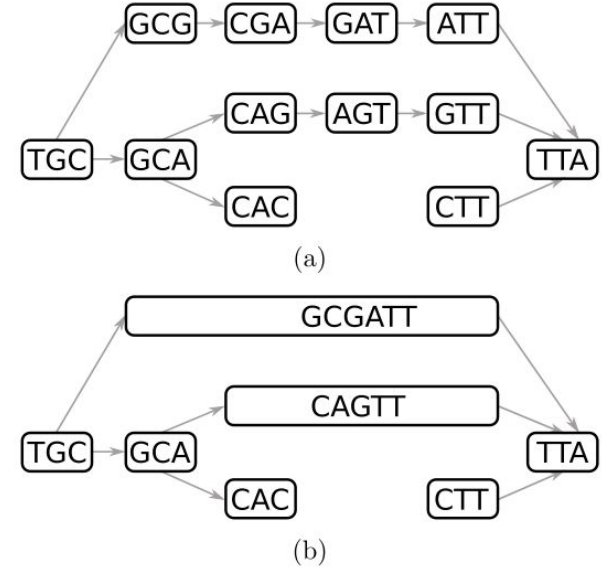# Insights from Compacted De Bruijn Graphs

Ömer Faruk Özdemir

De Bruijn graphs

OLC framework is the defacto standard but,

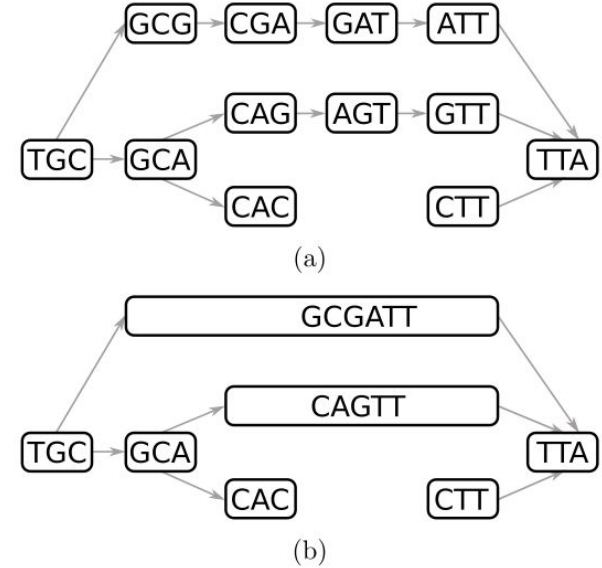De Bruijn is still used to assemble and correct long reads

Fig. 1 A de Bruijn graph in **a** and its compacted counterpart in **b** using 3-mers. For simplicity, reverse-complements are not considered

De Bruijn graphs

OLC framework is the defacto standard but,
De Bruijn is still used to assemble and correct long reads

Problems such as

- **de novo transcriptome assembly**
- variant calling
- short read compression
- short read correction
- long read correction
- short read mapping

Fig. 1 A de Bruijn graph in **a** and its compacted counterpart in **b** using 3-mers. For simplicity, reverse-complements are not considered

