

Mapping of loci associated with tocopherol composition **and oleic acid content** using genotyping-by-sequencing approach in sunflower

Rim Gubaev, Stepan Boldyrev, Alina Chernova, Elena Martynova, Tatiana Kovalenko, Svetlana Goryunova, Tatiana Peretyagina, Cecile Ben, Laurent Gentzbittel, Philipp Khaitovich & **Yakov Demurin**



Skoltech

Сколковский институт науки и технологий



ВНИИМК

This project was supported by the RFBR grants №20-316-90051 and №20-316-80002.

Why oleic acid and tocopherols?



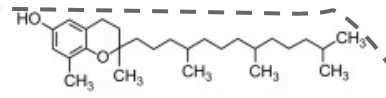
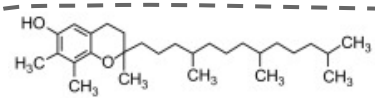
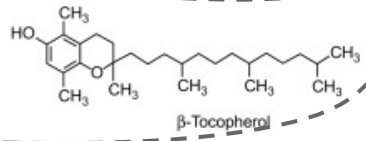
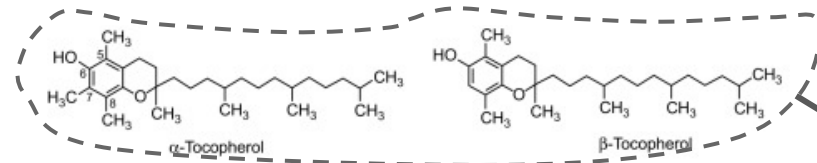
Confectionery
sunflower



Oilseed
sunflower

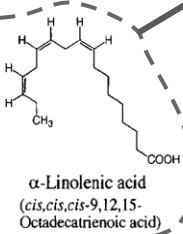
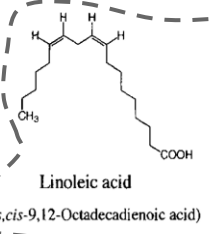
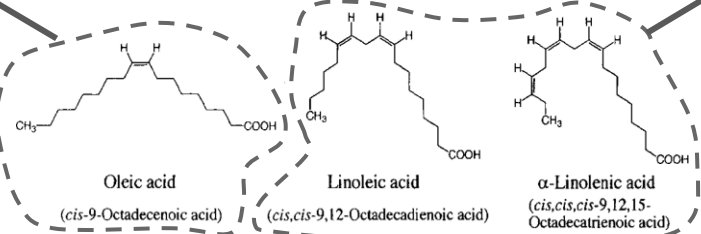


Frying oils



Salad oils

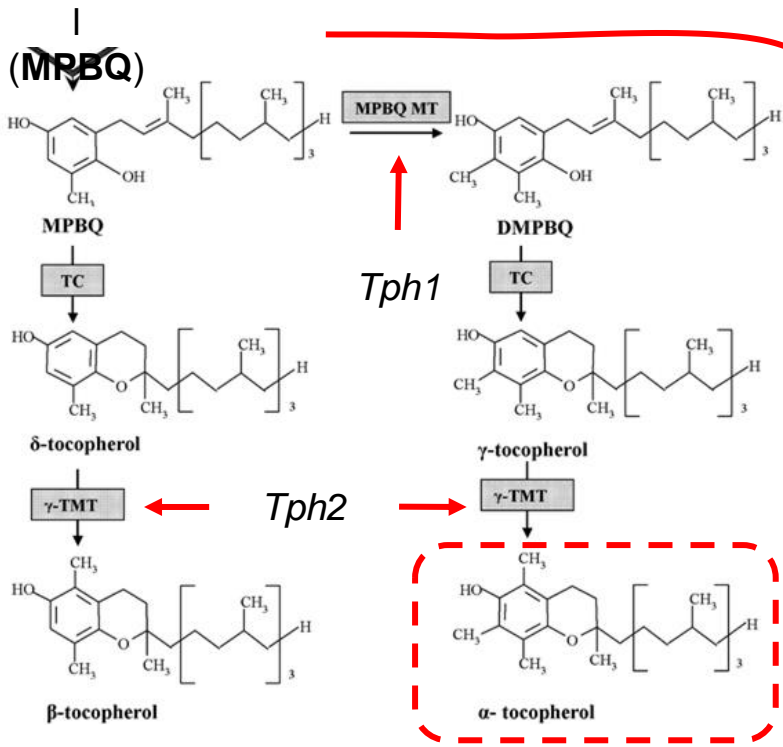
- Resistant to oxidation
- High content of oleic acid
- Gamma and delta tocopherols



