PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads

Ryan R. Wick*, Louise M. Judd, Claire L. Gorrie, Kathryn E. Holt

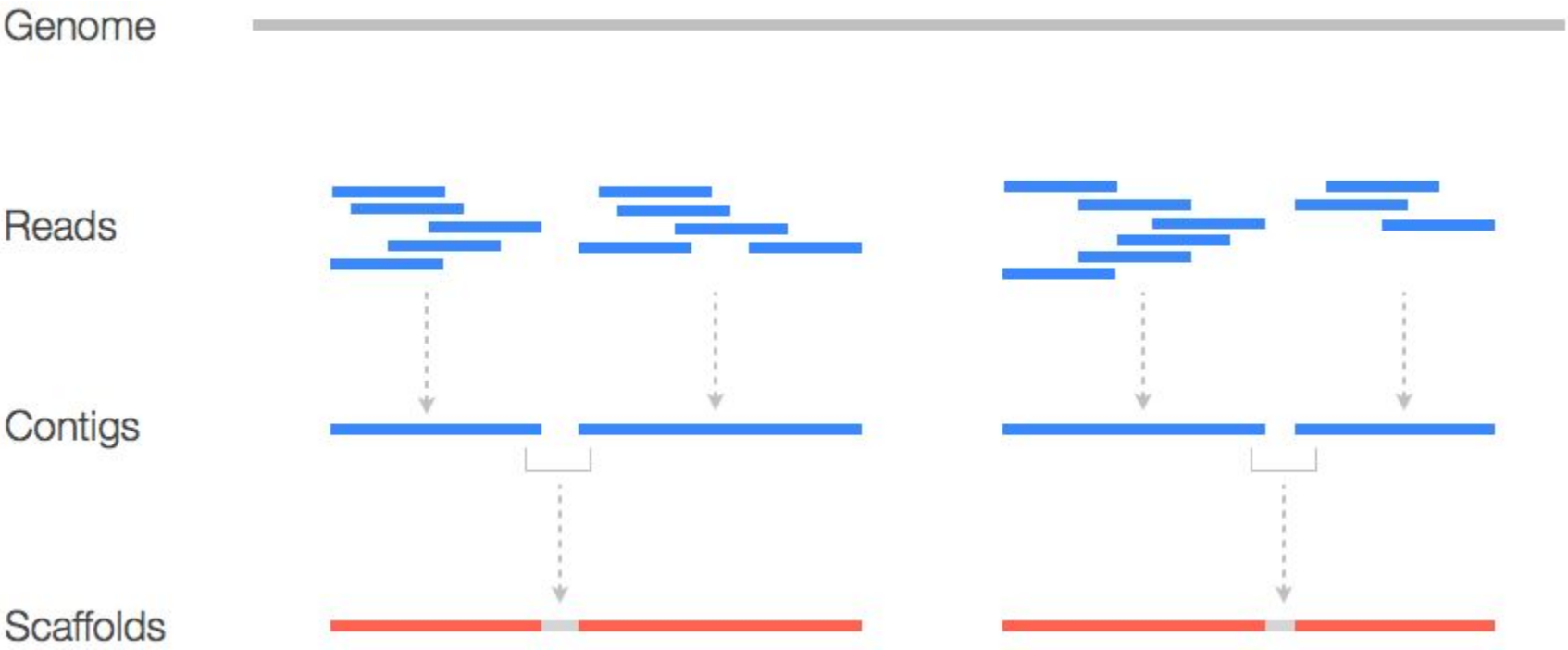
Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Victoria, Australia

Vladimir Nikolic[1,2], Diana Lin[1,2]

