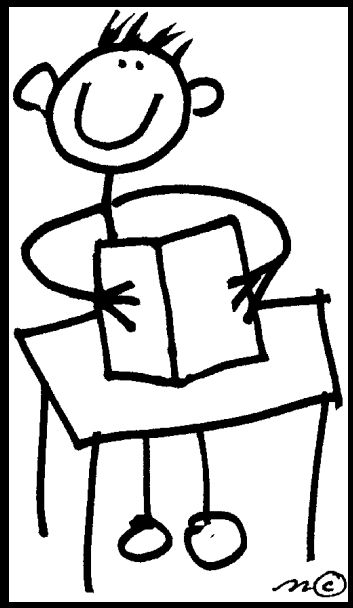
- … for Bacteria

  (a few Mb)



- … for Humans

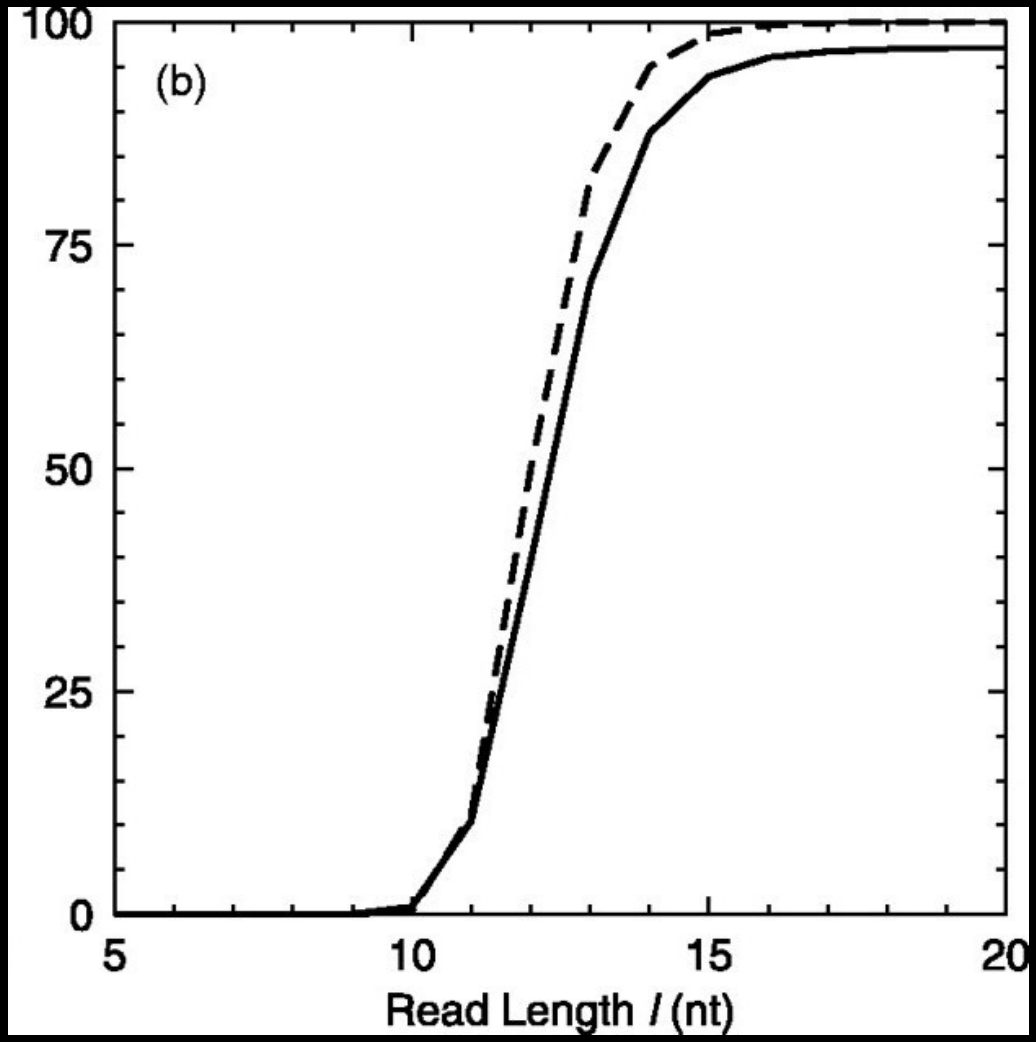
  (almost a Gb)