- Vitamin E + polyunsaturated fats
- Moderate content of oleic acid
- Alpha and beta tocopherols

Biosynthesis and causal genes

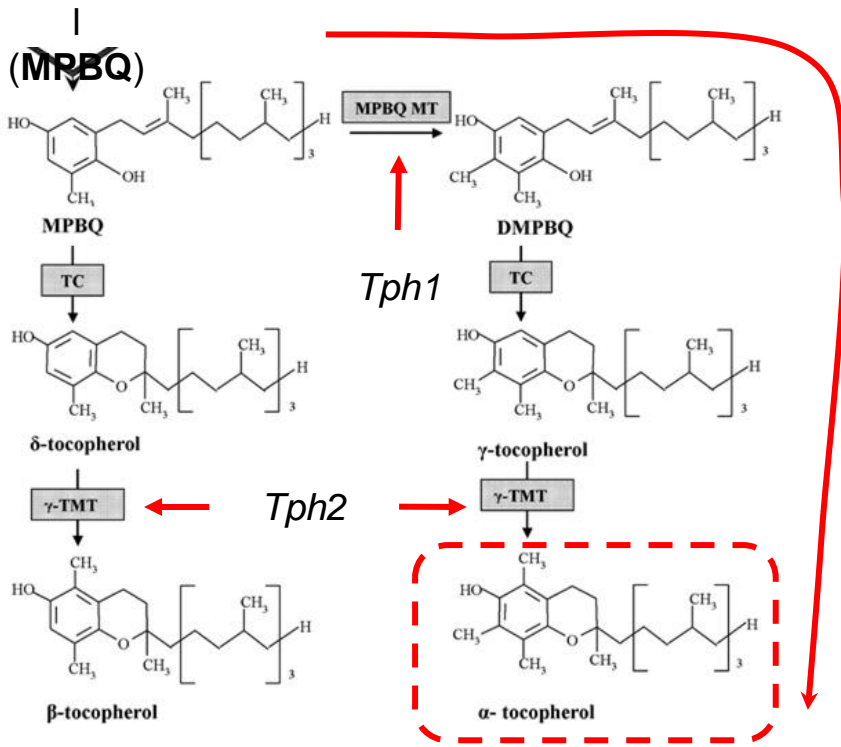
2-methyl-6-phytylbenzoquinone
Tocopherols biosynthesis



The *Tph1* and *Tph2* mutants were obtained by screening genetic collection including one from VIR

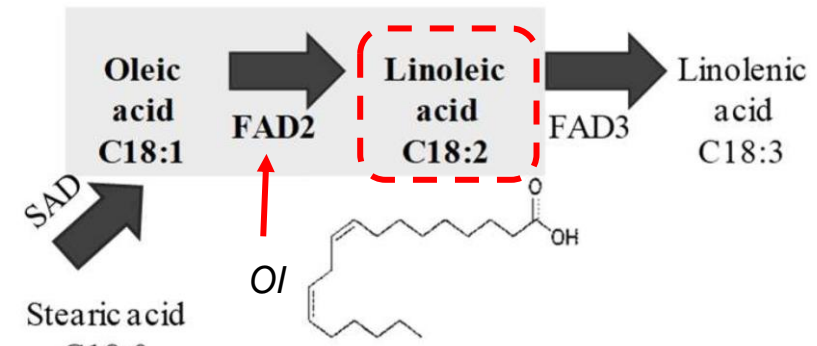
Biosynthesis and causal genes

2-methyl-6-phytylbenzoquinone
Tocopherols biosynthesis



The *Tph1* and *Tph2* mutants were obtained by screening genetic collection including one from VIR

Oleic acid (FA18:1)
biosynthesis



The *OI* mutant plant was obtained in VNIIMK by chemical mutagenesis in 1970s

It looks like there should be a **three** genes! **Two** for tocopherol composition and **one** for oleic acid content!