September 12, 2019

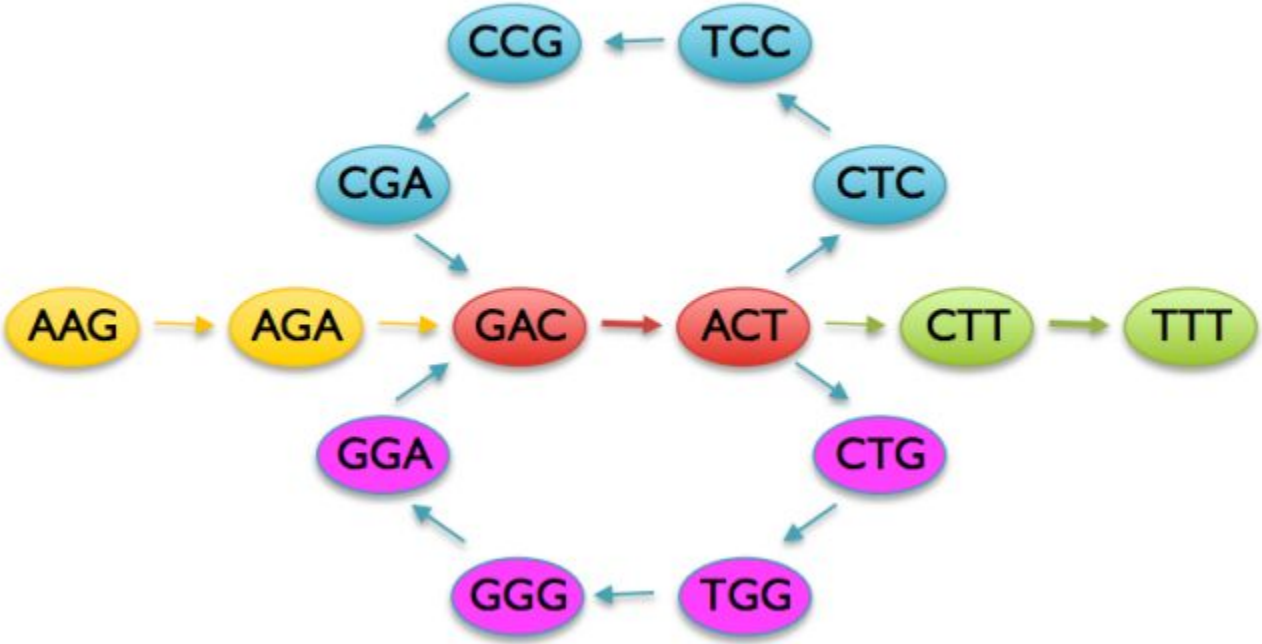
[1] Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada
[2] Bioinformatics Graduate Program, University of British Columbia, Vancouver, BC, Canada

What is genome assembly?



