Possibilities and challenges of the potato genome sequence

Richard GF Visser, Wageningen UR Plant Breeding

EAPR 2014, July 6-11, Brussels

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Overview

- A sequence is available what can we do with it?
- Sequencing broad or deep?
- Genotyping by Sequencing in Potato
    - Genome re-sequencing in Potato
    - Haplotype reconstruction in polyploids
- Sequencing and breeding
    - BreeDB in Potato


OUTLOOK
Alzheimer's disease
nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
The DNA sequence of the South American tuber eaten around the world PAGE 189
THE POTATO GENOME
HISTORY
PURE JOY
Arcane mathematics that changed the world
PAGE 166
EVOLUTION
GIANT DINOSAURS
Seeds of greatness in small sauropod ancestors
PAGE 159
NEUROSCIENCE
SPINAL CORD REGENERATION
Restoring breath control after neck injury
PAGES 170 & 196
NATURE.COM/NATURE
14 July 2011 £10
Vol. 475, No. 7355

Two potatoes have been sequenced

- DM doubled monoploid has been sequenced
- Heterozygous diploid RH has been partly sequenced
- At best three alleles per gene known for a selection of traits
- Is this a problem?
  - No: Data sufficient to design SNP markers
  - Yes: No info on maximum number of alleles per gene

Deliverables and direct use of sequence

- SNPs; useful for marker development on arrays
- Dense genetic map construction in every type of material (not always same SNPs!)
- Improved QTL mapping
- Enabling link of genetic to physical map
- Potential ID of candidate genes
- Forward and backward searches for gene/trait ID possible

Potato SNP array

- Infinium array ~18,000 SNPs
  - Selected from 129 K genomic DNA sequence variants over 83 cultivars
  - Stringent quality criteria
  - Low redundancy – one-per-gene
  - Including low-frequency SNP alleles
- Mapping populations not involved in choice of SNPs
  - Expecting: enough segregating SNPs for mapping

Samples hybridized with this array

- ~2500 potato samples:
  - Tetraploid mapping population: ~250 individuals
  - Parents 2, x 2 replicates
  - Grandparents (3), great-grandparent (1)
  - >500 diploid accessions
  - ~550 tetraploid cultivars and breeding lines (GWAS)
  - 1100 other tetraploid progenies

Sequences

- More spurious data of different potato genotypes has become available; ~30 tetraploids
- Different micro arrays are constructed a.o. SOLCAP and Wageningen Infinium array
- ~100 million different SNP positions
- Difference between exonic and intronic snp density

Allele frequencies

- Anywhere between 8 to 16 alleles per gene
- Nucleotide diversity
  - ~ 1 SNP/50 bp
    between two alleles
  - ~ 1 SNP/16 bp across alleles
  - On average 1 SNP/20 bp

‘Deep’ sequencing

- Necessary to get a ‘perfect’ (template) sequence, without sequencing mistakes
- Will improve the assembly (which is easier if sequence is better)
- Is important for a better gene identification and annotation of genes
- Calls for more ‘cleaning’ of data

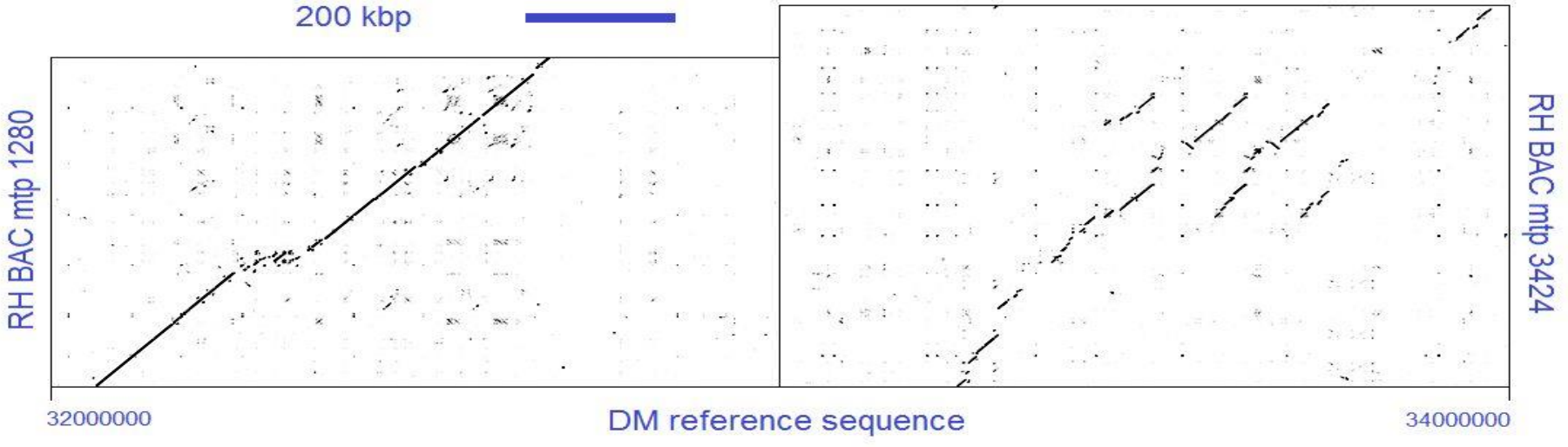
‘Broad’ sequencing

- Necessary to get full insight in the number of alleles of a particular gene
- Will allow a better prediction of haplotypes
- Is currently done in a number of studies incl. all tuber bearing potato accessions for a selected number of genes incl. earliness

