

Graph genome theory

Jordan Eizenga, Tobias Marschall, Erik Garrison

Methods of making genome graphs

Adding variation

CHR	POS	REF	ALT
1	5	G	T
1	20	GA	G

ATCCGATACCAGATACCATGTACCAGTACAG

Methods of making genome graphs

Adding variation

CHR	POS	REF	ALT
1	5	G	T
1	20	GA	G

ATCC G ATACCAGATACCAT G TACCAGTACAG

Methods of making genome graphs

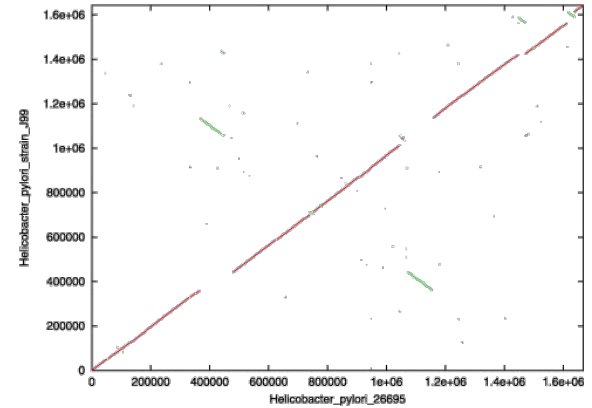
Adding variation

CHR	POS	REF	ALT
1	5	G	T
1	20	GA	G



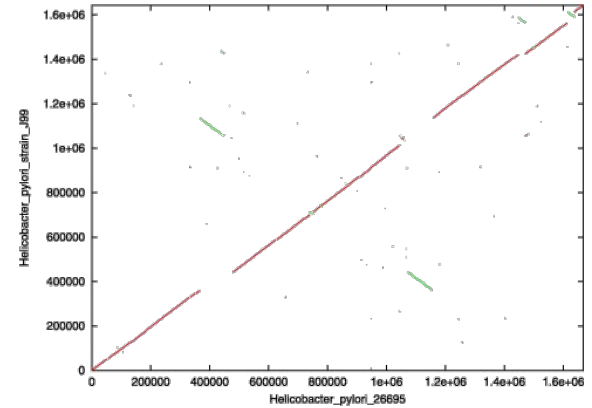
Methods of making genome graphs

Alignment



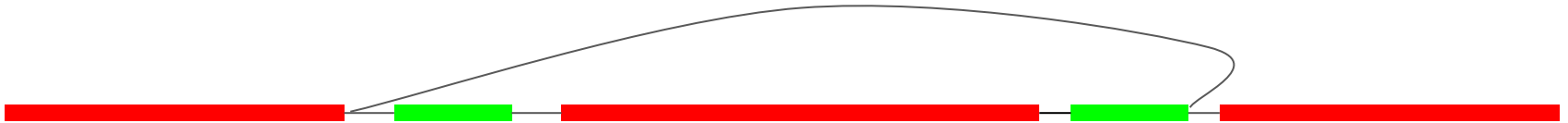
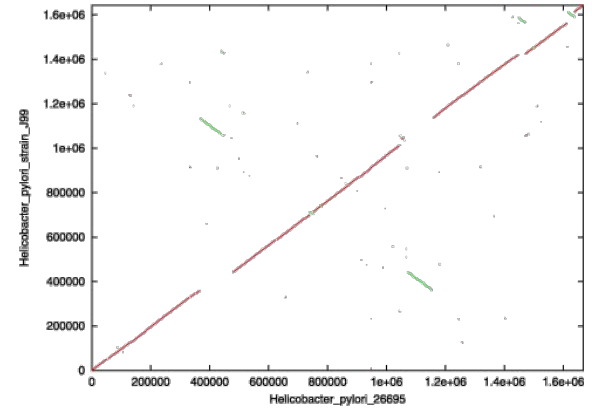
Methods of making genome graphs

Alignment



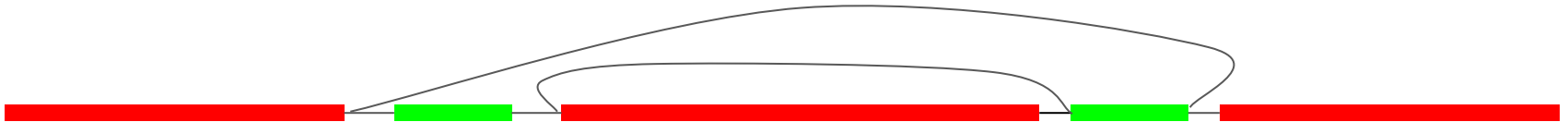
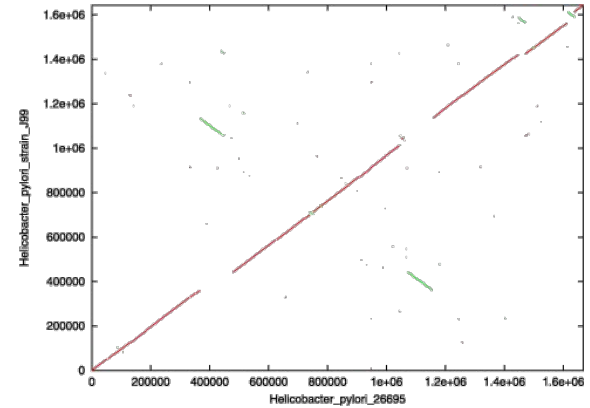
Methods of making genome graphs