https://www.biostars.org/p/253222/

de Bruijn Graph



http://bioinformaticsinstitute.ru/sites/default/files/slides2.pdf

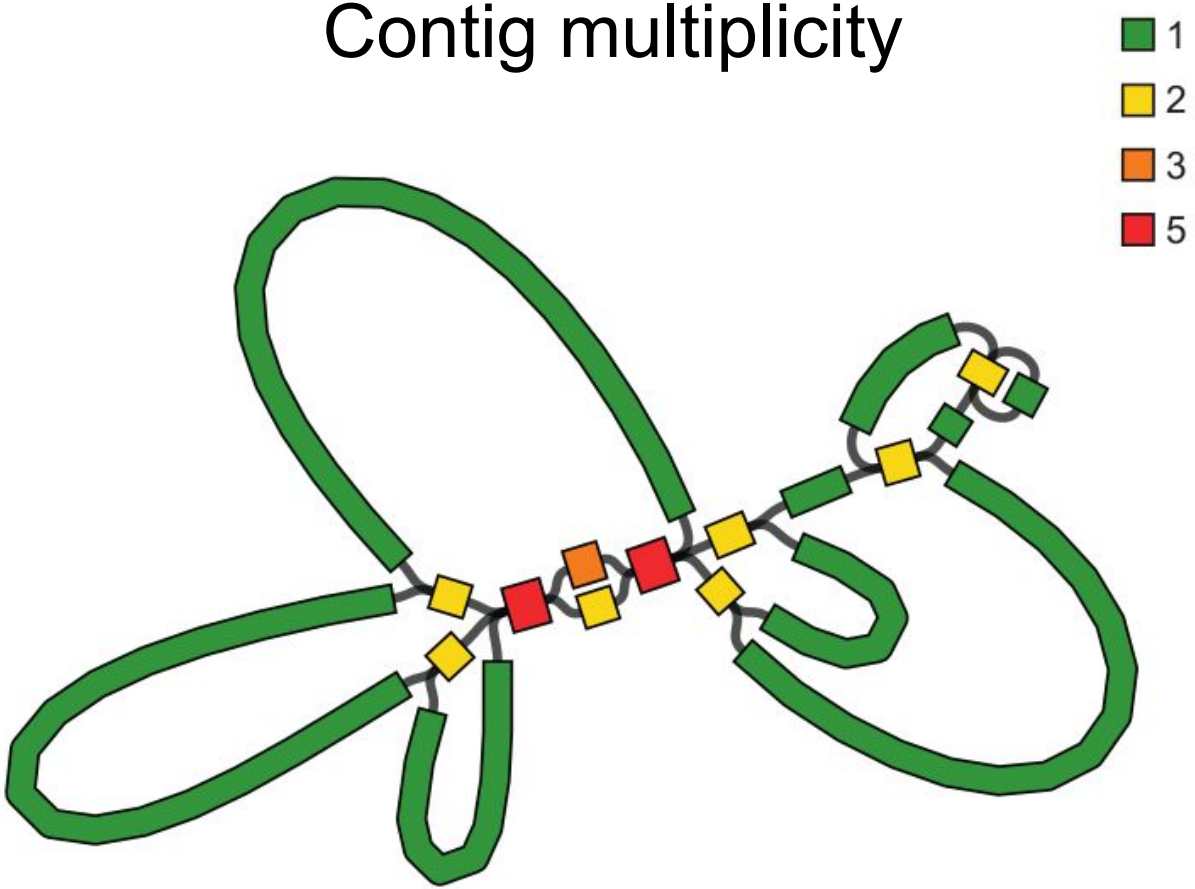
Introduction

- Why perform a hybrid genome assembly?

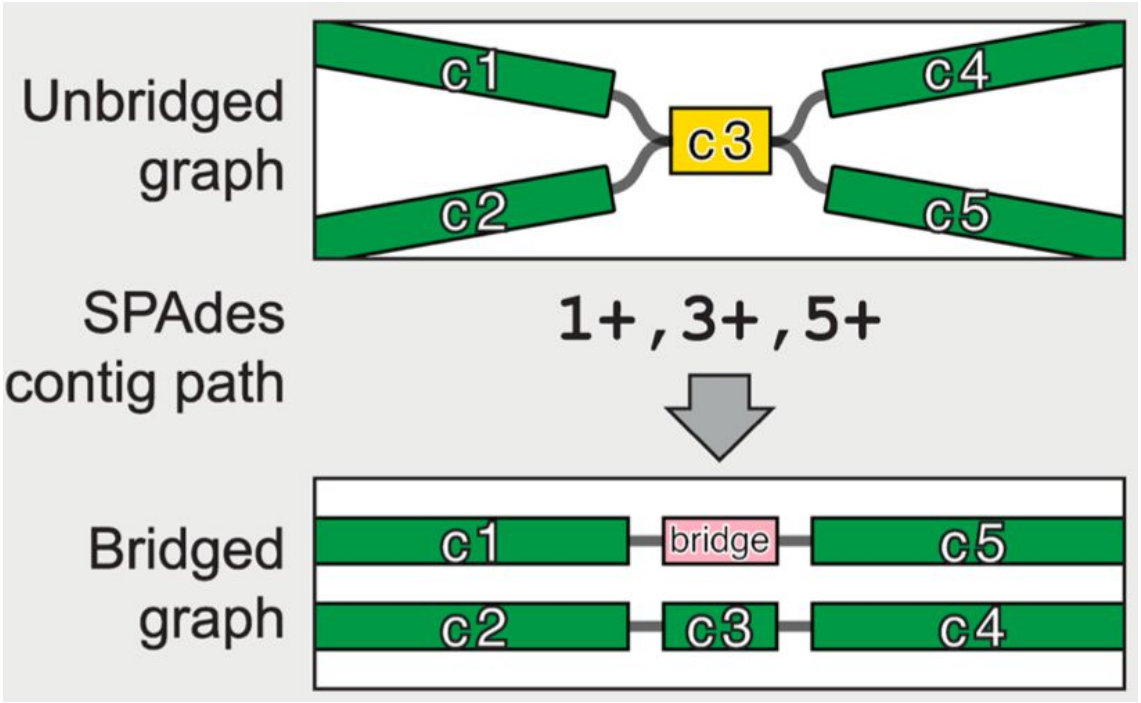
SHORT READS		LONG READS	
Pros	**Cons**	**Pros**	**Cons**
● Low cost per base ● < 1% per-base error rate	● <= 500 bp, shorter than most repetitive regions	● >= 10 kbp, longer than most repetitive regions	● High cost per case ● 5-15% per-base error rate
Fragmented assembly for more genomes		Complete assembly for fewer genomes	