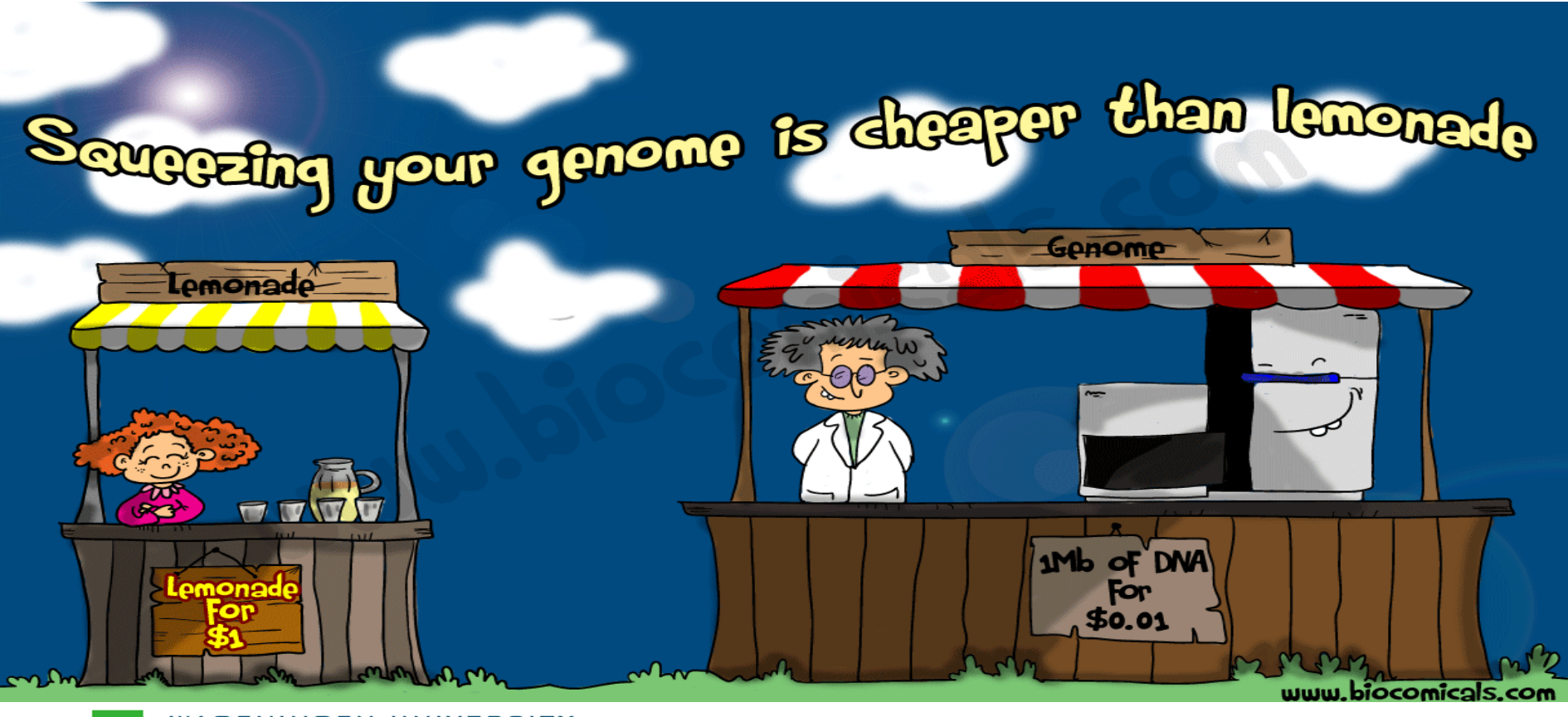
‘Broad’ sequencing

- Many SNP identified; ca 100 million
- Sequence of one genotype is not always indicative for the other sequence
- Potato = tomato =/= potato
- Analysis awaits further work but already many questions → how many, how deep, what technique?