Alignment



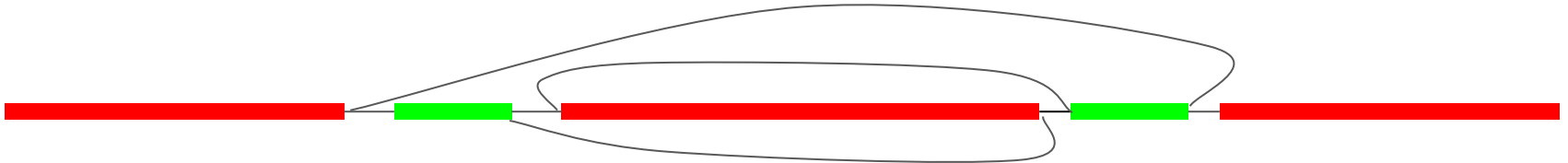
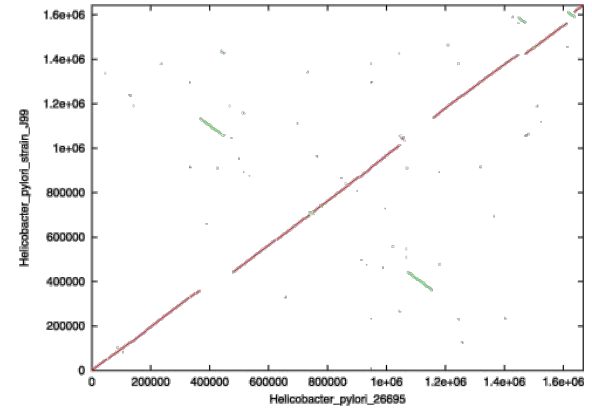
Methods of making genome graphs

Alignment



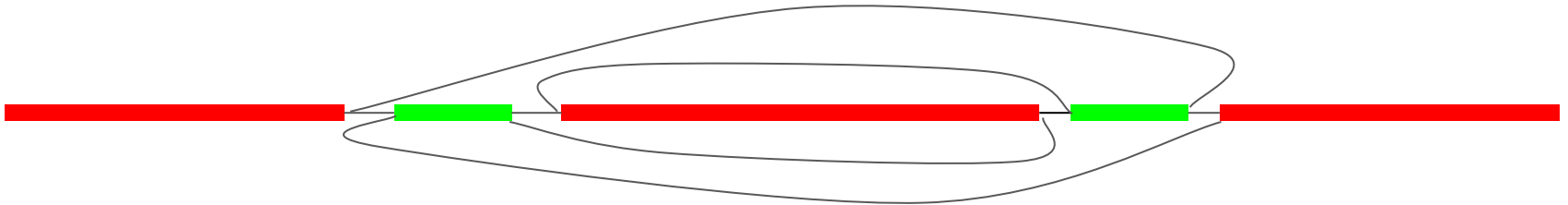
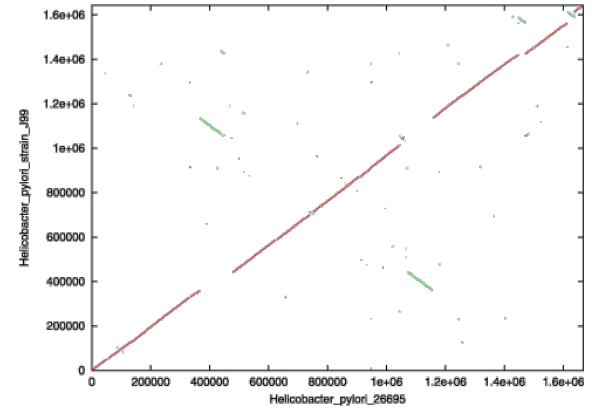
Methods of making genome graphs

