

Loci and variants

How do we define a genetic locus?

Reference

ATGTCATCTAGGAGGGGCGCGTAGGATAAATAATTCAATTAAGATGTCGTTTTGCTAGTATACGCCTAGGCGTCACCGGCCATCTGTGTGCAGATGGG

Loci and variants

How do we define a genetic locus?

- Coordinate on a reference

Chr1:25

Reference

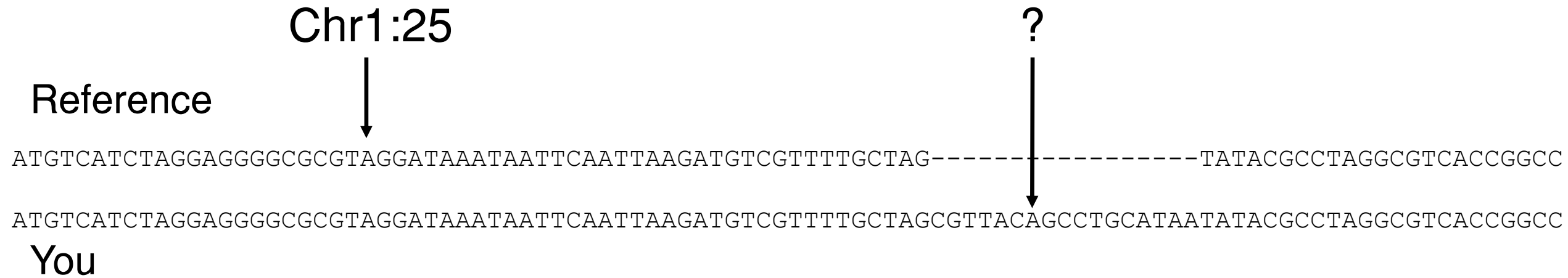


ATGTCATCTAGGAGGGGCGCGTAGGATAAATAATTCAATTAAGATGTCGTTTTGCTAGTATACGCCTAGGCGTCACCGGCCATCTGTGTGCAGATGGG

Loci and variants

How do we define a genetic locus?

- Coordinate on a reference
- But...



Loci and variants

How do we define a genetic variant?

Reference

ATGTCATCTAGGAGGGGCGCGTAGGATAAATAATTCAATTAAGATGTCGTTTTGCTAGTATACGCCTAGGCGTCACCGGCCATCTGTGTGCAGATGGG

Loci and variants

How do we define a genetic variant?

- Edit operations to the reference

CHR	POS	REF	ALT	INFO
1	24	TA	T	.

Chr1:25

Reference



ATGTCATCTAGGAGGGGCGCGTAGGATAAATAATTCAATTAAGATGTCGTTTTGCTAGTATACGCCTAGGCGTCACCGGCCATCTGTGTGCAGATGGG

ATGTCATCTAGGAGGGGCGCGT-GGATAAATAATTCAATTAAGATGTCGTTTTGCTAGTATACGCCTAGGCGTCACCGGCCATCTGTGTGCAGATGGG

You

Loci and variants

How do we define a genetic variant?

- Edit operations to the reference
- But...

Chr1:25

Reference



ATGTCATCTAGGAGGGGCGCGTAGGATAAATAATTCAATTAAGATGTCGTTTTGCTAGTATACGCCTAGGCGTCACCGGCCATCTGTGTGCAGATGGG

ATGTCATCTAGGAGGGGCGCGT-GGATAAATAATTCAATTAAGATGTCGTTTTGCTAGTATACGCCTAGGCGTCACCGGCCATCTGTGTGCAGATGGG

You

CHR	POS	REF	ALT	INFO
1	24	TA	T	.
1	24	T		SVTYPE=DEL;END=26
1	24	T	T[1:25]	SVTYPE=BND
1	26	G	[1:25]G	SVTYPE=BND

Loci and variants

Which variants can co-occur?

Reference

ATGTCATCTAGGAGGGGCGCGTAGGATAAATAATTCAATTAAGATGTCGTTTTGCTAGTATACGCCTAGGCGTCACCGGCCATCTGTGTGCAGATGGG

Loci and variants

Which variants can co-occur?