Alignment of sequences shows large differences

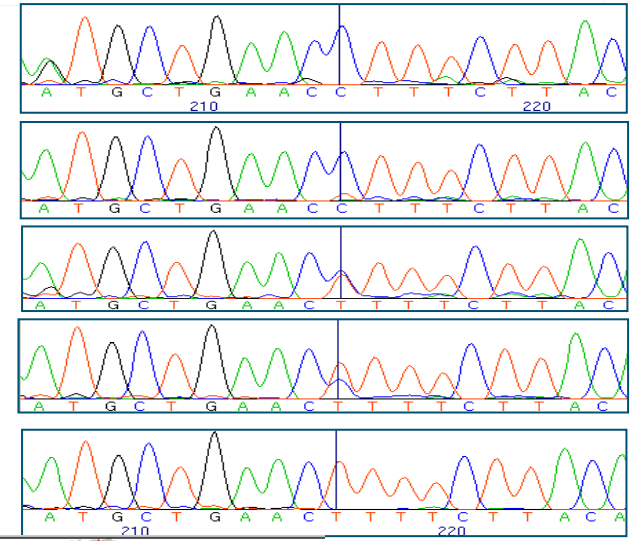


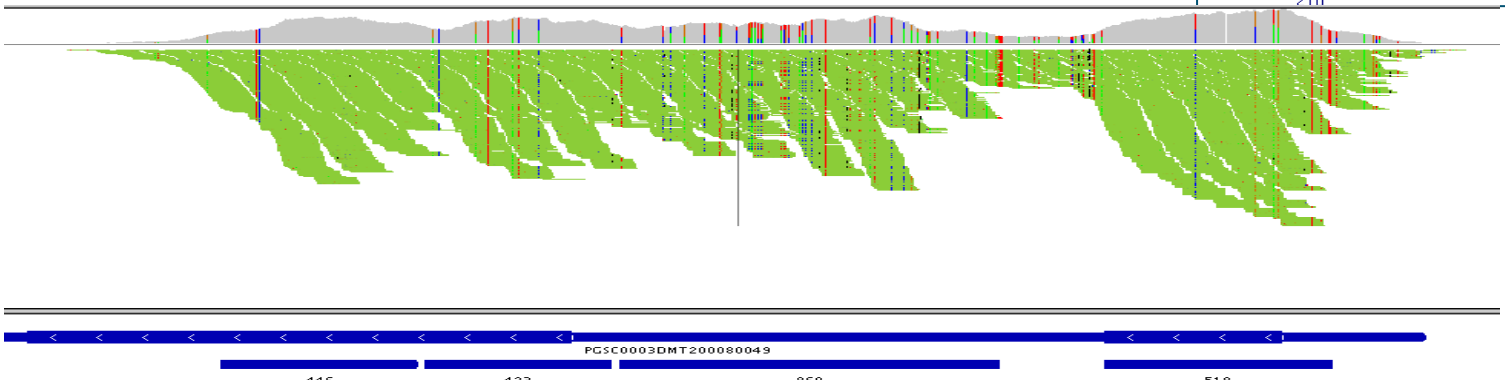
Genotyping by Sequencing



Genotyping-by-sequencing in tetraploids

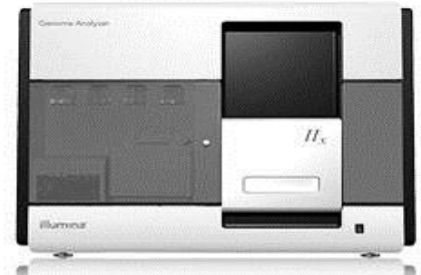
- Traditional Sanger reads
  - Peak ratio
- Next generation reads
  - Read depth ratio





Sequencing of tetraploid potato

- 800 genes selected in 83 cultivars
- Distributed over the genome (all chromosomes)
- Capture technology (sure select)
- Sequenced > 2 Mb; seq depth 75-100 x
- 130.000 SNPs, 30.000 in coding regions



PLANT BREEDING
WAGENINGEN UR

Haplotype diversity in cultivated potato germplasm

- Many rare alleles
- Many introgressed alleles
- High sequence divergence
- Average 3.2 per genotype