Alignment



Methods of making genome graphs

Alignment

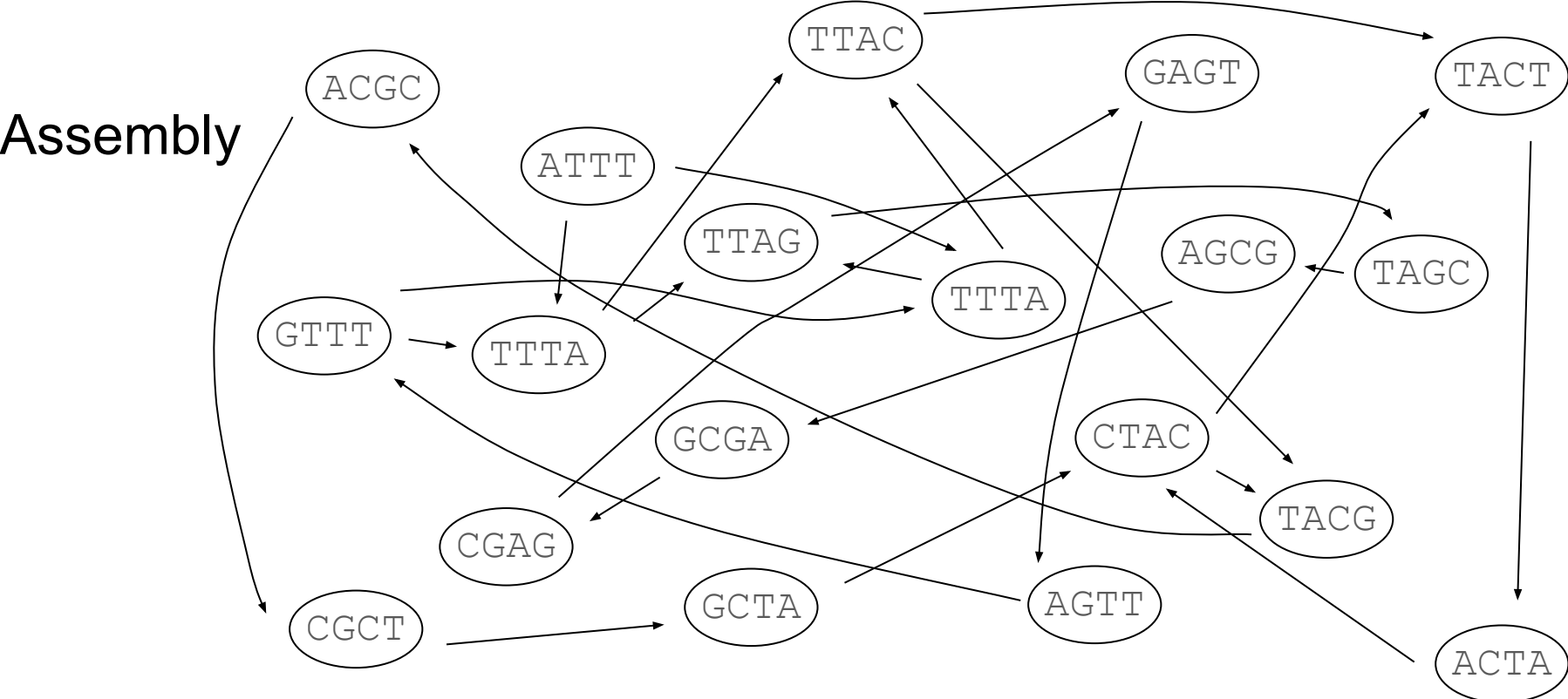


Methods of making genome graphs

Assembly

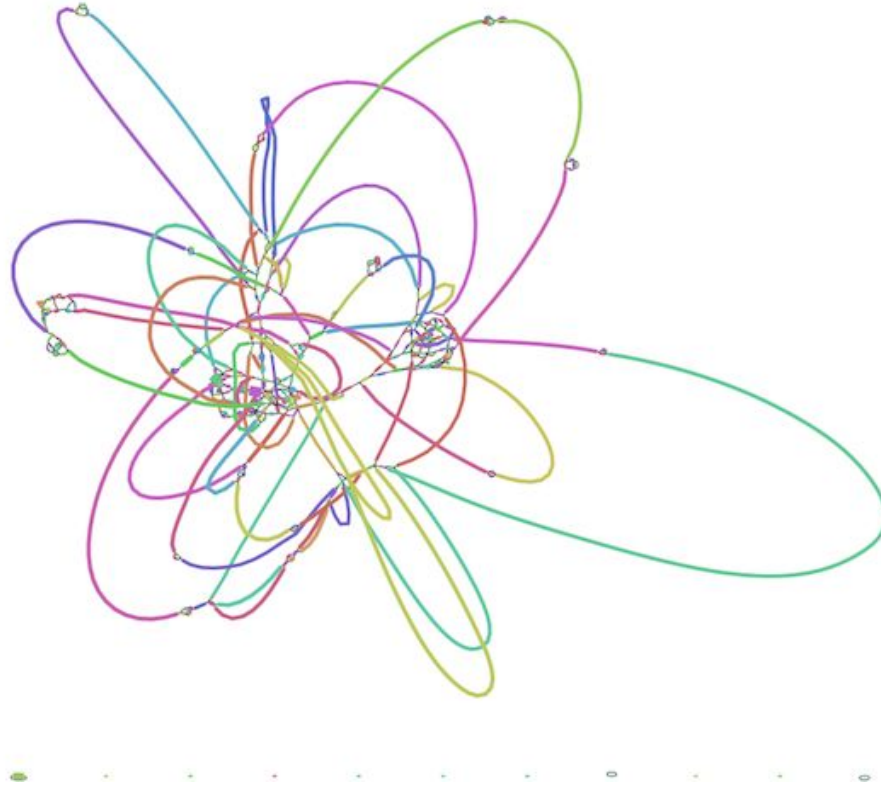
ACGC TTAC GAGT TACT
ATTT
TTAG AGCG TAGC
GTTT TTTA TTTA
GCGA CTAC
CGAG TACG
CGCT GCTA AGTT
ACTA

