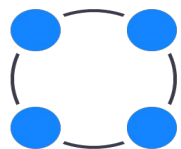


Автоматический анализ ошибок сборки SPAdes

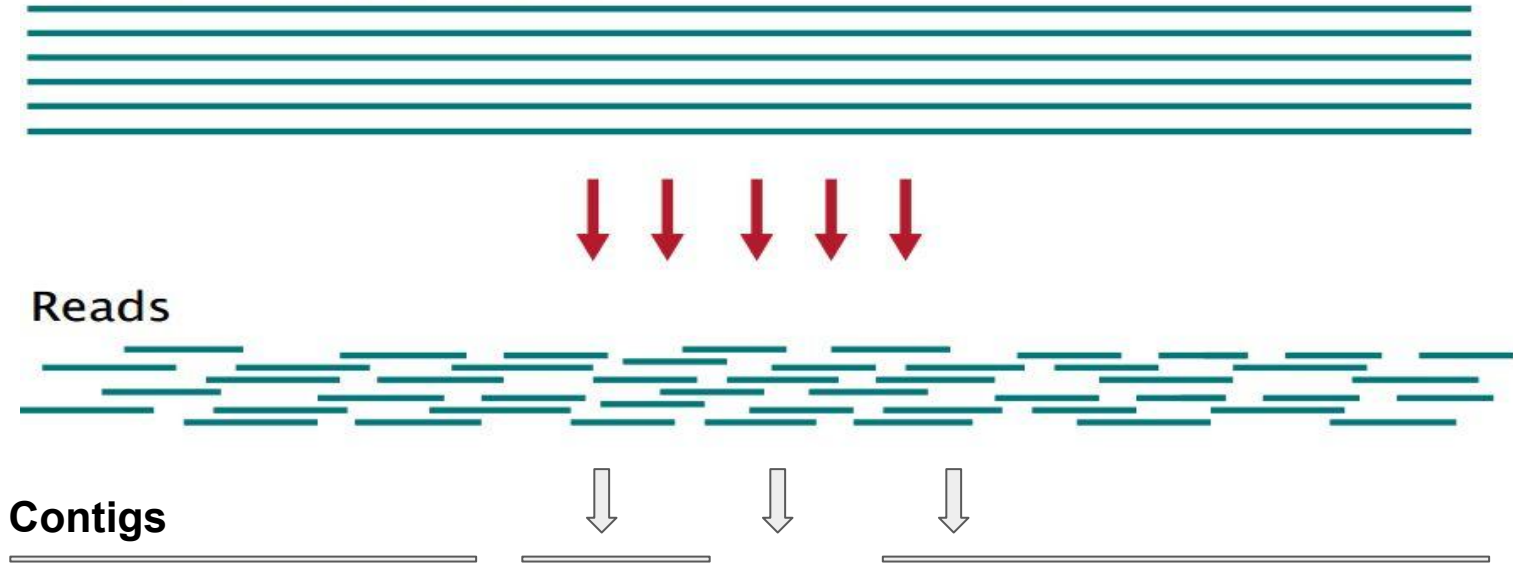


ИНСТИТУТ
БИОИНФОРМАТИКИ

Руководитель проекта: Горшков Юрий
Центр алгоритмической биотехнологии

Студент: Колесниченко Андрей.