- Only two variants at the same locus

Reference

ATGTCATCTAGGAGGGGCGCGTAGGATAAATAATTCAATTAAGATGTCGTTTTGCTAGTATACGCCTAGGCGTCACCGGCCATCTGTGTGCAGATGGG

Loci and variants

Which variants can co-occur?

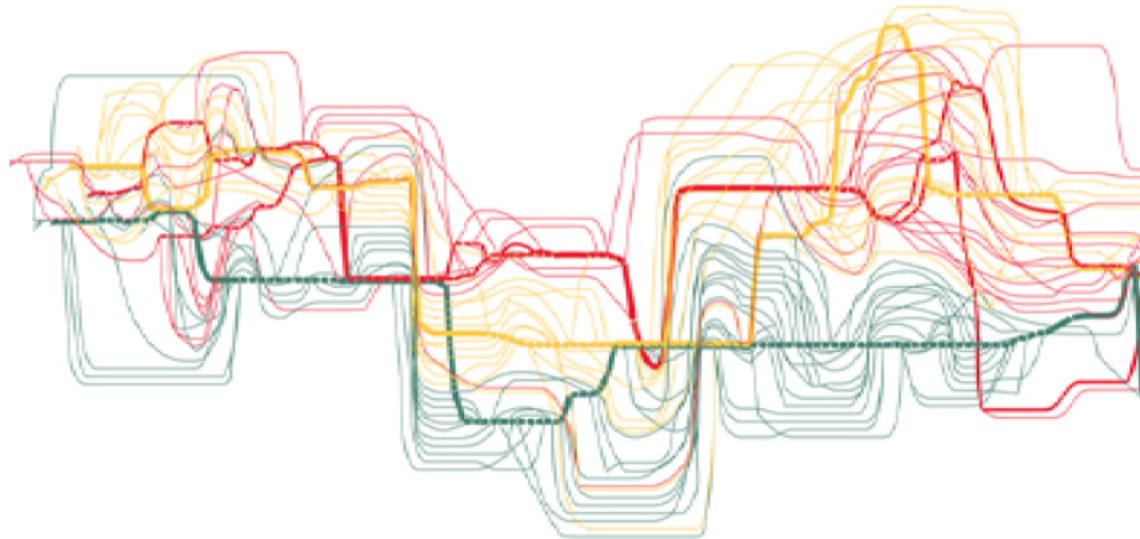
- Only two variants at the same locus
- But... violations ~1x / megabase in 1000 Genomes call set

Reference

ATGTCATCTAGGAGGGGCGCGTAGGATAAATAATTCAATTAAGATGTCGTTTTGCTAGTATACGCCTAGGCGTCACCGGCCATCTGTGTGCAGATGGG

Loci and variants

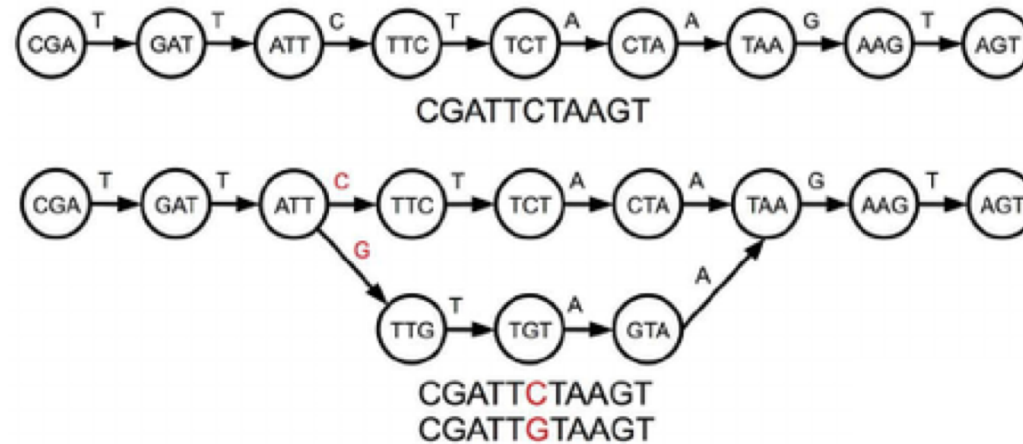
Where are the loci in a genome graph?



