Assembly of heterozygous Fodder beet

Philipp Peters

University of Natural Resources and Life Sciences, Vienna

31st TBI Winterseminar
February 18th, 2016

Sugar beet (Beta vulgaris)

- Diploid with n = 9 chromosomes
- Estimated genome size: 731 Mbp
- Reference assembly[1] was created with a double-haploid plant

[1] Dohm et al., The genome of the recently domesticated crop plant sugar beet (Beta vulgaris), Nature, 2014

Assembly of heterozygous Fodder beet

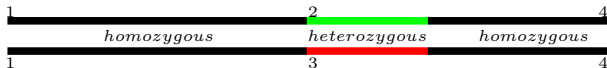**Fodder beet**
- Diploid
- Estimated genome size $\approx$ sugar beet
- Assembly size: 955 Mbp

  Assembled by J. Dohm, using Roche's Newbler

  - Highly heterozygous
    (Different alleles at one or more loci on homologous chromosomes)
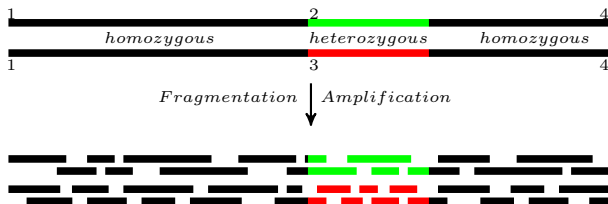- Task: Resolve the heterozygous structure of the genome
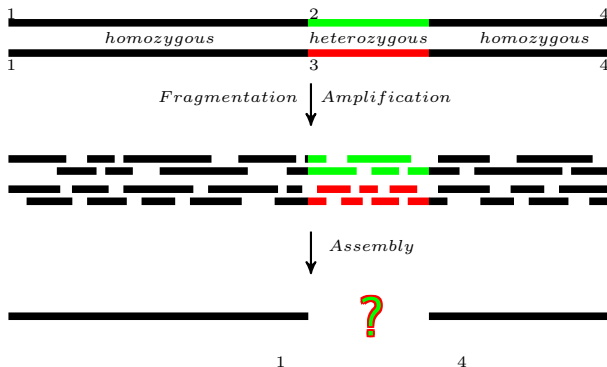
Assembly of heterozygous Fodder beet

Diploid Genome

Assembly of heterozygous Fodder beet
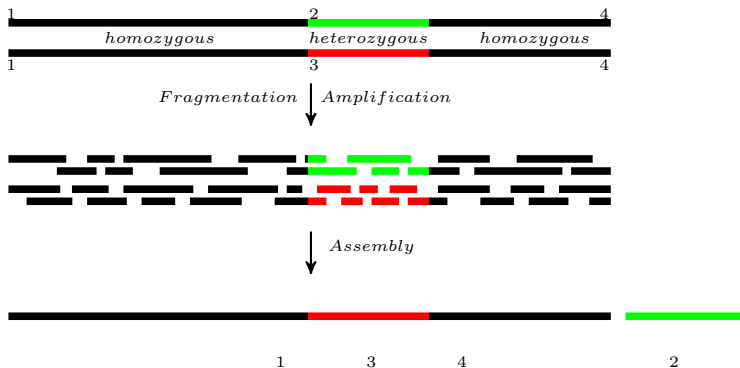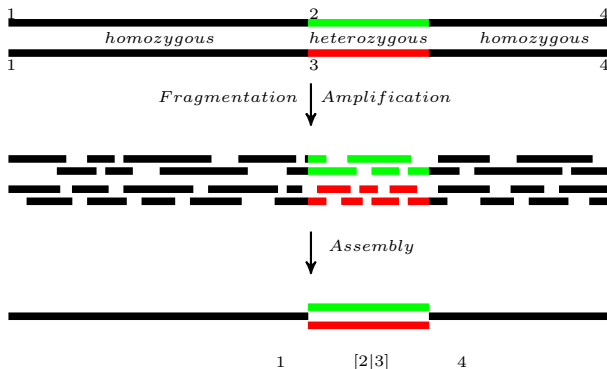
Diploid Genome

Assembly of heterozygous Fodder beet

Diploid Genome

Diploid Genome
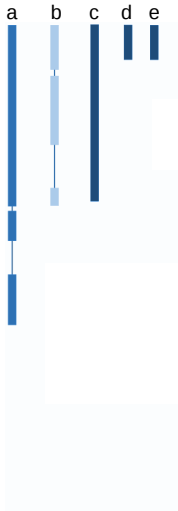
Assembly of heterozygous Fodder beet

- Raw Data
  - Moleculo reads (8-9x Coverage)
  - Illumina Mate-Pairs

- Assembled Data (Newbler)
  - Assembly (955 Mbp)
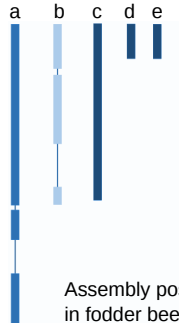 - Additional information on
    Contig connections and coverage