Genetic control is more complex for tocopherols and oleic acid

Theor Appl Genet (2006) 113:783–799
DOI 10.1007/s00122-006-0320-3

ORIGINAL PAPER

Ty3/gypsy-like retrotransposon knockout of a 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase is non-lethal, uncovers a cryptic paralogous mutation, and produces novel tocopherol (vitamin E) profiles in sunflower

Shunxue Tang · Catherine G. Hass · Steven J. Knapp

Received: 6 February 2006 / Accepted: 13 May 2006 / Published online: 9 August 2006
© Springer-Verlag 2006

Theor Appl Genet (2006) 113:767–782
DOI 10.1007/s00122-006-0320-4

ORIGINAL PAPER

Three non-allelic epistatically interacting methyltransferase mutations produce novel tocopherol (vitamin E) profiles in sunflower

Catherine G. Hass · Shunxue Tang · Scott Leonard · Maret G. Traber · Jerry F. Miller · Steven J. Knapp

2002

B. Pérez-Vich · J.M. Fernández-Martínez
M. Grondona · S.J. Knapp · S.T. Berry

Stearoyl-ACP and oleoyl-PC desaturase genes cosegregate with quantitative trait loci underlying high stearic and high oleic acid mutant phenotypes in sunflower

García-Moreno et al. *BMC Plant Biology* 2012, 12:71
<http://www.biomedcentral.com/1471-2229/12/71>

RESEARCH ARTICLE

Open Access

Genetic basis of unstable expression of high gamma-tocopherol content in sunflower seeds

María J García-Moreno, José M Fernández-Martínez, Leonardo Velasco and Begoña Pérez-Vich*

Mol Breeding (2016) 36:106
DOI 10.1007/s11032-016-0527-2



Mapping quantitative trait loci controlling oil content, oleic acid and linoleic acid content in sunflower (*Helianthus annuus* L.)

Ameena Premnath · Manivannan Narayana · Chandirakala Ramakrishnan · Senthil Kuppasamy · Vanniaraajan Chockalingam

Received: 7 September 2015 / Accepted: 4 July 2016 / Published online: 14 July 2016
© Springer Science+Business Media Dordrecht 2016

Breeding Science 68: 596–605 (2018)
doi:10.1270/jbbs.18051

Research Paper

Construction of a high-density genetic linkage map and QTL mapping of oleic acid content and three agronomic traits in sunflower (*Helianthus annuus* L.) using specific-locus amplified fragment sequencing (SLAF-seq)

Fei Zhou^{1,2}, Yan Liu^{1,2}, Chunbo Liang², Wenjun Wang², Cen Li², Yongli Gao², Jun Ma², Ying Yu², Lijuan Fan², Yubo Yao², Dongsheng Zhao², Xuemei Liu² and Xutang Huang^{2*}

¹ College of Life Science, Northeast Forestry University, Harbin, 150040, China
² Institute of Industrial Crops, Heilongjiang Academy of Agricultural Sciences, Harbin, 150086, China

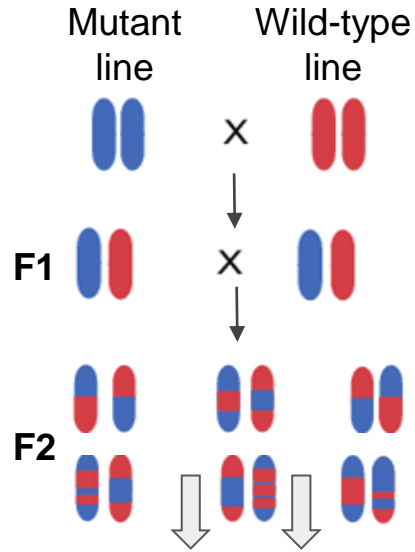
Additional factors:

- Non-allelic genes with moderate effects
- Loci with minor effects
- Genetic background

Goals of the present study:

- Map major effect loci
- Check whether the minor loci are also specific to VNIIMK lines
- Find potential markers for MAS

Experimental crosses and phenotyping



Wild-type lines (VK303 and VK101):

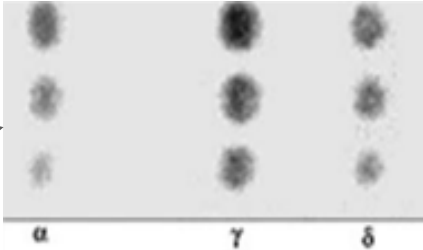
- ~80-99% of alpha tocopherol
- ~32-55% oleic acid

Mutant lines (VK195 and VK876):

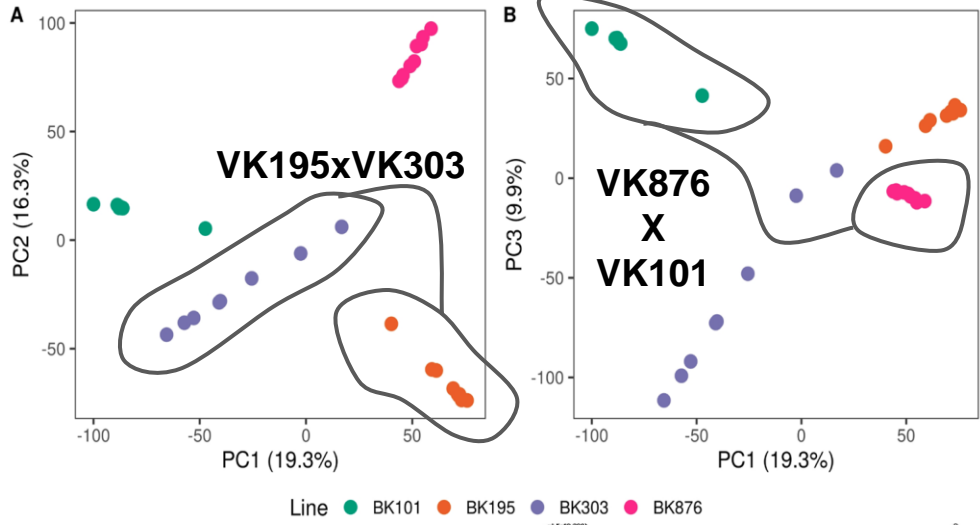
- ~40-60% of gamma tocopherol
- ~40-60% of delta tocopherol
- ~83-93% of oleic acid

144 F2 plants + 7 plants for each of the parents collected For each of the population!

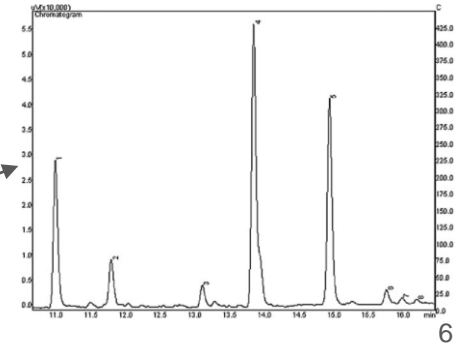
Tocopherol profiling:
Thin-layer chromatography with subsequent densitometry