Method

- Take complete genome
- Calculate all possible reads of predefined length
- Reassemble all the reads

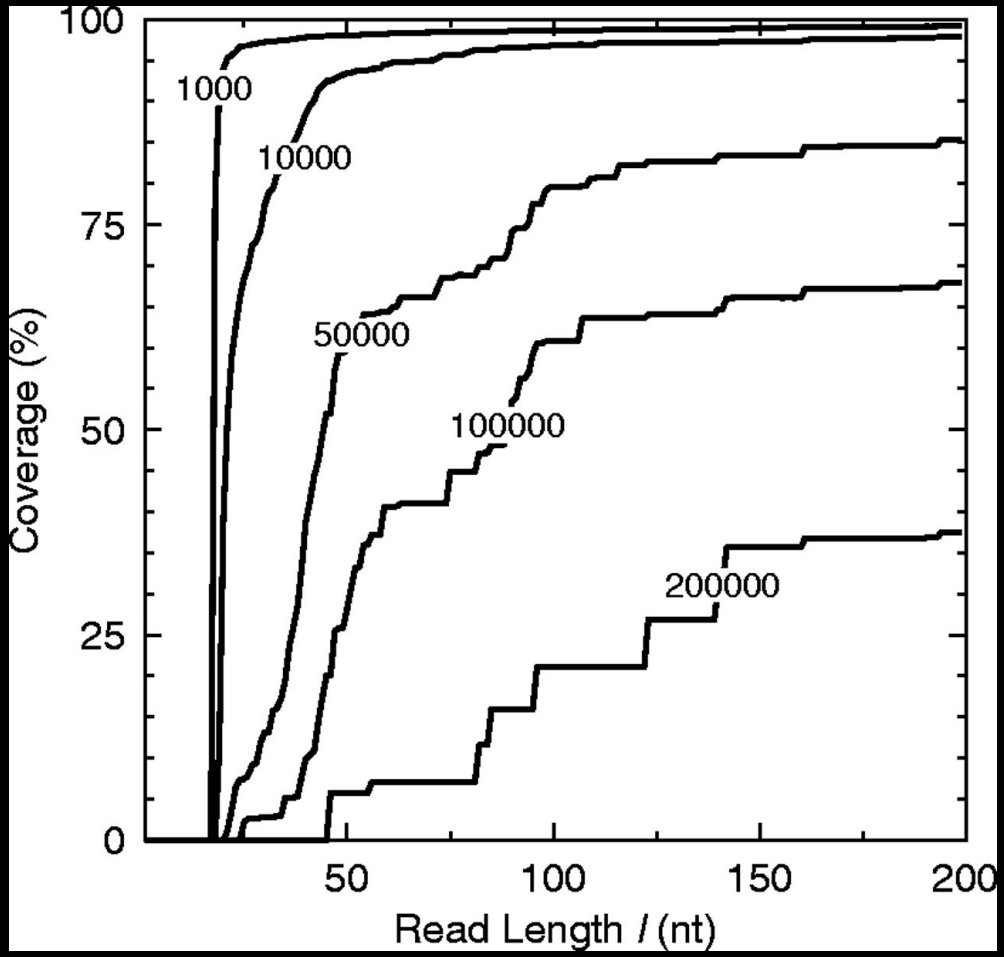
How many of the reads are unique?



E. Coli

Note the sigmoidal shape!