Assembly of heterozygous Fodder beet

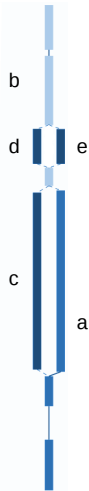assembled output, 59 kbp:

Assembly of heterozygous Fodder beet

assembled output, 59 kbp:

a b c d e

heterozygous structure:

b

d e

c

a

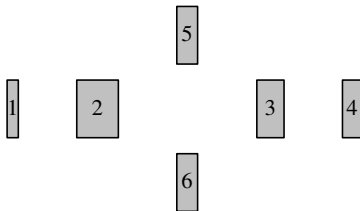Assembly postprocessing in fodder beet to resolve the heterozygous genome structure.

total length: 38 kbp

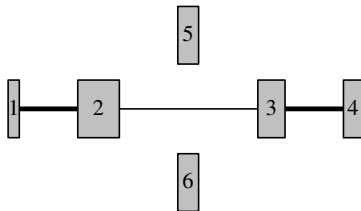# Assembly of heterozygous Fodder beet

Create Assembly Graph

▶ Contigs as vertex set

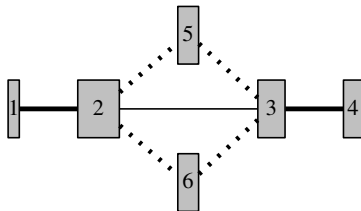# Assembly of heterozygous Fodder beet

Create Assembly Graph

- Contigs as vertex set
- Scaffold-connections of Newbler-assembly as edge set (solid)

Assembly of heterozygous Fodder beet

Create Assembly Graph

- Contigs as vertex set
- Scaffold-connections of Newbler-assembly as edge set (solid)
- Additional Links from Newbler-output as edge set (dotted)

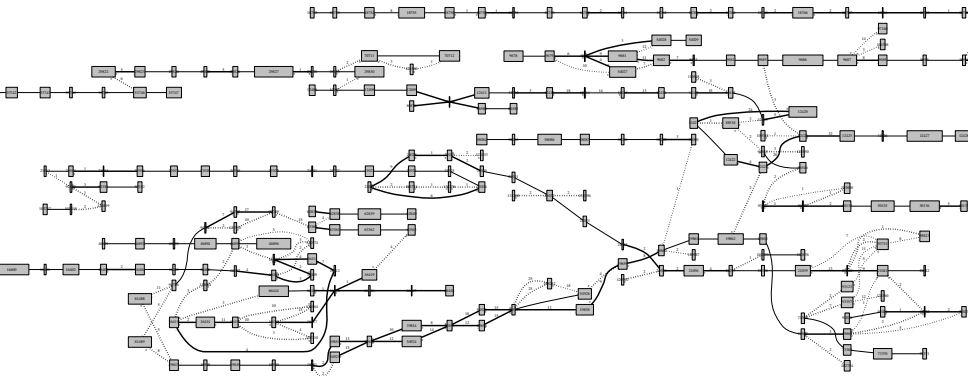# Assembly of heterozygous Fodder beet

Create Assembly Graph

- Contigs as vertex set
- Scaffold-connections of Newbler-assembly as edge set (solid)
- Additional Links from Newbler-output as edge set (dotted)

Assembly of heterozygous Fodder beet

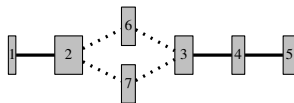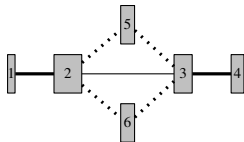**Cycles**
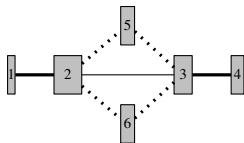
Assembly of heterozygous Fodder beet

Cycles

- Find all cycles in graph $\rightarrow$ NP-complete Problem
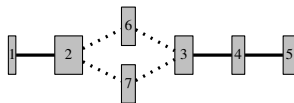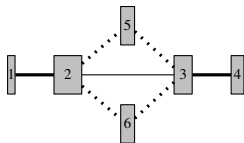
Cycles

- Find all cycles in graph $\rightarrow$ NP-complete Problem
- Which cycles are the good/correct cycles?

Assembly of heterozygous Fodder beet

Splits
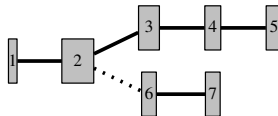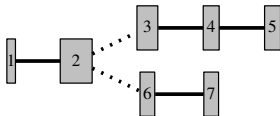
- Which splits are the good/correct splits?