Sequence identity

95.6-99.9% within potato

94% tomato vs. potato

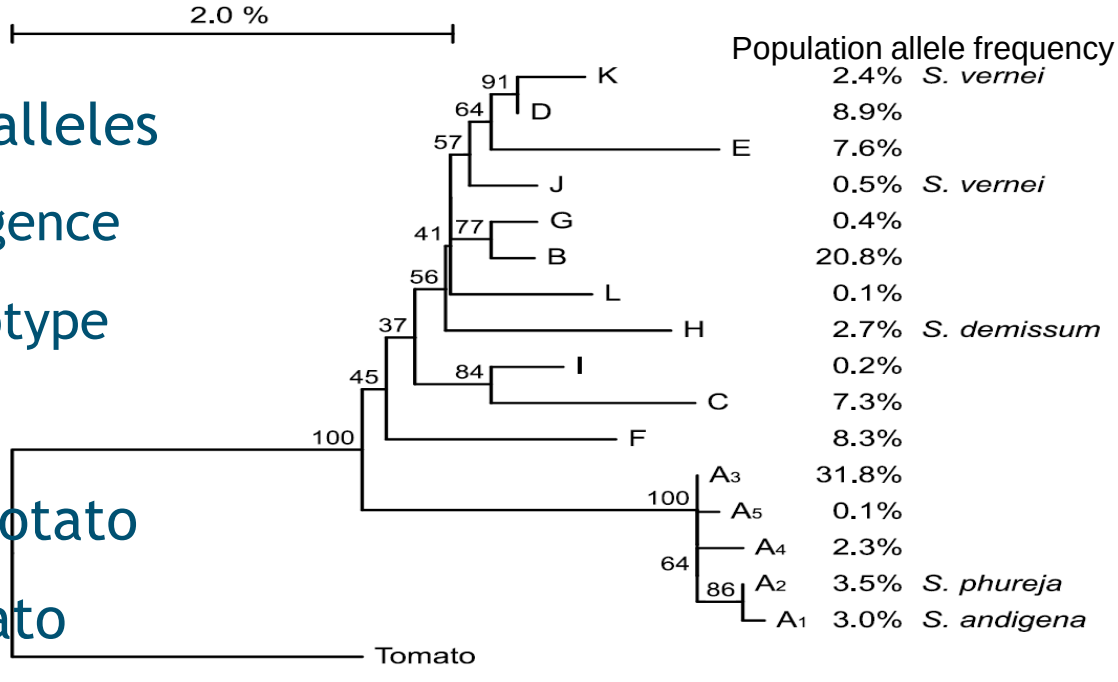


FIGURE 2. Dendogram of the 16 GWD haplotypes. The distances were computed using the Jukes-Cantor method and the tree inferred using the Neighbor-joining method. The percentage of replicate trees in which the associated haplotypes clustered together in the bootstrap test (1000 replicates) are shown next to the branches. For each allele the frequency and – when identified – the source is given. The tomato sequence was used as out-group to root the tree.

Potato Genome (Re-)Sequencing

Identify alleles underpinning phenotypic diversity across the entire genome and entire potato clade

(Re-)sequencing

Two different experiments:

1. use tetraploid varieties to assess the allelic variation in the gene encoding plant maturity using PCR cloning and Illumina sequencing.
2. Identify the sequence of the earliness locus CDF in tuber bearing wild species and landraces from north to South America using PACBIO

Plant maturity locus

From 500 tetraploid varieties 180 selected and finally 15 for PCR cloning and sequencing of intron and exon region of ~600 bp

Outcome:

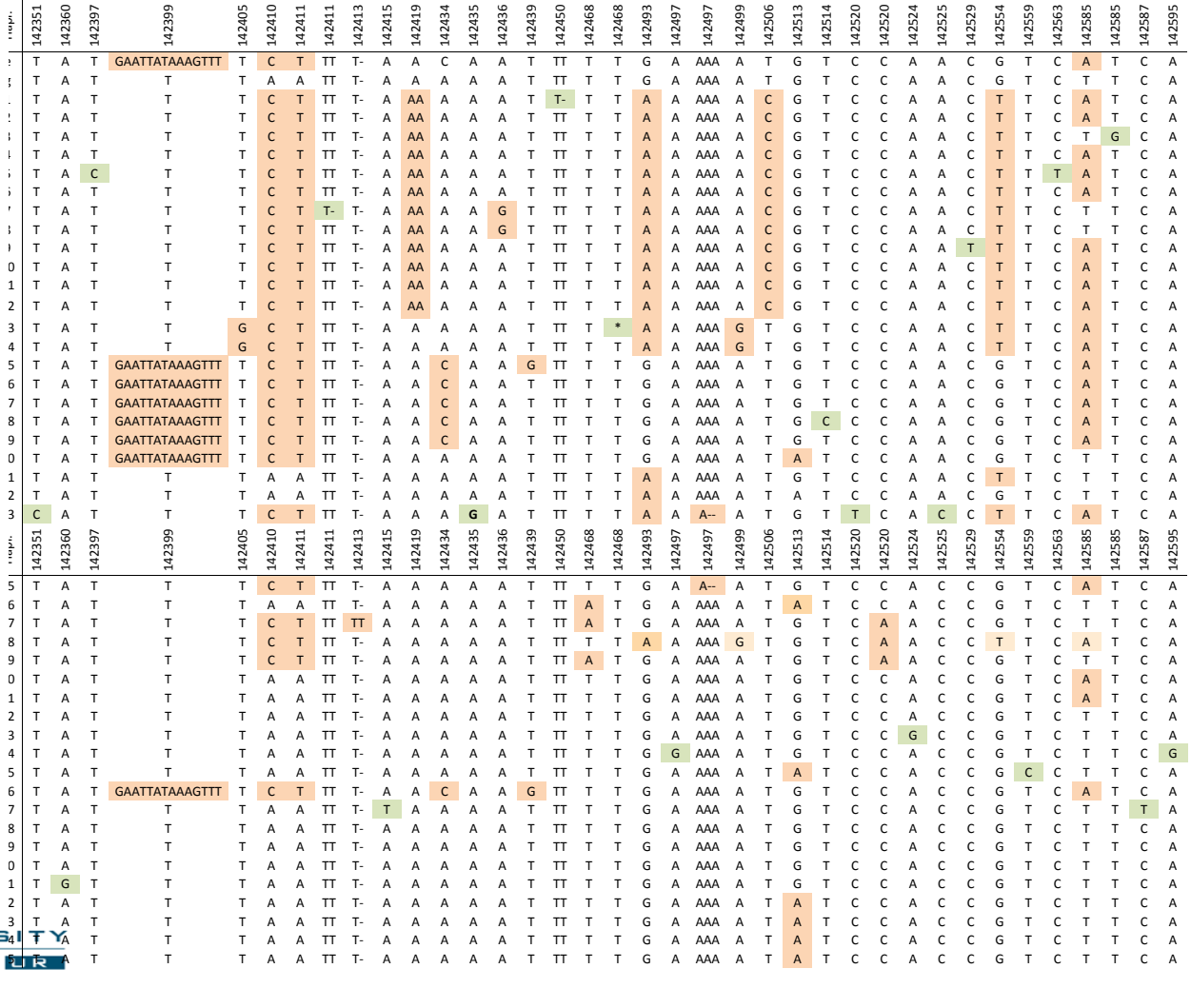
- ✓ 15 genotypes (varying from 3.5 till 8.3) show 3.4 different alleles
- ✓ 45 haplotypes
- ✓ Herald is 1278, Binella is 9,9,10,10
- ✓ 1 SNP per 8 bp!
- ✓ How many more haplotypes????