Multiple Genome Copies

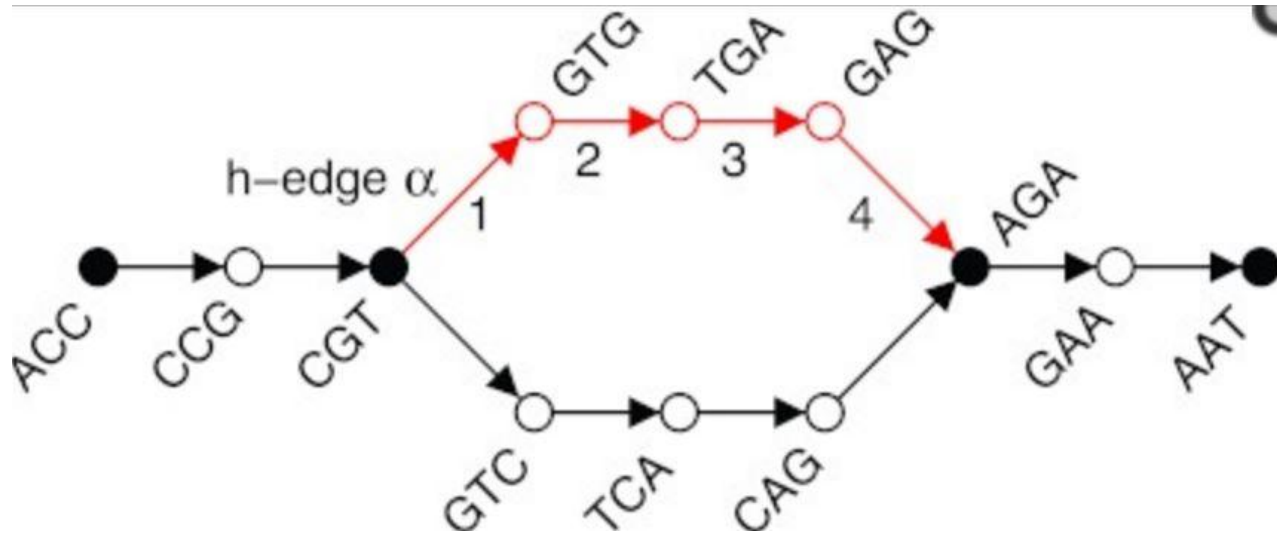


Issues:

- Dictionary contains only 4 letters
- Errors in reads
- Repeats
- Little context clues

SPAdes - genome assembler

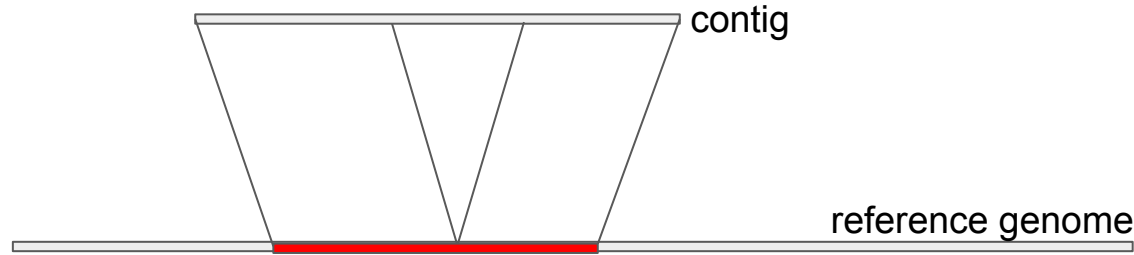
- One of the leading prokaryote assemblers.
- Uses De Bruijn graphs.



A de Bruijn graph on reads ACCGT**C**AGAAT and ACCGT**G**AGAAT

QUAST - assembly quality checker

- Can use a reference genome to check the assembly quality.
- Produces different metrics: quantitative, qualitative.
- Produces misassembly infos.

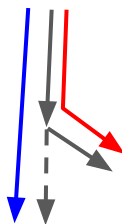
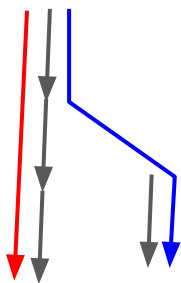


- Currently each misassembly case is investigated manually by looking at the assembly graph. That requires a lot of time.

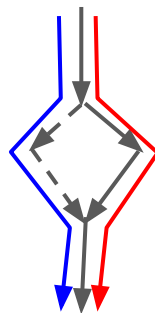
Project objectives:

- Use alignment information provided by QUAST to draw misassembled parts of the graph.
- Auto-detect misassembly types of those parts:

<ul style="list-style-type: none">• Bad or missing data	<ul style="list-style-type: none">• Wrong simplifications	<ul style="list-style-type: none">• Incorrect repeat resolving
---	---	--



Tip



Bulge

