

# Sequence-read extraction from Counting de Bruijn Graphs

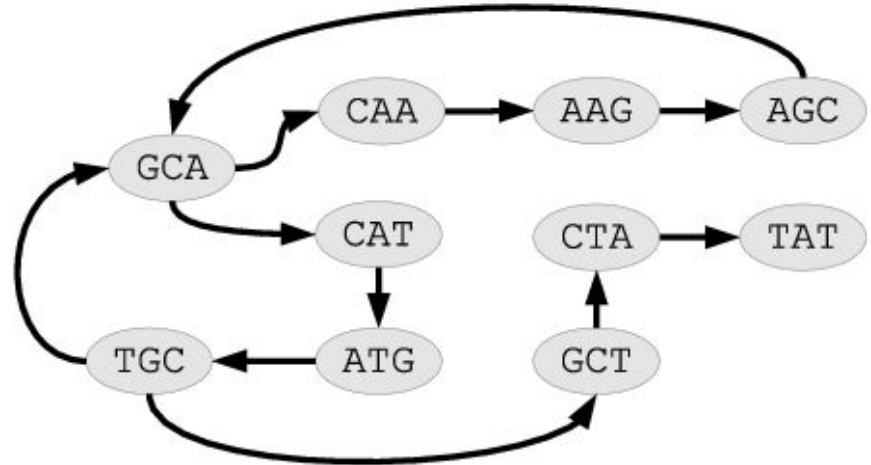
---

Dmytro Horyslavets

Mentor: Andre Kahles (ETH Zurich)

# Background

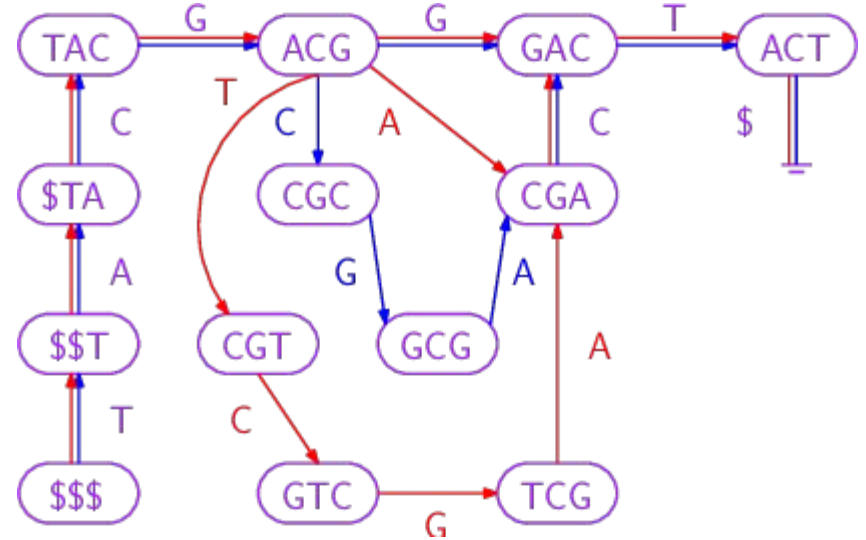
- **De Bruijn graph** is a directed graph of overlaps between sequences of symbols
- In bioinformatics DBGs are used for de novo assembly of sequencing read into a genome (and not only).



$R_1$ : TGCAAGCA  
 $R_2$ : AAGCATGC  
 $R_3$ : CATGCTAT

# Background

- A **Colored** (annotated) DBG is a generalization to multiple samples.
- A **Counting** DBG is a notion generalizing annotated DBG by supplementing each node-label relation with additional attributes.