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Haplotype	Cultivars
1	Herald
2	Herald Kerpondy
3	Ar96
4	Ar96
5	Early Rose
6	Ar96 IVP4X-144-2
7	Herald
8	Herald
9	Binella
10	Binella
11	IVP4X-144-2
12	IVP4X-144-2
13	IVP4X-144-2
14	IVP92-057-3
15	Alpha
16	Alpha Aurora
17	Wur038
18	Wur038
19	Wur038
20	Kepplestone kidney
21	Wur038
22	Aurora
23	Ar96

Haplotype	Cultivars
24	Ar96
25	Wur038
26	Karnico
27	IVP4X-061-1
28	Aurora
29	Early Rose
30	Civa
31	Clva
32	Clva
33	Alpha
34	Early Rose
35	Kerpondy
36	Alpha
37	Karnico
38	IVP92-057-3
39	Karnico
40	DM reference StCDF1.2 Early Rose IVP92-057-3 RH4X-549-3 **Urgenta** **Saskia** **Vr-98-377** **Vtn 62-33-3**
41	Alpha
42	Kerpondy
43	Kerpondy
44	Karnico IVP4X-061-1
45	Aurora

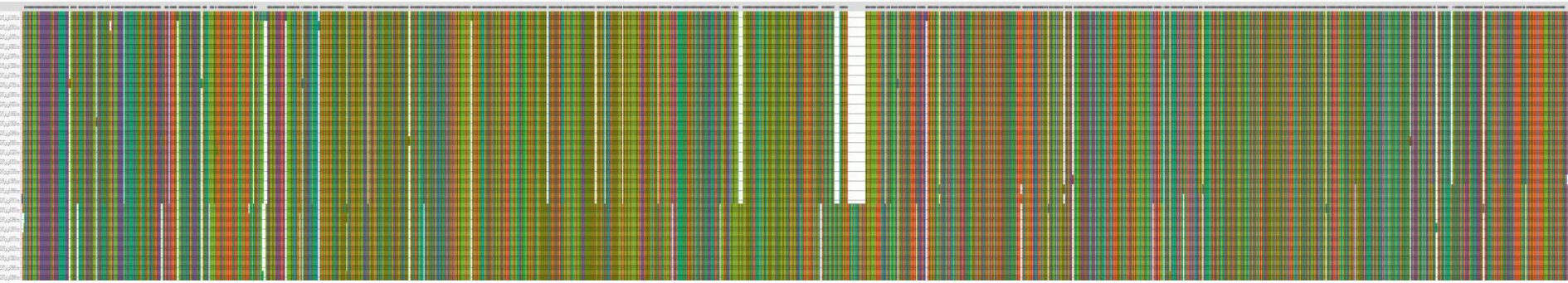
Difference in haplotype is often one SNP only, but more differences occur also



WAGENINGEN UNIVERSITY
WAGENINGEN UR

PacBio results of CDF gene

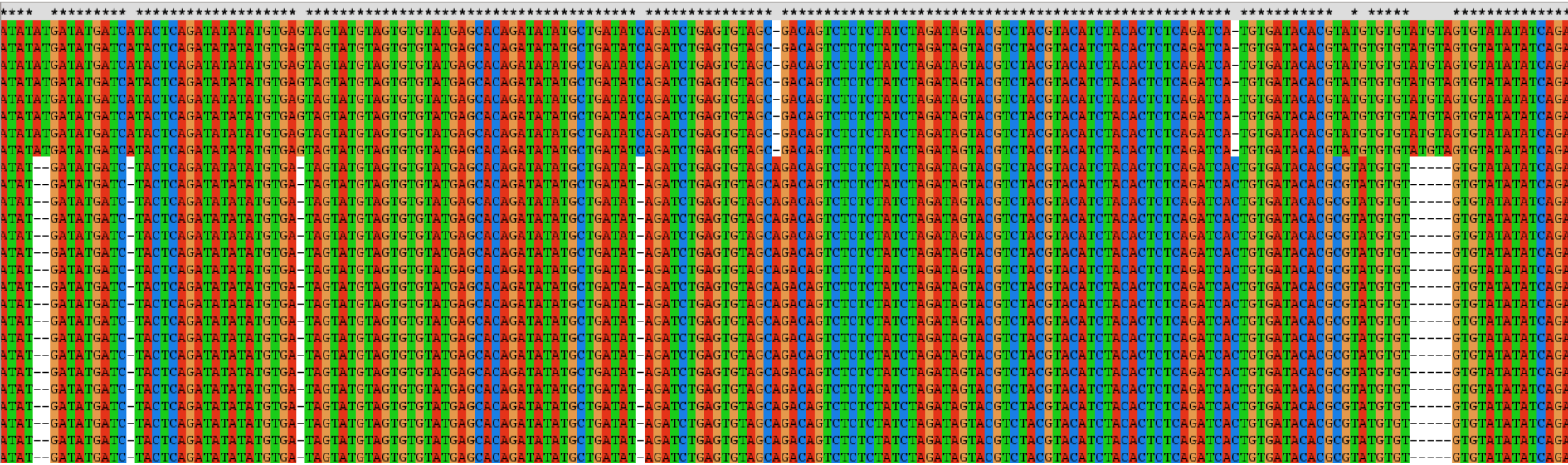
- ✓ About 150 genotpes analysed
- ✓ Sequence from one tetraploid genotype
- ✓ 2.5 kb after cleaning 1.7 kb
- ✓ 2 haplotypes