Assembly of heterozygous Fodder beet

Filtering Criteria

- 3′- and 5′-ends consistent
- Distances consistent / realistic

- Structural criteria
- Coverage
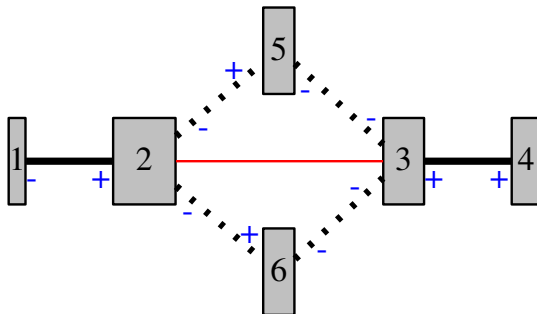
Assembly of heterozygous Fodder beet

Cycle Validation

- Mapping of Fodder beet Contigs against RefBeet 1.2
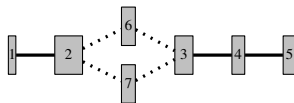- Two Contigs on same position in RefBeet
    - ⇒ **Possible heterozygous pair**
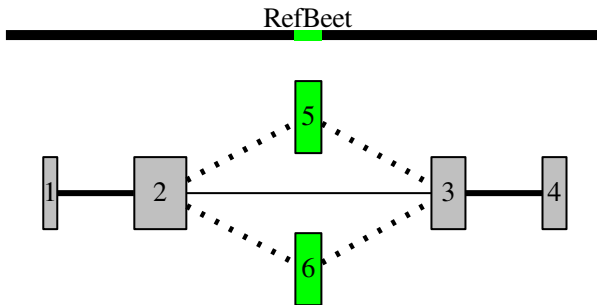
Assembly of heterozygous Fodder beet

Cycle Detection

	Scaffolds Only	
♯ **Contigs**	106732	
♯ **Cycles**	3429	
♯ **Filtered Cycles**	1733	
♯ **Validated Cycles**	1511 87.2 %	
Length Component	676.5 Mbp	
Length Filt. Cycles	34.8 Mbp 5.1 %	
Length Hetero	2×9 Mbp 2×1.33 %	
Runtime	8.4 h 1.4 GHz 4-core	

Assembly of heterozygous Fodder beet

Cycle Detection

	Scaffolds Only		Scaffolds & Contigs	
♯ **Contigs**	106732		163961	
♯ **Cycles**	3429		19452	
♯ **Filtered Cycles**	1733		3976	
♯ **Validated Cycles**	1511	87.2 %	3425	86.1 %
Length Component	676.5 Mbp		761.1 Mbp	
Length Filt. Cycles	34.8 Mbp	5.1 %	64.9 Mbp	8.5 %
Length Hetero	2×9 Mbp	2×1.33 %	2×14.2 Mbp	2×1.87 %
Runtime	8.4 h 1.4 GHz 4-core		10.8 h 1.4 GHz 4-core	

Assembly of heterozygous Fodder beet

Scaffold Reduction

	Scaffolds Only	
♯ **Scaffolds**	55093	
♯ **Scaffolds Before** (consisting of $\geq$ 4 Contigs)	27196	
♯ **Scaffolds After**	24882 (-8.5 %)	
♯ **Connected Scaffolds** (consisting of > 1 Contig)	816	

Assembly of heterozygous Fodder beet

Scaffold Reduction

	Scaffolds Only	Scaffolds & Contigs
♯ **Scaffolds**	55093	55093
♯ **Scaffolds Before** (consisting of ≥ 4 Contigs)	27196	32811
♯ **Scaffolds After**	24882 (-8.5 %)	29408 (-10.4 %)
♯ **Connected Scaffolds** (consisting of > 1 Contig)	816	1038

Assembly of heterozygous Fodder beet
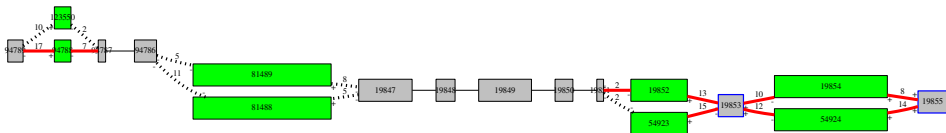
Outlook

- ▶ Find more/better criteria for Cycles and Splits
- ▶ Add more layers of information (coverage, neighborhood,...)
- ▶ Add Phasing
- ▶ Improve Validation
- ▶ Use information/output from other Scaffolders
- ▶ Data Set Simulation for Testing
- ▶ Make it usable (not only for me...)

Acknowledgements

Bioinformatics BOKU Vienna
- ▶ Heinz Himmelbauer
- ▶ Juliane Dohm

Funding and Data Access
- ▶ illumına˚
- ▶ BMBF project AnnoBeet