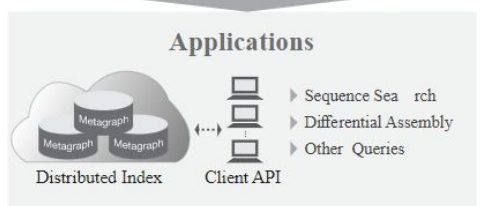
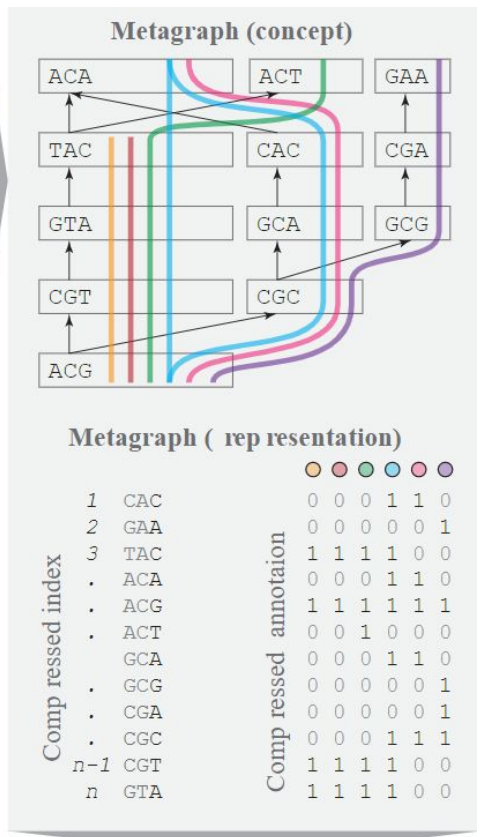
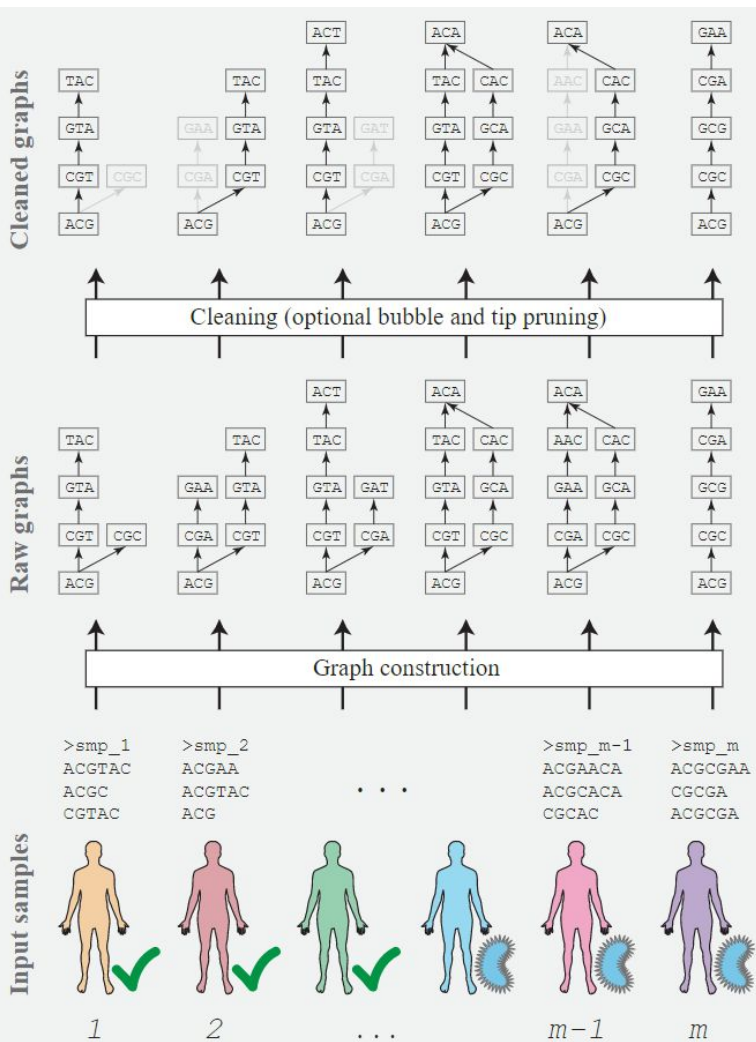


# Background

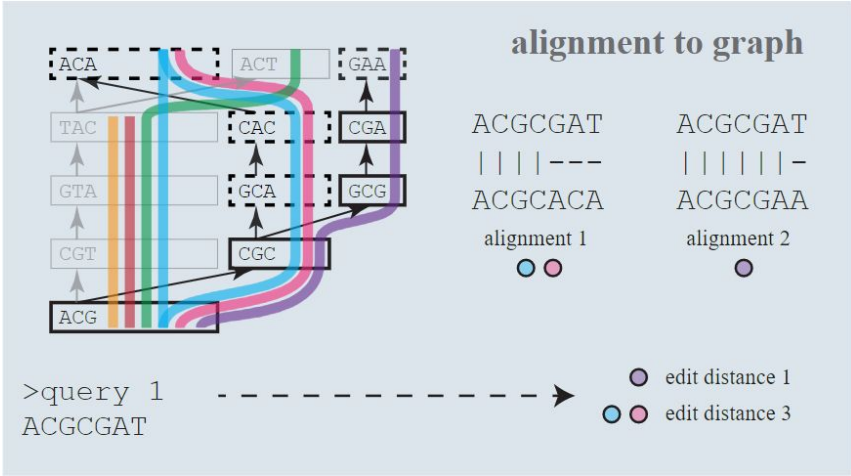
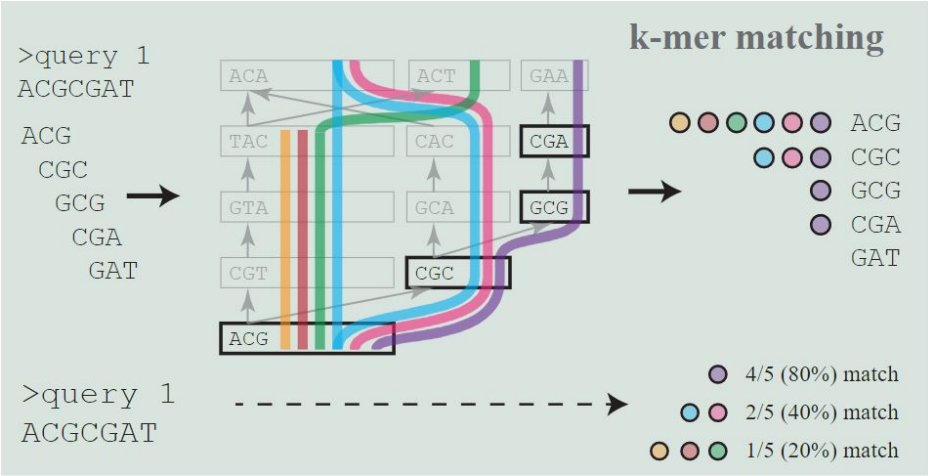
**Metagraph** is a versatile framework for the scalable analysis of extensive sequence repositories.

Can be used for:

- DNA Search
- Sequence alignment
- Differential assembly



# Background



Sequence alignment using Metagraph

# Goal

Given an annotated Counting de Bruijn Graph and a set of kmers, return the set of all input read sequences that overlap at least one of the kmers in the given set.

That will help to perform more refined graph cleaning during the assembly step and also be useful for further downstream analysis after the sequence search.

Thank you!

---