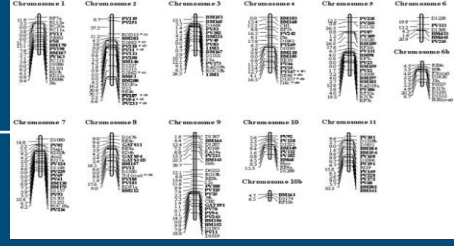
PLANT BREEDING
WAGENINGEN UR

PacBio results

- ✓ 200 bp fragment enlarged
- ✓ Still problems in aligning sequences, manual handling



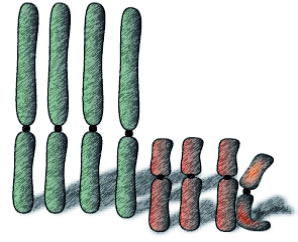
Markers in breeding / research



- Diploids
  - Genetic maps ; mapping qualitative / quantitative traits
  - Marker assisted introgression, marker-assisted selection, genetic distance/similarity, purity, pedigree relationships, etc.
- Tetraploids
  - Allopolyploids: inheritance as in diploid, markers/genes not always unique
  - Autopolyploids: much more complex genetics

Tetraploid analysis methods and software

- SNP genotyping: dosage scoring
  - 5 classes: nulliplex - quadruplex
  - fitTetra (Voorrips et al. 2011)



- Linkage analysis and QTL mapping
  - TetraploidMap (Hackett & Luo, 2003; Hackett et al. TAG 2014) limitations (a.o. max 50 markers / chromosome)
  - New high-throughput tools needed
- Software to simulate tetraploid progenies
  - PedigreeSim (Voorrips & Maliepaard, 2012) for development and validation of new tools

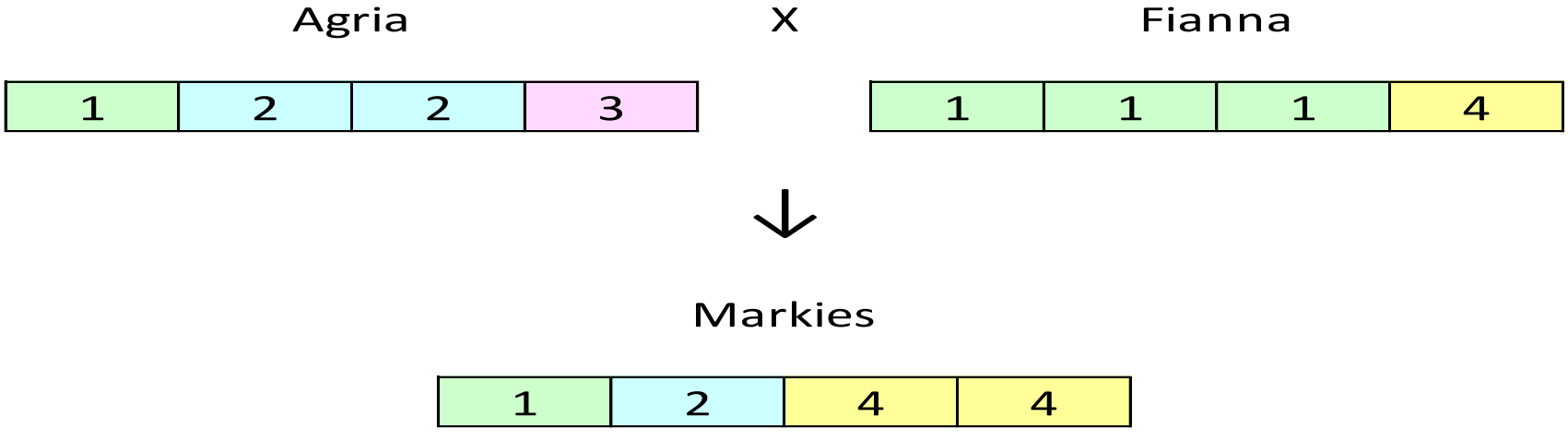
Why so complex ?

- *Different dosages of alleles possible*
  - Extra problem if also null alleles are possible
- *Different alleles:* multiple allelism
- *More than 1 allele per locus* passed on to gamete
  - Presence of one allele ≠ absence of the other

Why so complex ?