- Solution: UNICYCLER
  - Short reads to produce accurate contigs
  - Long reads used to scaffold and simplify the graph

Contig multiplicity



Short Read Bridging



- Find contig path (from SPAdes) that are between single-copy contigs
- Bridge the graph (directly, and by elimination)

Fig 1. Key steps in the Unicycler pipeline.

Long Read Bridging - Unbridged Path

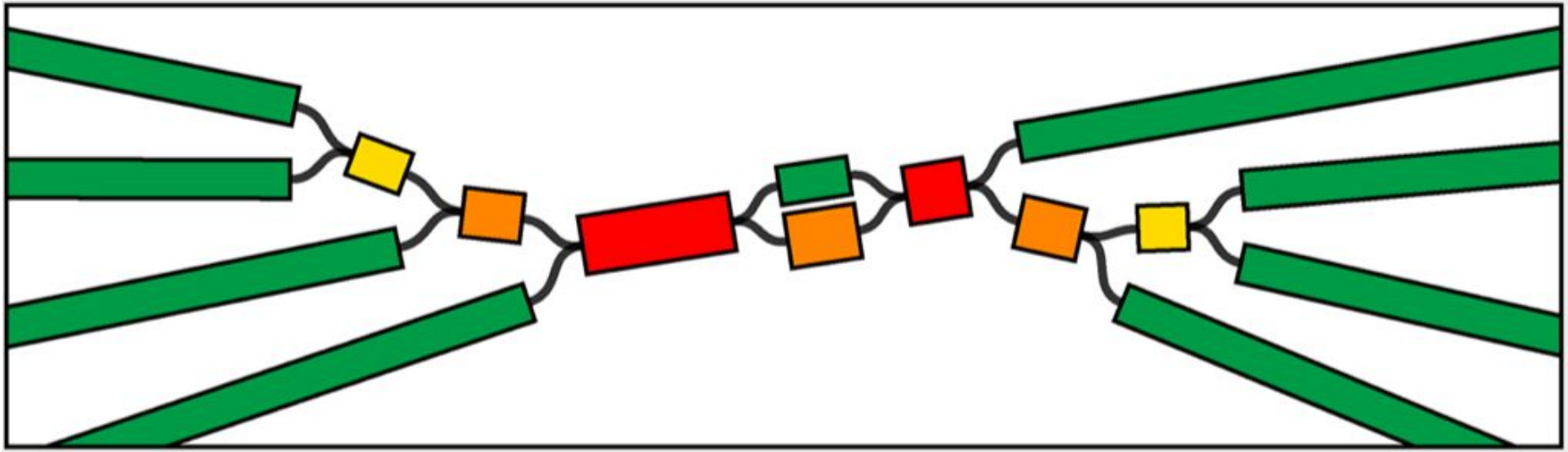


Fig 1. Key steps in the Unicycler pipeline.