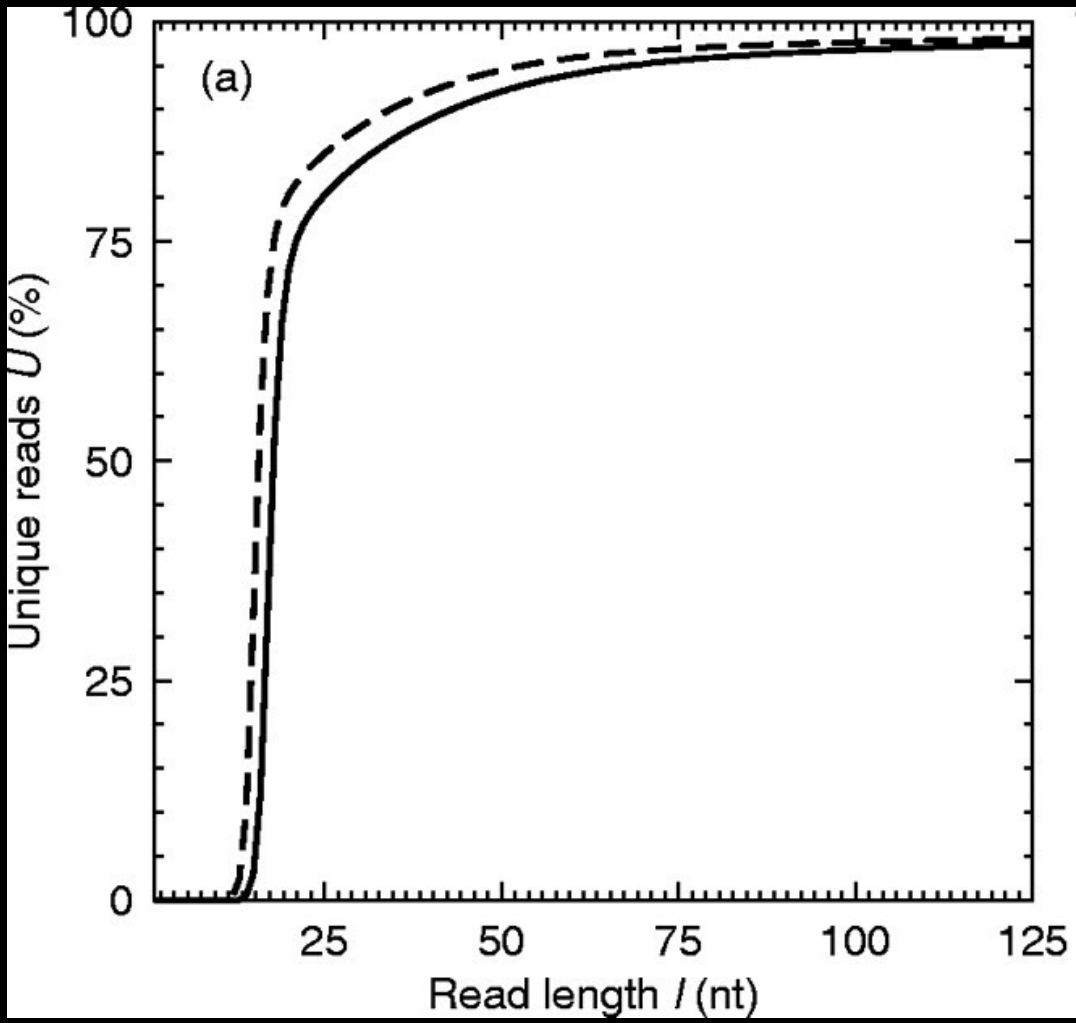
How good does reassembling work?