- *More possibilities phase* of linked loci
  - Coupling and three times repulsion in a tetraploid
- *Recombination possible over more than two homologs*
  - Bivalents: preferential or random pairing homologs
  - Quadrivalents: possibility of double reduction

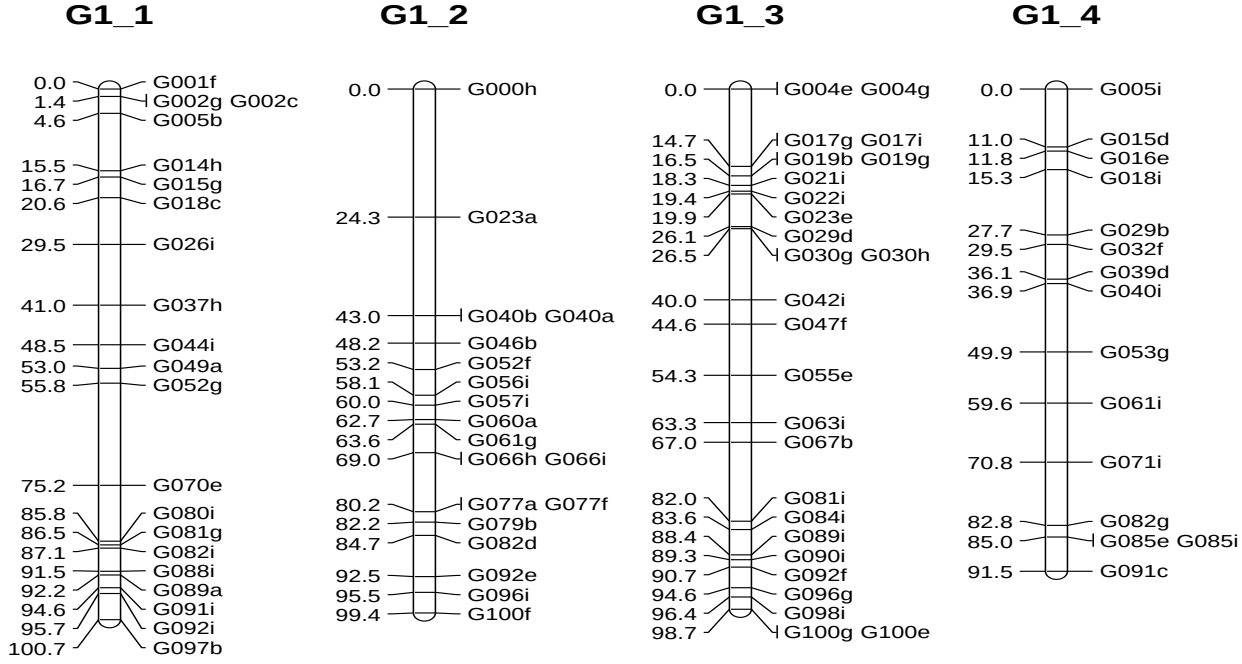
Haplotype inheritance for genes can be followed based on pedigree info



Markies has received two copies of haplotype 4 from Fianna:
Gene is near the end of Chr.1 → example of double reduction.

Linkage map construction

- Map the markers in each parental haplogroup



Research and breeding aims

- Understanding the genetic structure of potato germplasm
  - How many haplotypes are there in cultivars at different loci?
  - What is the best combination of alleles for a given trait?
- What loci are involved in (quantitative) quality traits?
  - Haplotypes with negative/positive effects
  - Haplotypes with dominant/recessive/additive effects
  - -> marker assisted selection

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Uitdewilligen et al. (2013). A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. **PLoSONE 8(5): e62355**.

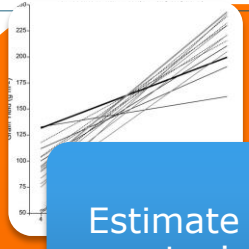
Allele effects (kinship)

LogP_ kinship = 6.939

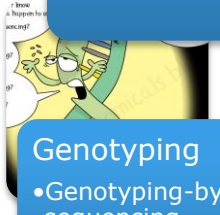
allele	abs freq	rel freq	Effect_Starch_PO4_nmol	se_Starch_PO4_nmol
A	300	0.414	0.00	0
B	147	0.203	1.01	0.439
C	62	0.086	1.66	0.671
D	61	0.084	2.65	0.584
E	51	0.070	0.01	0.575
F	65	0.090	-0.20	0.506
G	4	0.006	0.06	2.213
H	15	0.021	6.10	1.061
I	2	0.003	-1.80	2.707
J	2	0.003	2.30	2.383
K	15	0.021	1.12	1.075
L	0	0.000	*	*

Breeding with databases:BreeDB

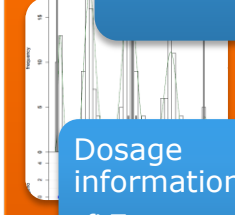
Field trial
- Observations
- ~omics analysis



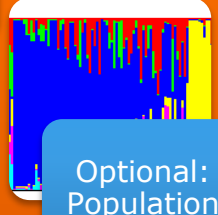
Estimate genotypic means



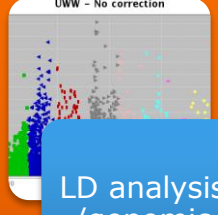
Genotyping
- Genotyping-by-sequencing
- SNP arrays