Methods of making genome graphs

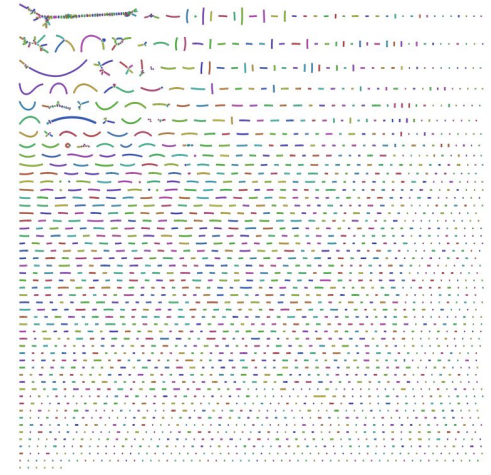
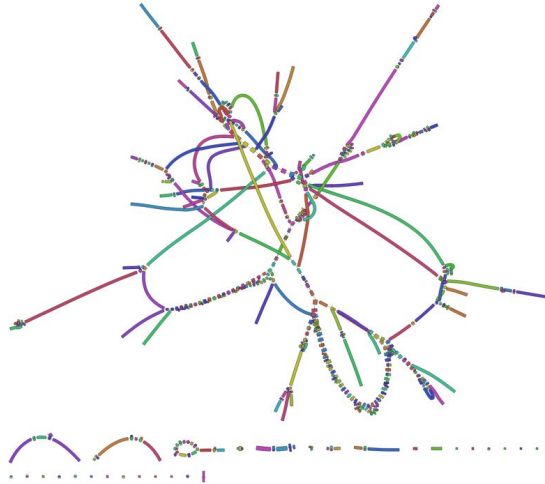
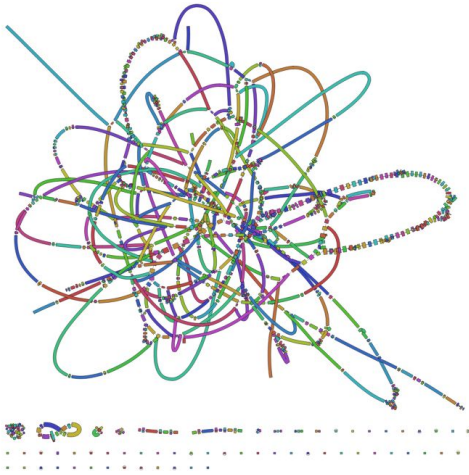


Methods of making genome graphs

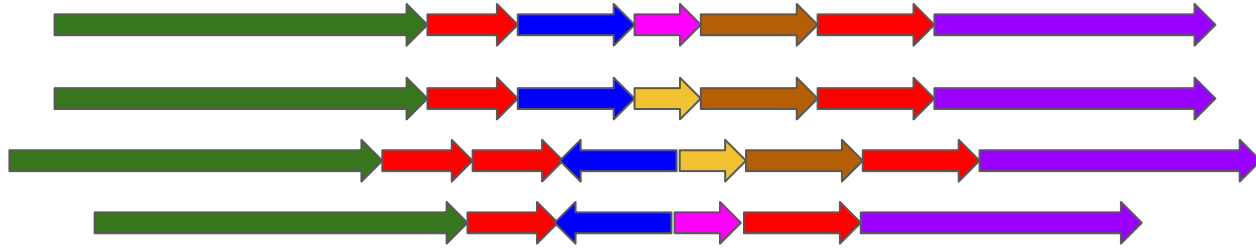
Assembly



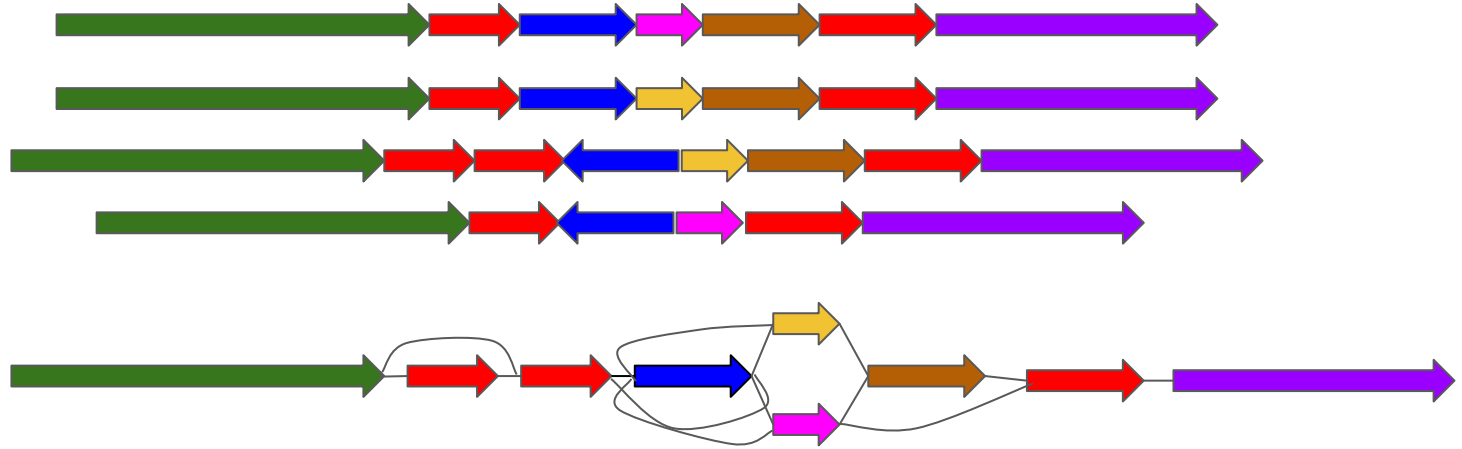
What makes a “good” genome graph?