Long Read Bridging: Long Read Alignment

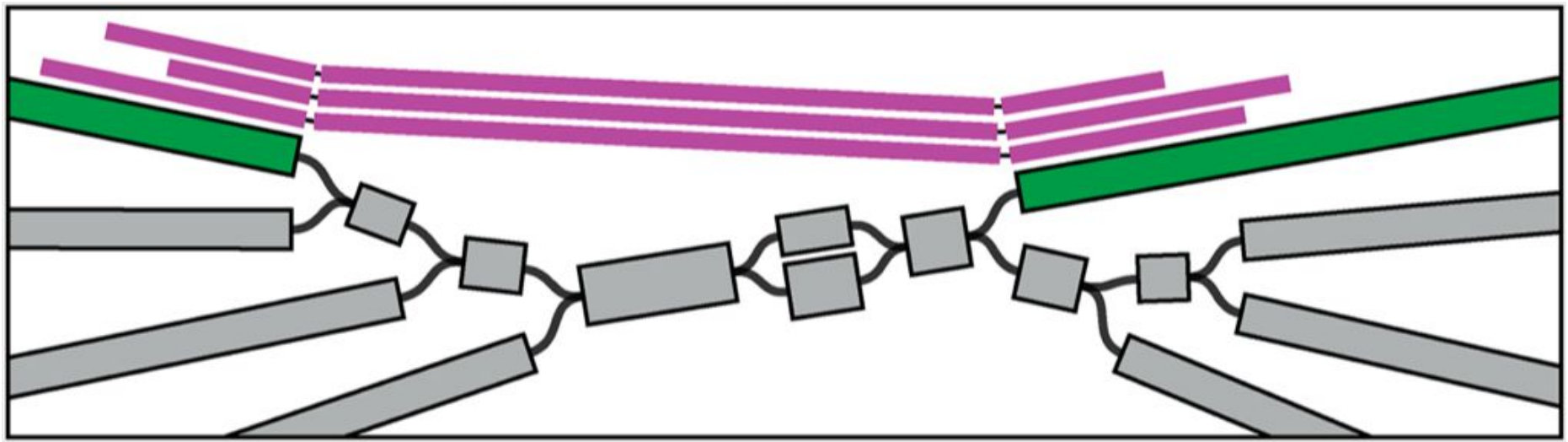


Fig 1. Key steps in the Unicycler pipeline.