Dosage information
- fitTetra



Optional: Population Structure correction



LD analysis /genomic selection

Marker selection

BreeDB

Precision Breeding

Research Focus: *in sillico* cultivar design

Precision Breeding

- Well characterized (elite) germplasm collections
- Design the desired genotype behind the computer
    - Breeding goals as input
- Obtain the most effective breeding strategy, starting with your current (elite) germplasm
- Cross (and if necessary genotype the offspring)
- Select candidates similar to the *in silico* designed line
    - Characterize and add to germplasm collection & Re-train predictive algorithms

Concluding

- Haplotyping is possible but should be automated
- Alignment of NGS sequences is a quite usable approach for haplotyping in tetraploid potato
  - Haplotypes of 1-1.5 kbp length were easily obtained in this project
  - For individual genes, spending time on 'manual' data analysis will significantly improve results
  - For genome-wide analyses, the automated haplotyping results are too irregular and fragmented ..... and it is necessary to use longer sequences → they are slowly comming eg PacBio
- Best allele combination predictions per trait are next challenge

Acknowledgements

Wageningen UR Plant Breeding:

Jose Abelenda, Christian Bachem, Jan de Boer, Theo Borm, Herman van Eck, Richard Finkers, Chris Maliepaard, María del Mar Pérez Nicolás, Roeland Voorrips, Peter Vos, Jan Uitdenwilligen, Annemarie Wolters

Biometris WUR: Joao Paulo, Fred van Eeuwijk,

Potato breeding, growing & processing companies

Agrico, Averis, HZPC, C. Meijer, KWS potato, VAVI, HPA

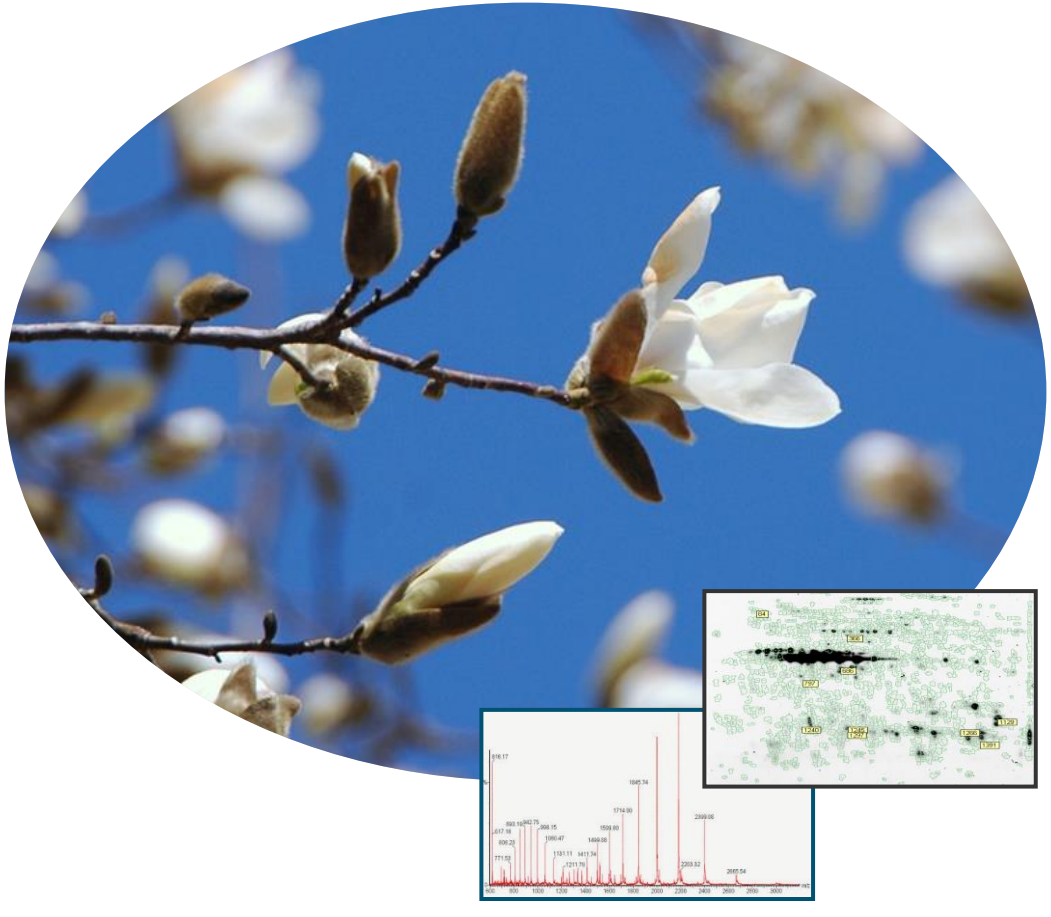
WAGENINGEN UNIVERSITY
WAGENINGEN UR

Sponsors

- Technology foundation STW
- Potato Breeding companies (2002-2012)
- Technological Top Institute - Green Genetics
- Ministry of Economic Affairs, Agriculture & Innovation
- Centre for BioSystems Genomics/NGI
- Wageningen University and Research Centre

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Questions?



WAGENINGEN UNIVERSITY
WAGENINGEN UR