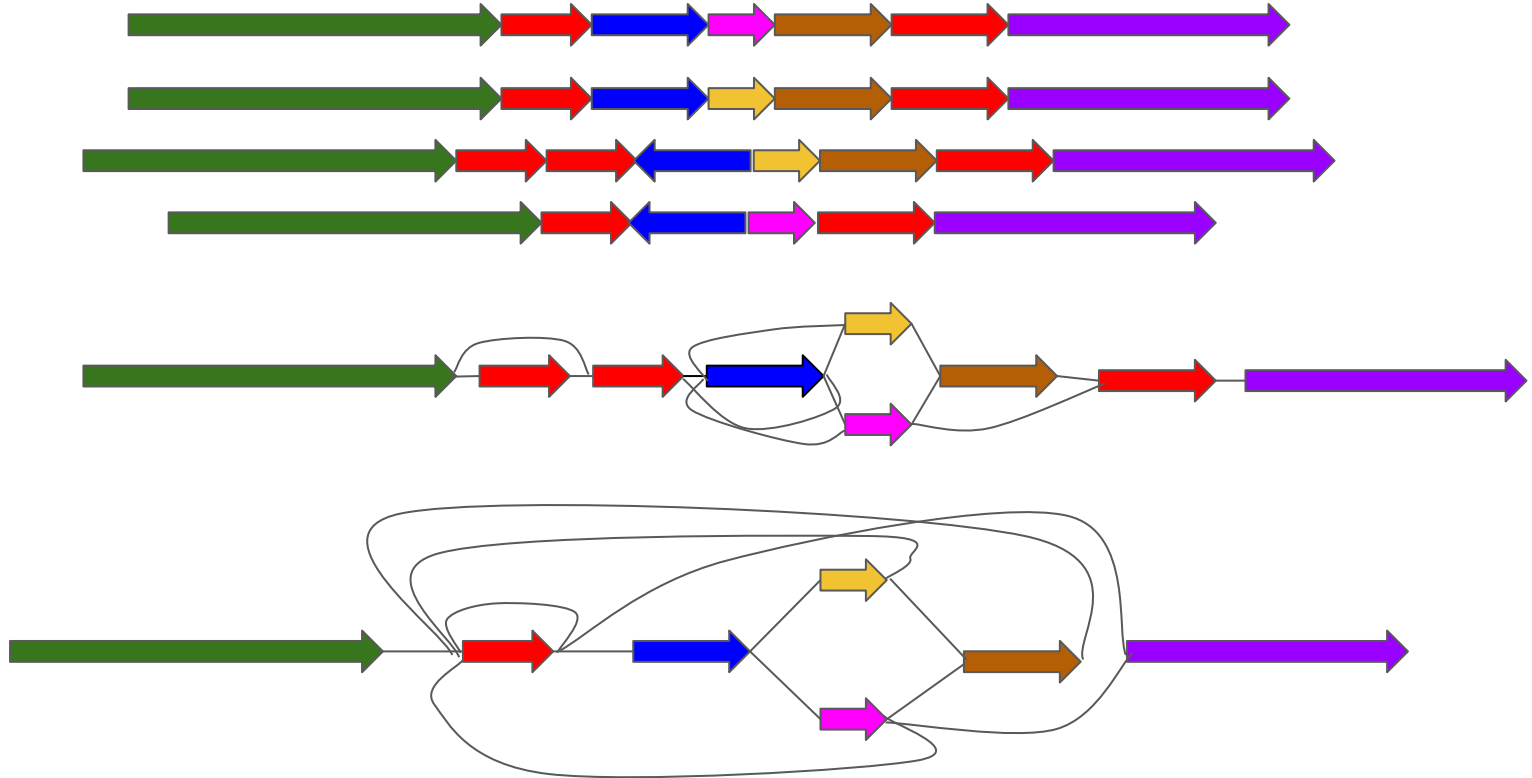
What makes a “good” genome graph?



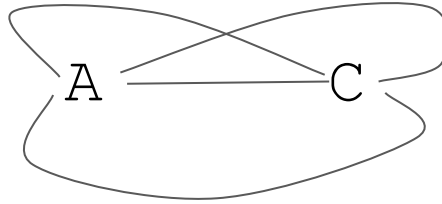
What makes a “good” genome graph?



What makes a “good” genome graph?

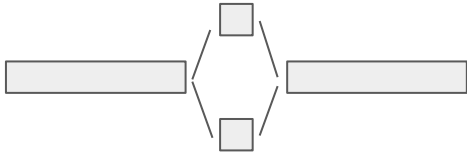


What makes a “good” genome graph?