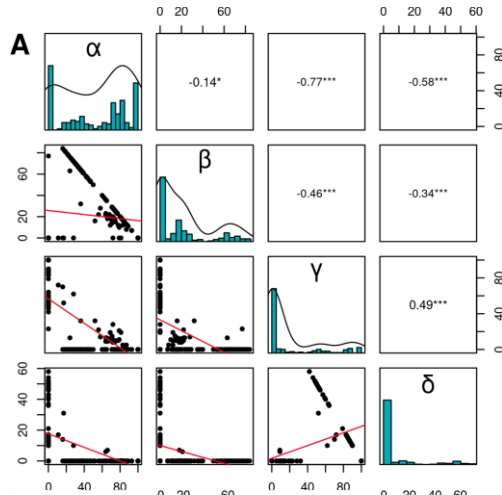
Genetic diversity of the studied lines



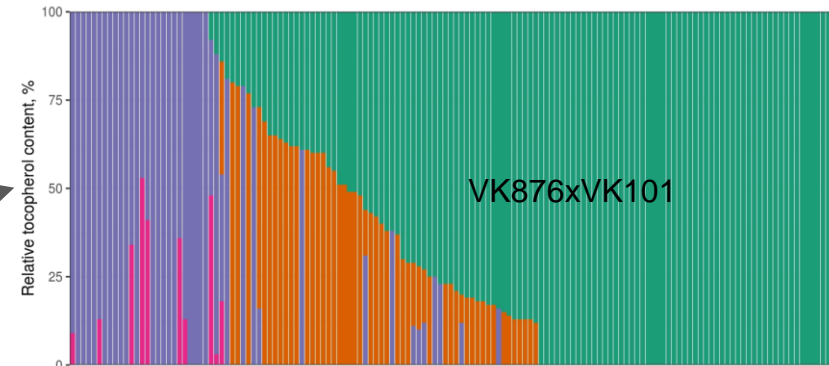
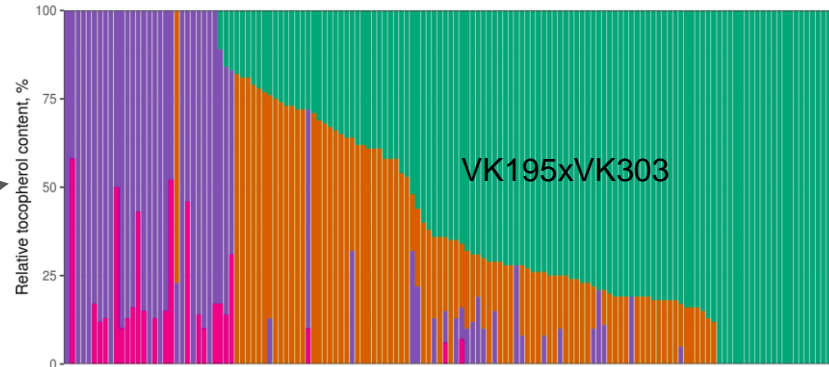
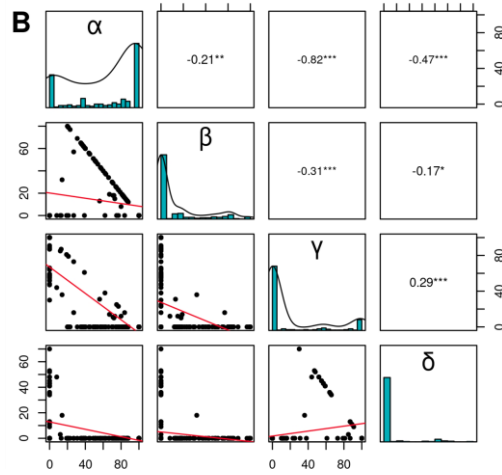
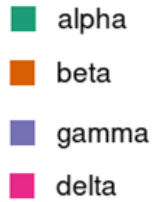
Oleic acid profiling:
gas chromatography followed by flame ionization detection