Long Read Bridging - Long Read Consensus

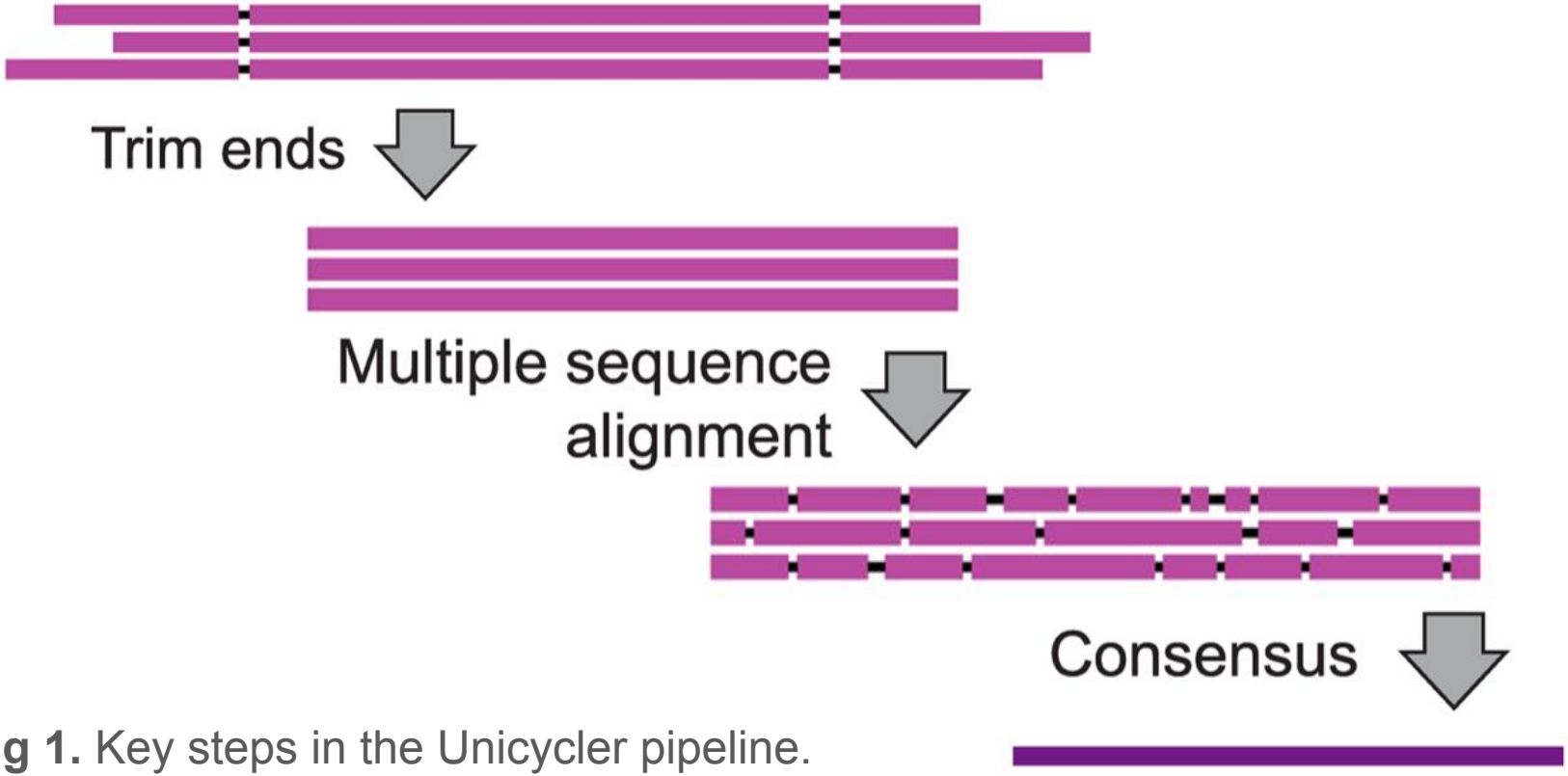


Fig 1. Key steps in the Unicycler pipeline.

Long Read Bridging: Finding the Path

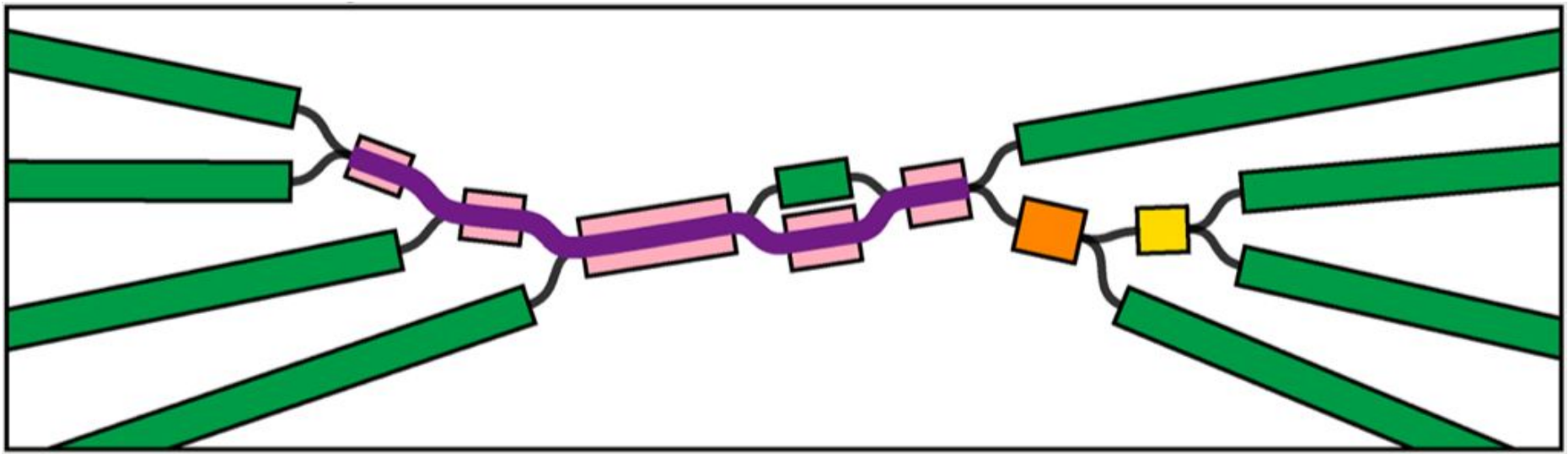


Fig 1. Key steps in the Unicycler pipeline.