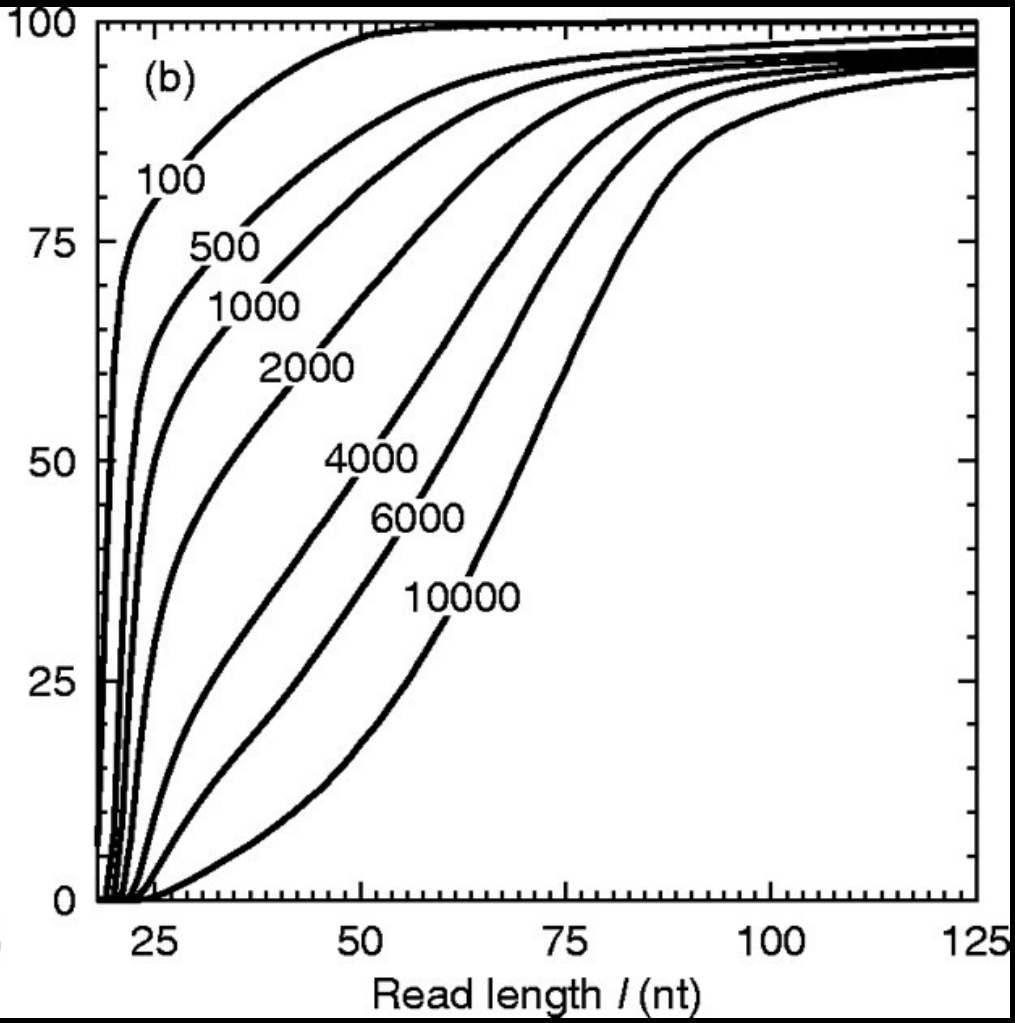


Percentage of E. Coli genome covered by contigs* greater than a threshold length

*Contig = a contiguous fragment of sequence data
(there must not be two or more potential reads that would extend a contig differently)

… and on the human genome:

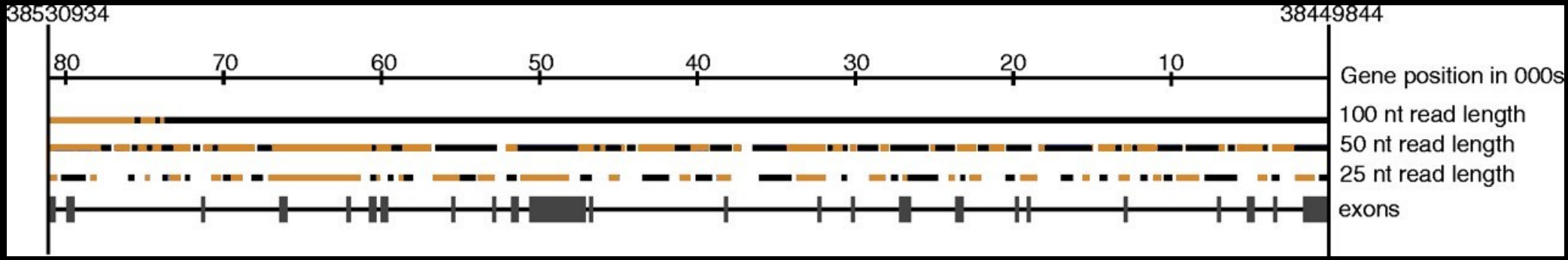




Dashed: chr1
Solid: whole human genome

Percentage of chr1 covered by contigs greater than a threshold length

Reassembling a gene



Reassembled contigs longer than 200bp in the 81 090bp of the BRCA1 gene (chr17)

Conclusion

- Lengths of 18-25bp are enough for Re-Sequencing/De-Novo sequencing of viral and bacterial genomes

- Human genome “reasonably” re-sequenceable

Beware!

- repeated sequences
- read quality

- Usually you don't have all possible reads

That's all, folks

Any Questions?

Forget Google, ask me!