Heterozygous Genome Display
Outline

- Background: Biology
- Alignment Approach
- Synteny Approach
Background: DNA

```
Background: DNA
```

```
RNA  →  Protein  →  Phenotype
```

```
Background: DNA
```

```
RNA  →  Protein  →  Phenotype
```
Background: Definitions

- **Diploid**: having two copies of DNA
- **Homozygous**: both copies are the same
- **Heterozygous**: the two copies are different
- **Polyploidy**: having multiple copies of DNA (commonly 2, 4, or 6).
- **Variant**: A change in the DNA
- **Haplotype**: A set of variants that are in the same copy of DNA.
Background: Variants

- **SNPs**: Single nucleotide change (ACT -> ATT)
- **Insertions/Deletions**: Changes that add or remove bases (ATT -> AT)
  - Variable Tandem Repeats: change in the number of repeats (ACTACT -> ACTACTACT)
- **Rearrangements**: Changes that move sections of DNA to a new location (TAAT -> TTAA)
- **Inversions**: Changes that switch the direction of the DNA (ACTT -> TTCA)
Background: Assembly

ACTGGGCAACAAGCTGCGCGCGCGTTTACAGT
ACTGGGAACAAGCTGCGCGCGCGTTTACAGT

CGCGTTTT

ACTGGGC

GTTCACAGT

Sequencing

Gaacaagctgcccgc

ACTGGGC

TGGGAAC

Assembly

TGGGAAC

GTTCACAGT

CGCGTTTT

ACTGGGAACAAGCTGCGCGCGCGTTTACAGT

C
Heterozygous Genome Display

• What if our sequence is very heterozygous. As an extreme, what if it is so heterozygous that we can’t form a consensus sequence? (Raspberry)

We need a way to display heterozygous genomes.
Current Options

Alignment Approach: Highlights Differences

Synteny Approach: Highlights Similarities

Haplowser

Gbrowse

HuRef Browser

Synteny Browser
Synteny Browser

Pros:
- Displays rearrangements.
- Can display haplotype specific genes.
- Better overview

Cons:
- Requires perfect haplotypes.
- Large homozygosity results in mostly duplicate tracks.
- Polyploidy quickly becomes overwhelming.
Alignment View

Pros:
- Limits duplicate tracks
- Better for polyploidy
- Handles imperfect assemblies.

Cons:
- No rearrangements or inversions could be shown.
- Haplotype specific genes would be hard to display
- Generates a fake consensus sequence
Questions?