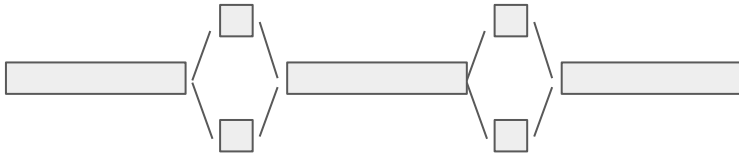
What makes a “good” genome graph?

2 paths



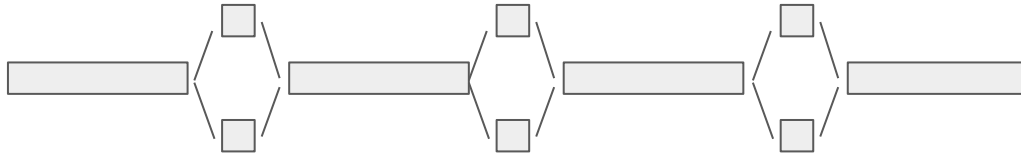
What makes a “good” genome graph?

4 paths



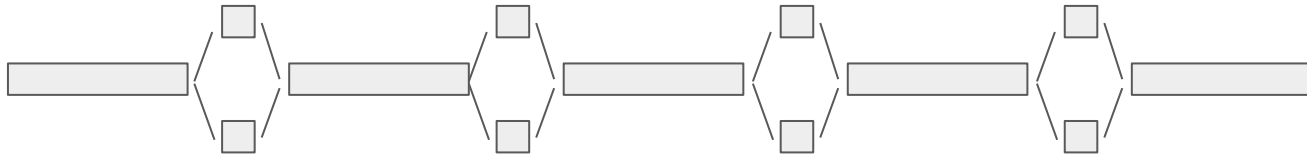
What makes a “good” genome graph?

8 paths



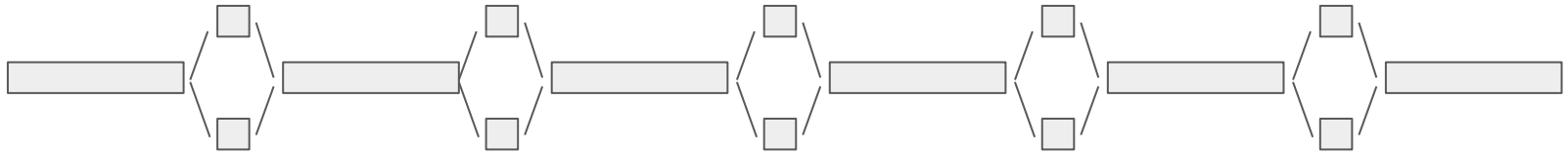
What makes a “good” genome graph?

16 paths



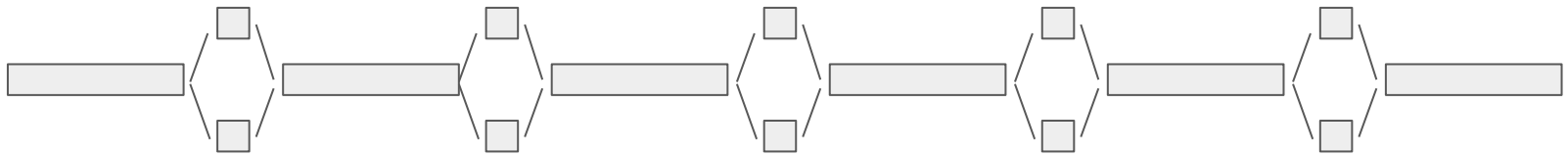
What makes a “good” genome graph?