Tocopherol composition

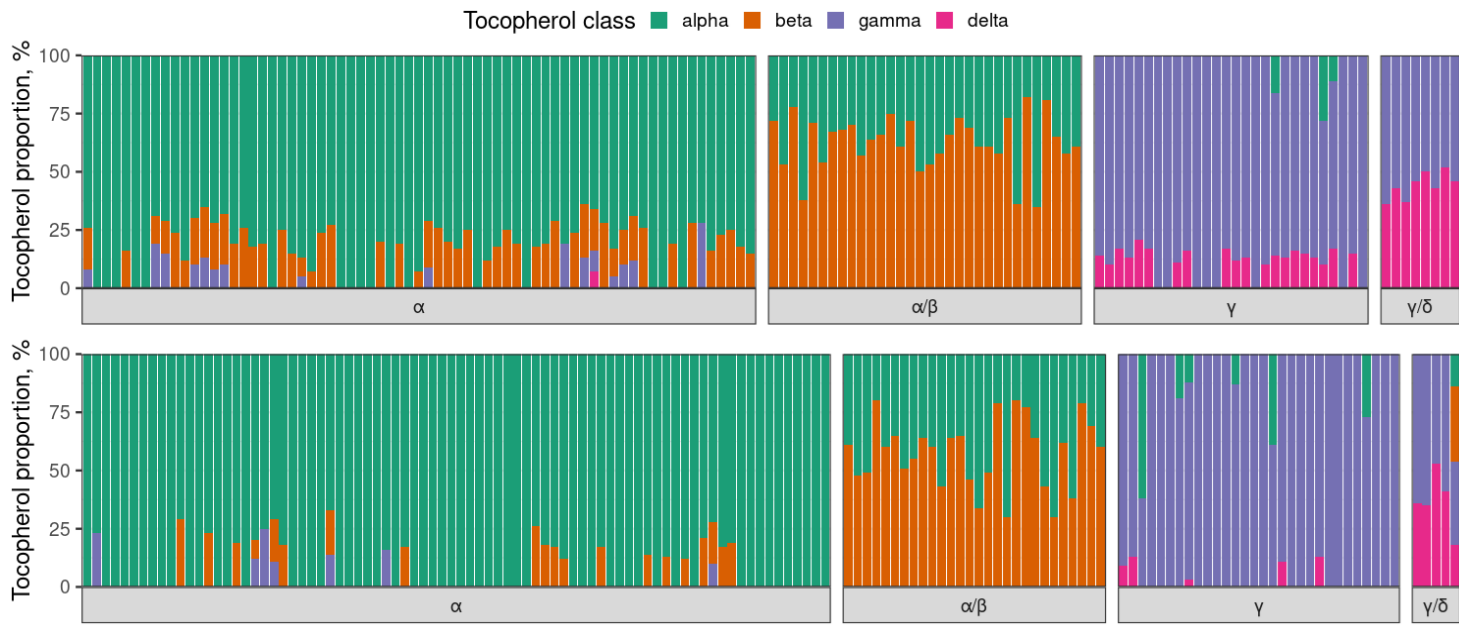


Tocopherol class



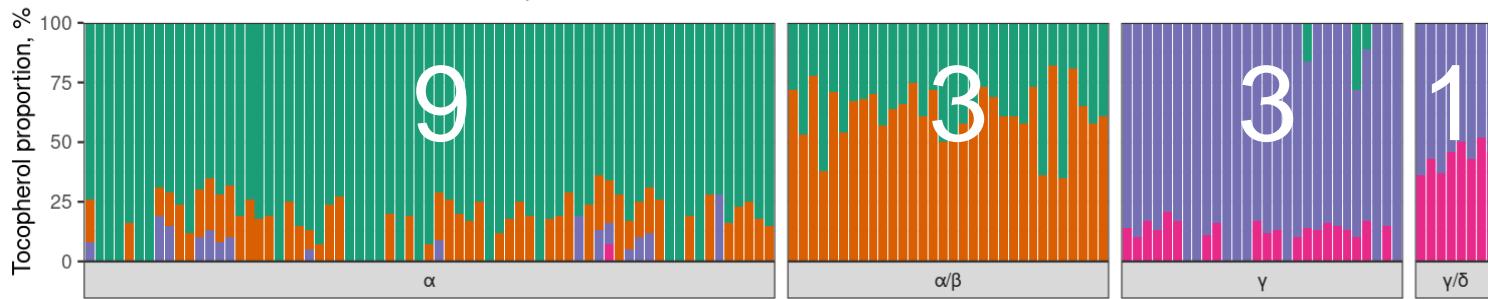
- Observations are dependent
- Distributions are not normal

Not so bad if we look through dihybridism model!