Loci and variants

Where are the loci in a genome graph?

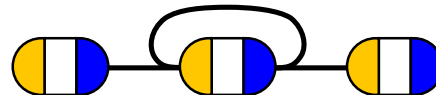
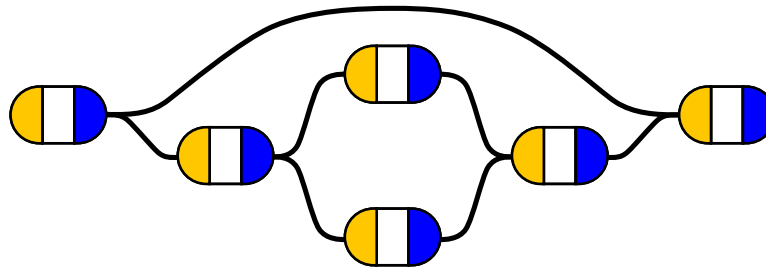
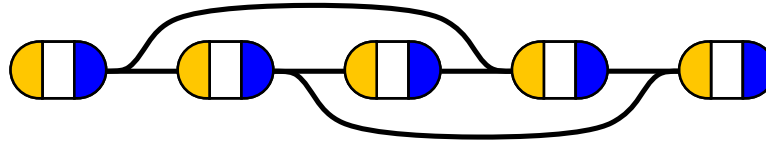
Bubbles



Loci and variants

Where are the loci in a genome graph?

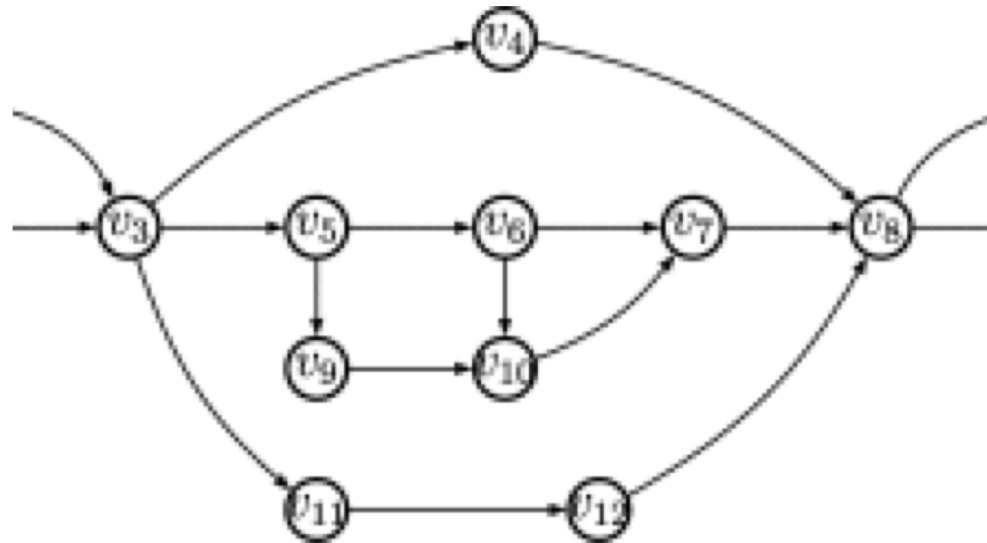
- But...



Loci and variants

Where are the loci in a genome graph?