Literature

- Faster
- Memory-efficient
- Smaller
- Structural Features(Bubbles, variant calling)
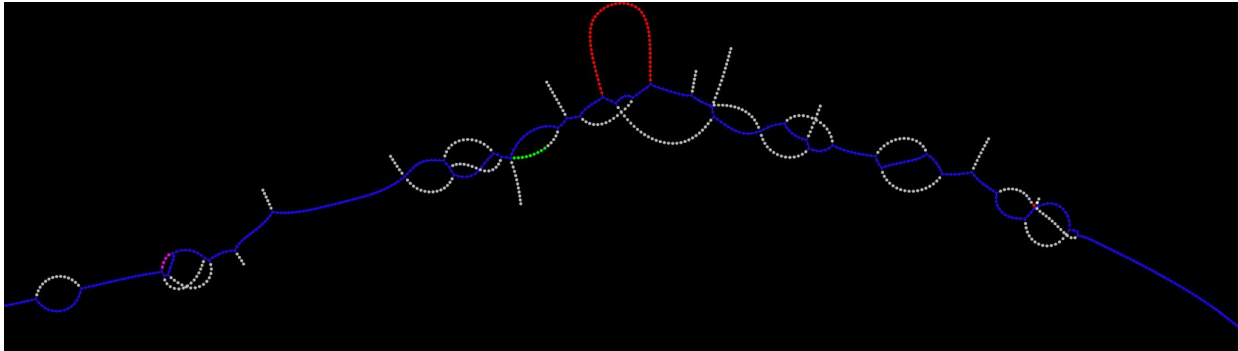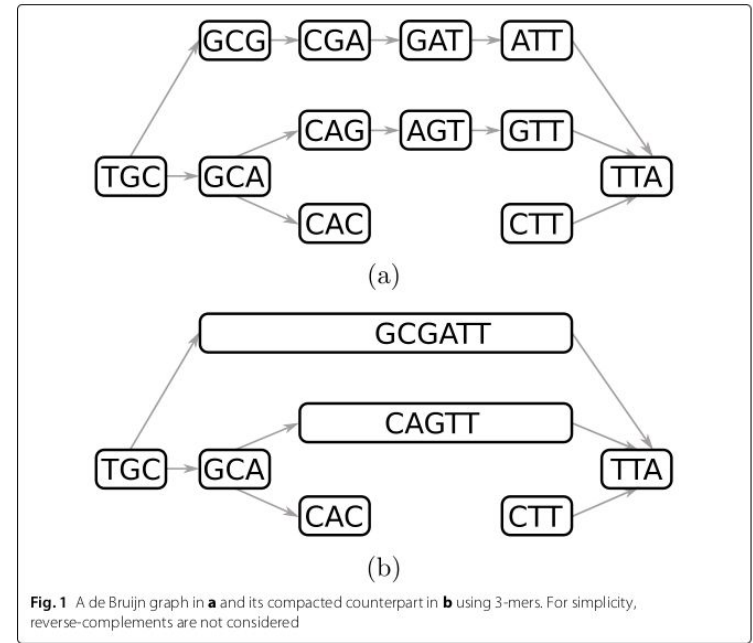- No work on graph properties
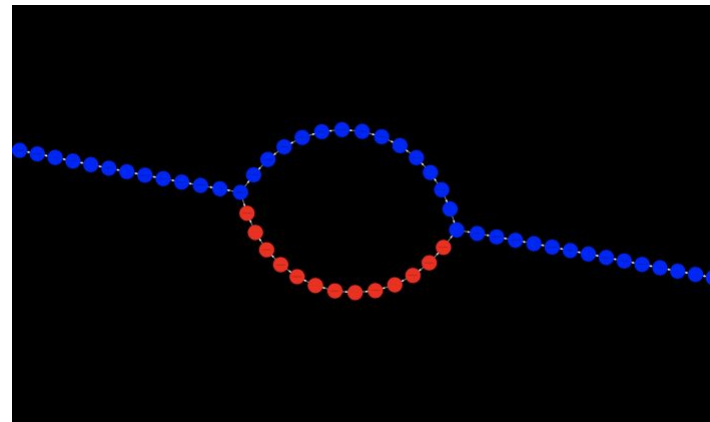
Applications of de Bruijn graphs in microbiome research
DOI: 10.1002/imt2.4

Opportunity

- Connectivity
- Paths
- Random walks
- Centers
- Node #
- Edge #
- ...

Fig. 1 A de Bruijn graph in **a** and its compacted counterpart in **b** using 3-mers. For simplicity, reverse-complements are not considered

Completed Steps

1. Literature survey
2. Install Bifrost
3. Generate cdbjgs
4. Working on analyzing GFA format

Future Steps

1. Complete graph analysis
2. Compare graph properties between different species
3. Compare graph properties between the same species
4. Try to find correlation between local/global graph properties and phenotypes&diseases

Lessons Learnt

1. Familiarity with toolchains
2. Importance of available libraries
3. Preplanning and discovery
4. Estimating cost of the prior steps

Insights from Compacted De Brujin Graphs

Ömer Faruk Özdemir
linktr.ee/FarukO