32 paths



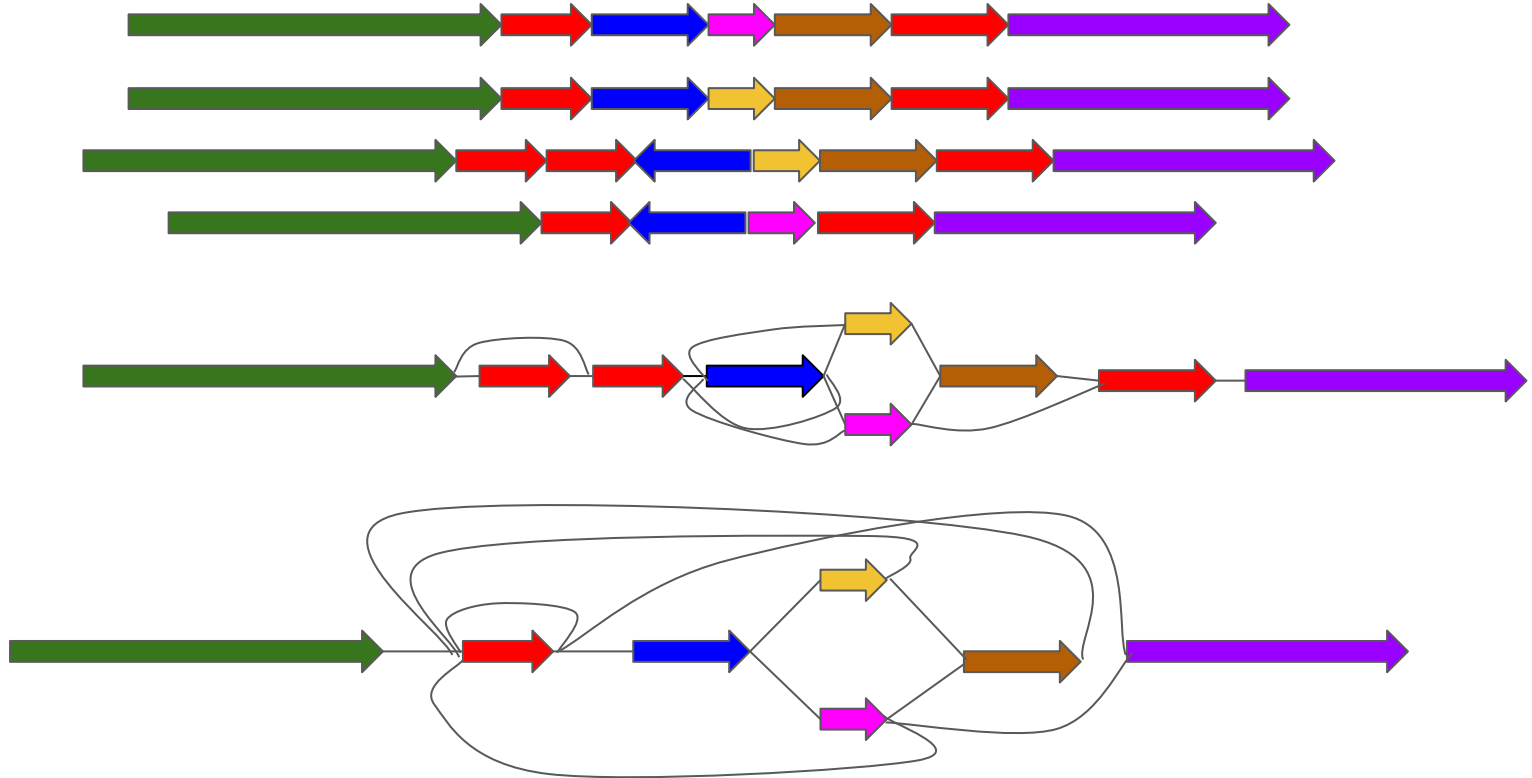
What makes a “good” genome graph?

32 paths



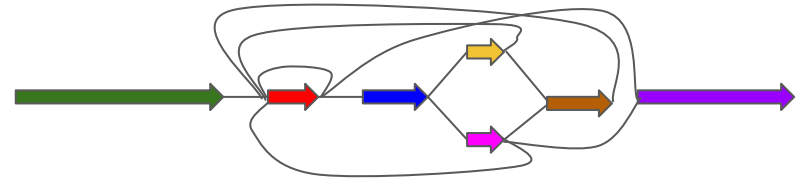
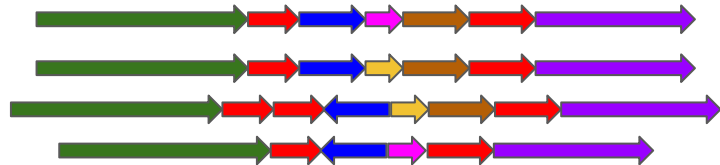
With 40 SNPs $>1,000,000,000,000$ paths

What makes a “good” genome graph?



What makes a “good” genome graph?

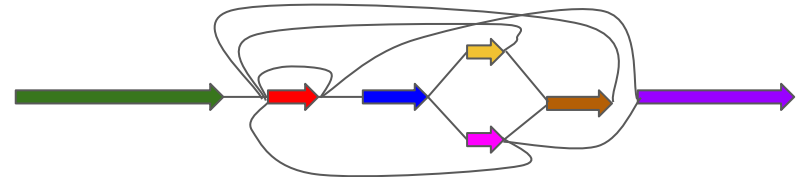
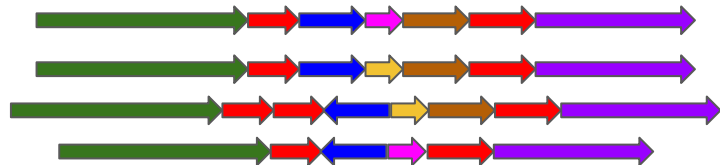
Fundamental tradeoff:



What makes a “good” genome graph?

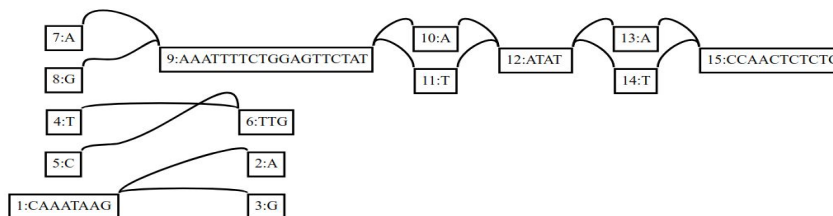
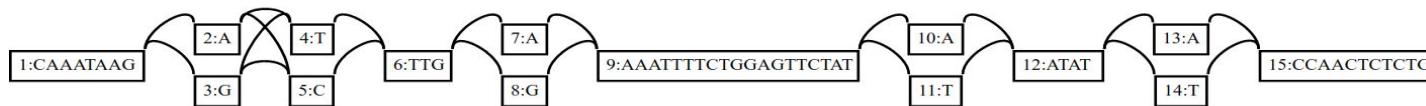
Fundamental tradeoff:

Homology ambiguity vs. Layout ambiguity



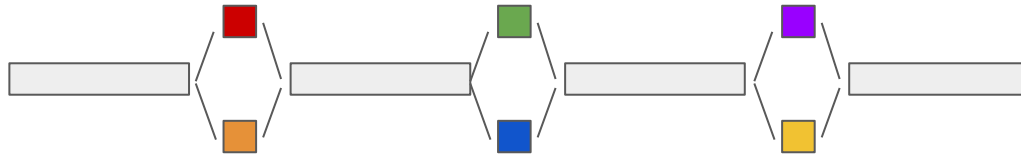
Methods for restraining complexity

Pruning



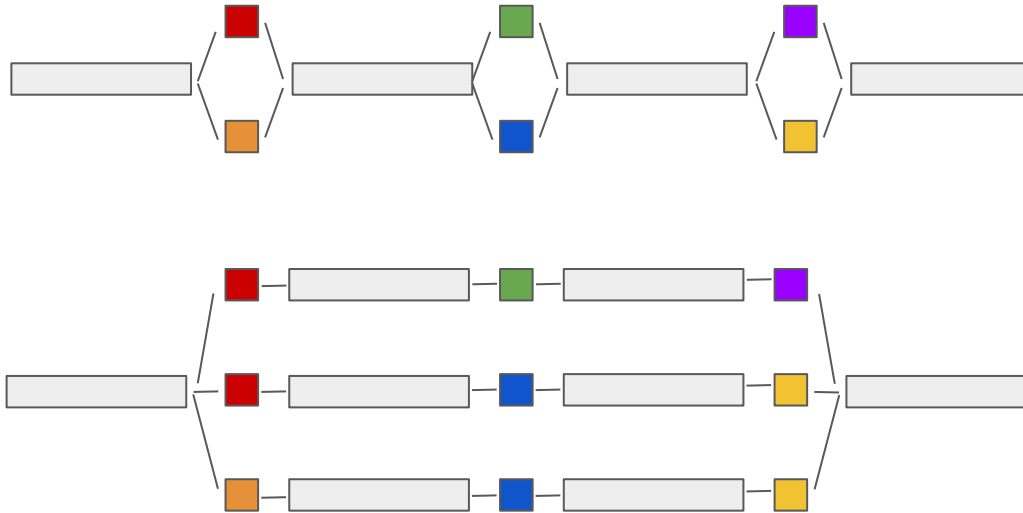
Methods for restraining complexity

Combining and duplicating



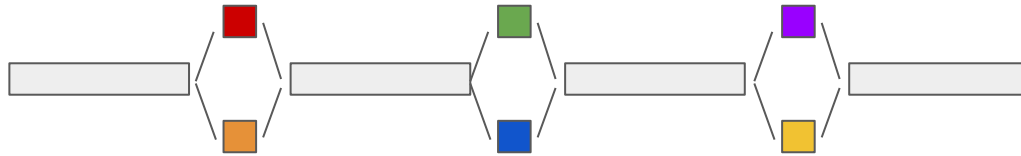
Methods for restraining complexity

Combining and duplicating



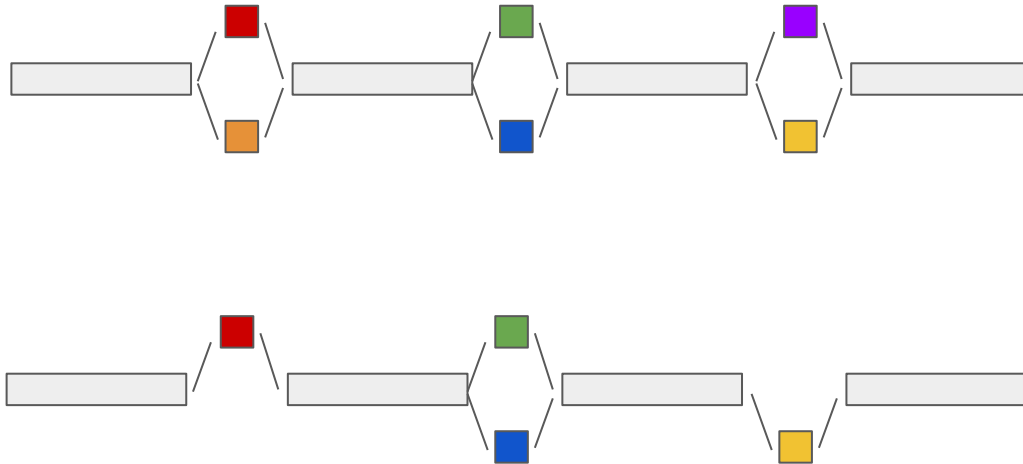
Methods for restraining complexity

Filtering



Methods for restraining complexity

Filtering



Methods for restraining complexity

Math sidebar?

Variants and loci

(go to PowerPoint)