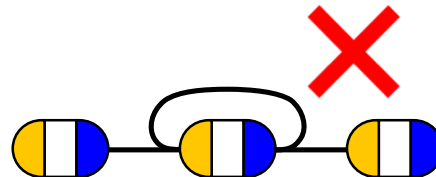
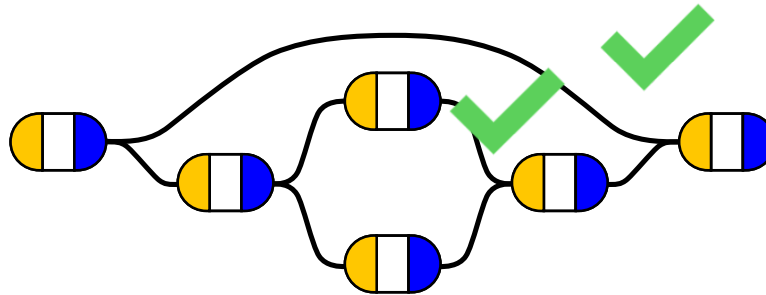
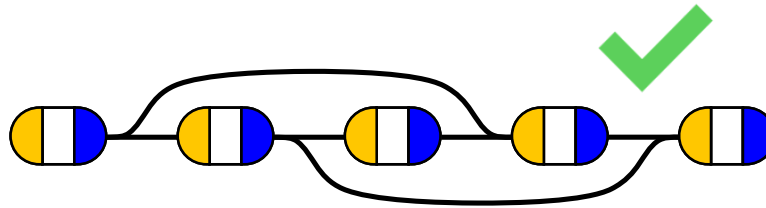
Superbubbles



Loci and variants

Where are the loci in a genome graph?

- But...

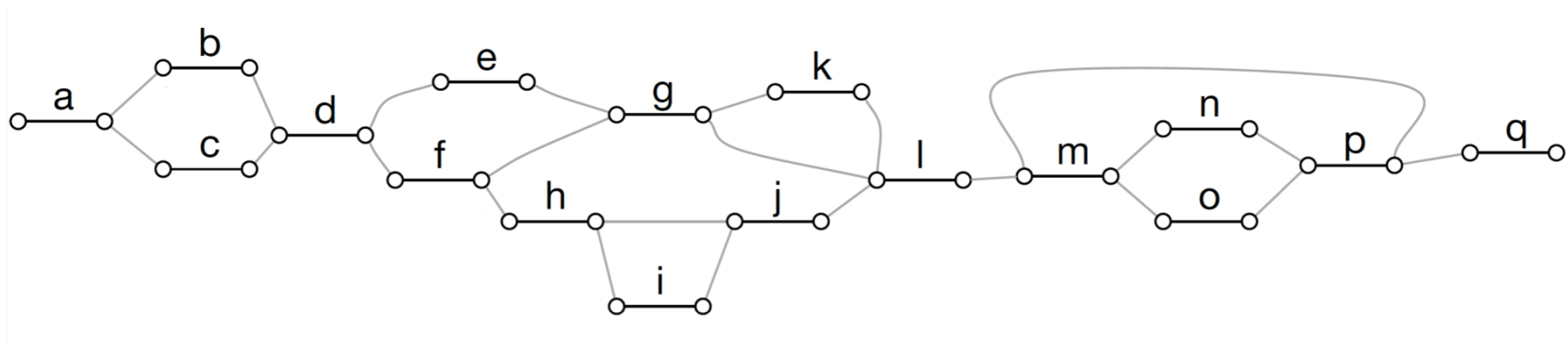


Loci and variants

Where are the loci in a genome graph?