Long Read Bridging: Finding the Path

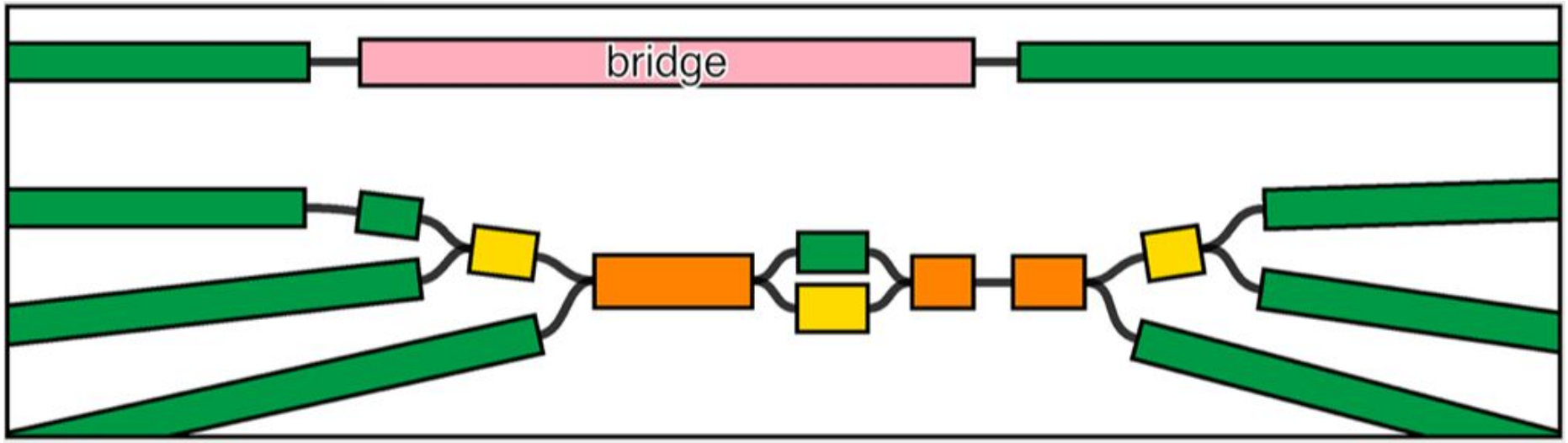


Fig 1. Key steps in the Unicycler pipeline.

Methods - Bridge Application

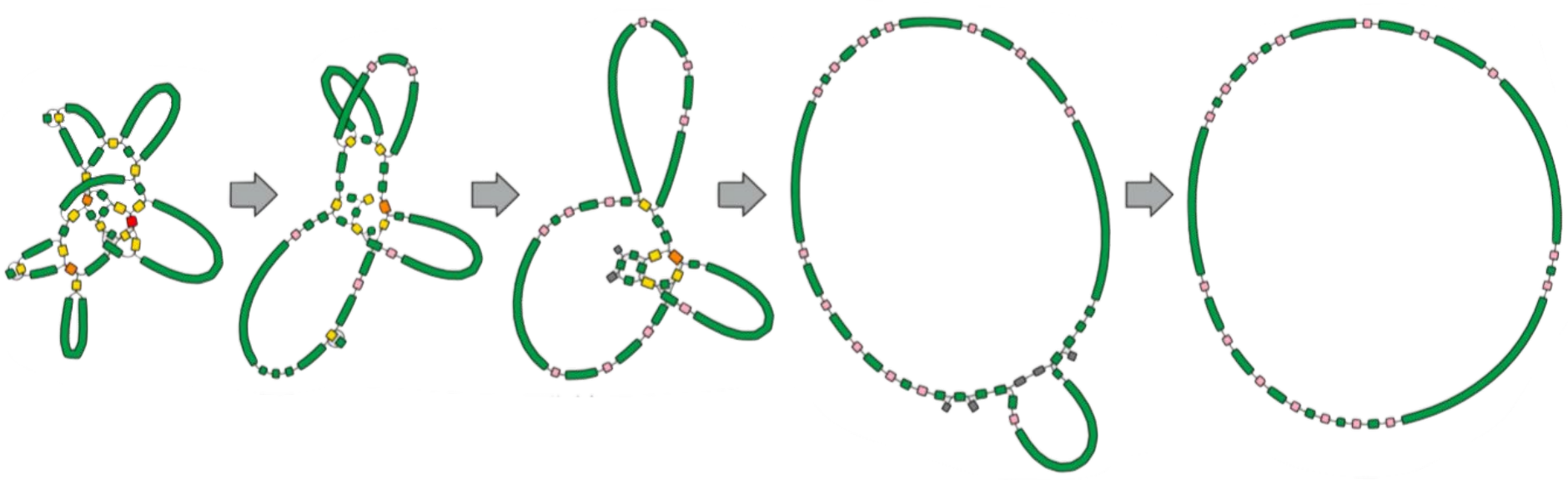


Fig 1. Key steps in the Unicycler pipeline.