Snarls

- Defined only in terms of being separable subgraphs

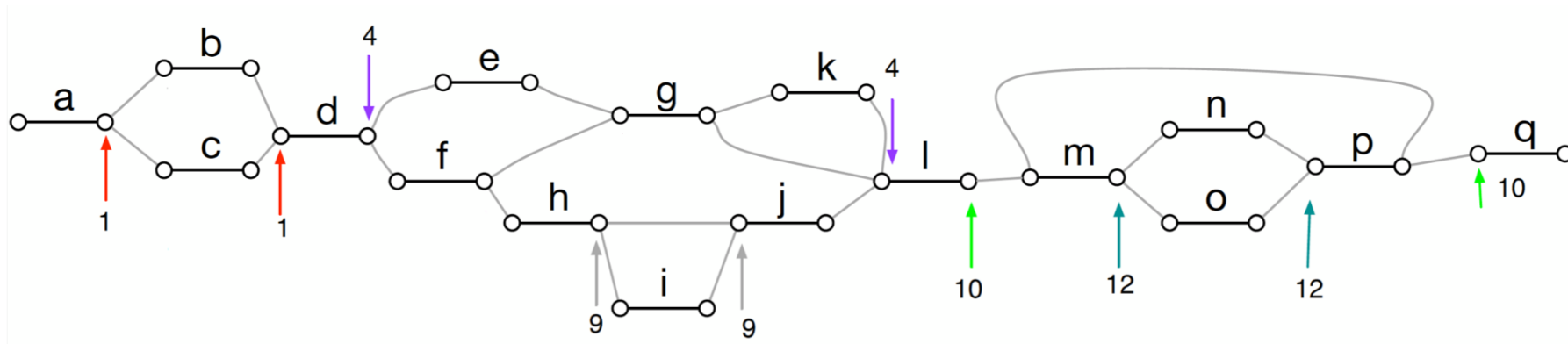


Loci and variants

Where are the loci in a genome graph?

Snarls

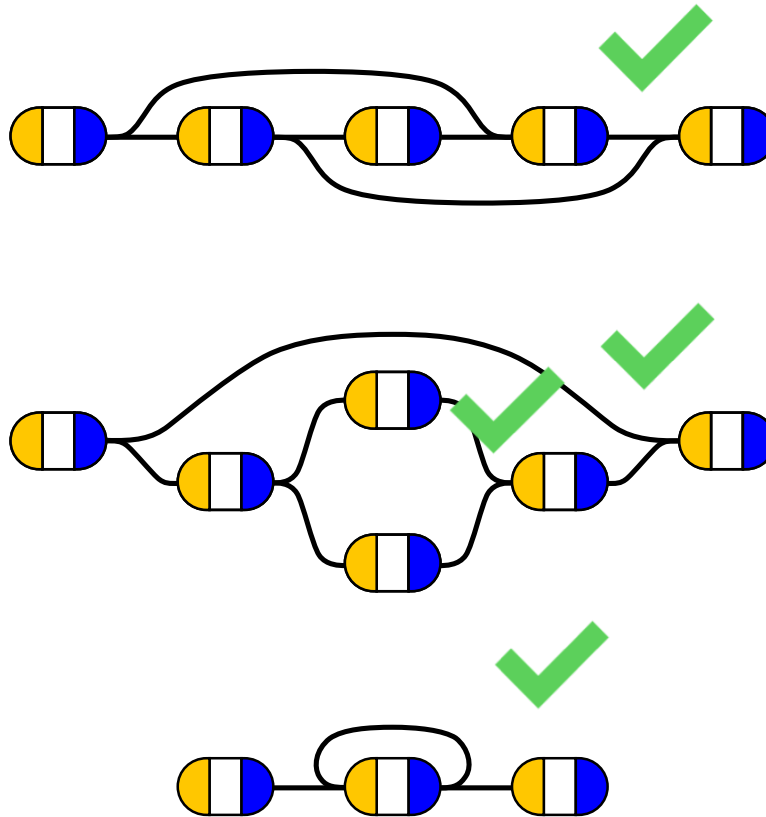
- Defined only in terms of being separable subgraphs



Loci and variants

Where are the loci in a genome graph?

- Yay!



Loci and variants

Initial results on chromosome 1

Structure	Nesting Level	Count	Coverage (bp)	Coverage (pct)
<i>Chains</i>	top	1	221,715,143	86.60
<i>Ultrabubbles</i>	top	5,554,903	12,539,619	4.90
<i>Snarls</i>	top	75	21,775,387	8.50
<i>Chains</i>	second	919	20,594,450	8.04
<i>Ultrabubbles</i>	second	533,252	1,199,777	0.47
<i>Snarls</i>	second	0	0	0
<i>Chains</i>	third	67	495	0.00
<i>Ultrabubbles</i>	third	694	1,623	0.00
<i>Snarls</i>	third	0	0	0

Loci and variants

In VG:

```
# compute the snarls and save them in a file  
> vg snarls -t graph.vg > graph.vg.snarls
```

```
# view the snarls in JSON format  
> vg view -Rj graph.vg.snarls
```

Loci and variants

In VG:

```
# the vg genotype command adds read edits into the graph (so that  
# they form snarls) and then genotypes the snarls whose ends fall  
# on a reference path  
> vg genotype -G mapped.gam -r refpathname -v graph.vg > genotypes.vcf
```