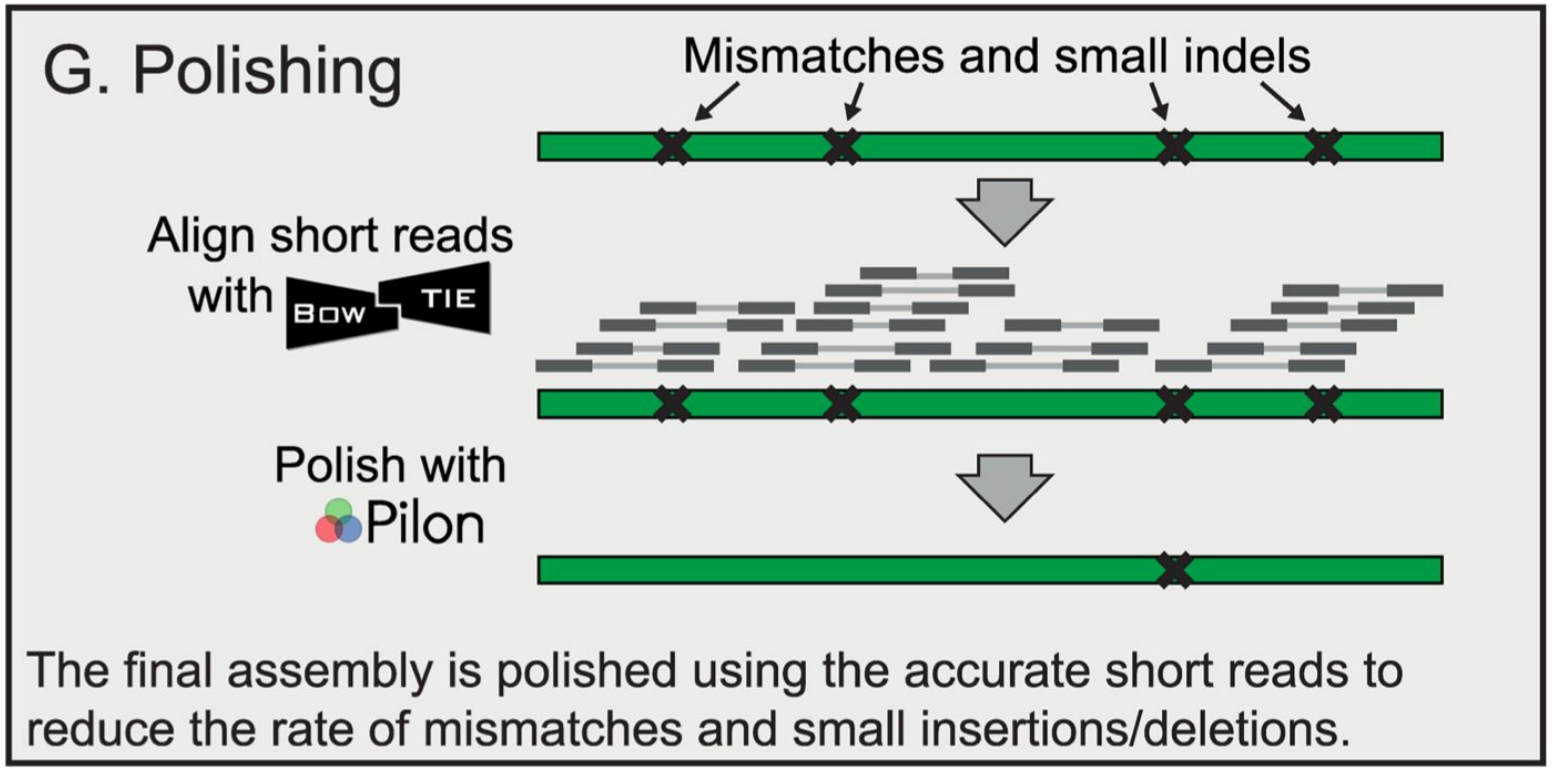


Fig 1. Key steps in the Unicycler pipeline.

Conclusion

Unicycler is a hybrid assembler that allows researchers to assemble a *large* number of **complete, yet accurate** bacterial genomes in a cost-effective manner, better than the assemblies achieved using short reads or long reads alone.