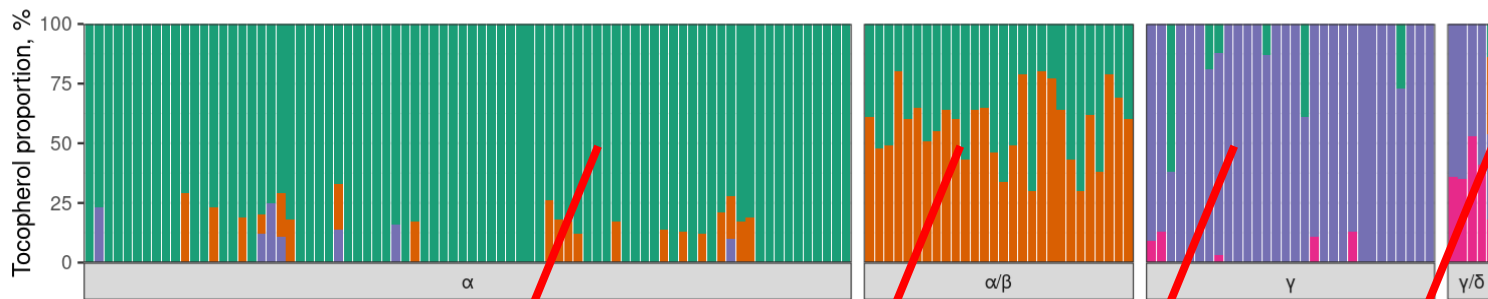
Not so bad if we look through dihybridism model!

Tocopherol class ■ alpha ■ beta ■ gamma ■ delta



VK195xVK303

Chi-square goodness of fit
test p -value p -value - **0.45**



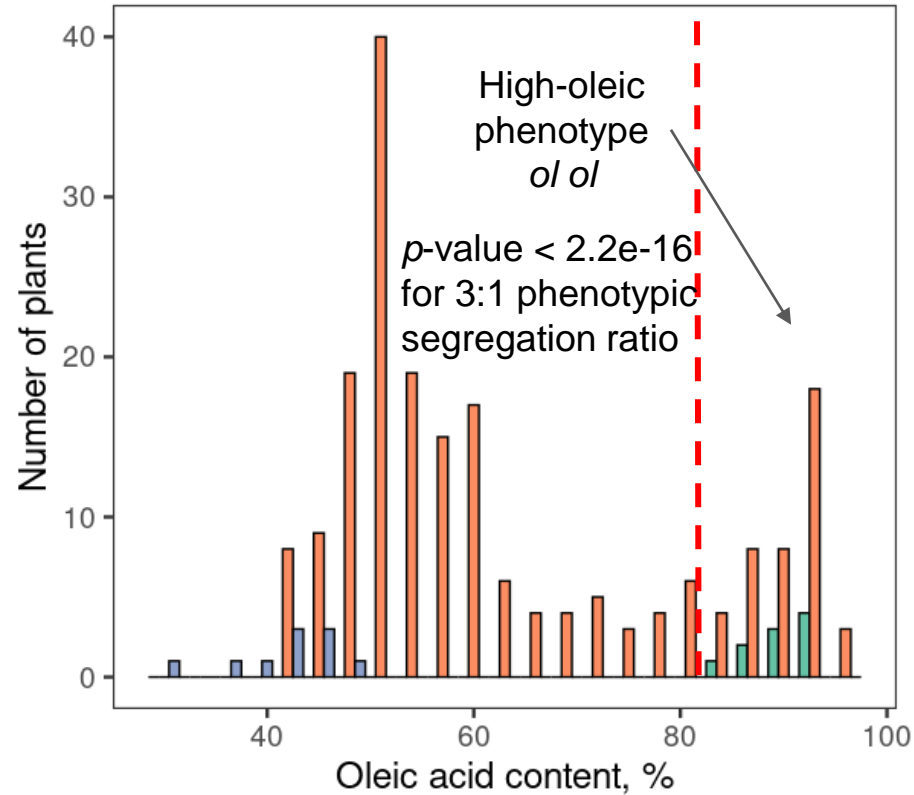
VK876xVK101

Chi-square goodness of fit
test p -value - **0.53**

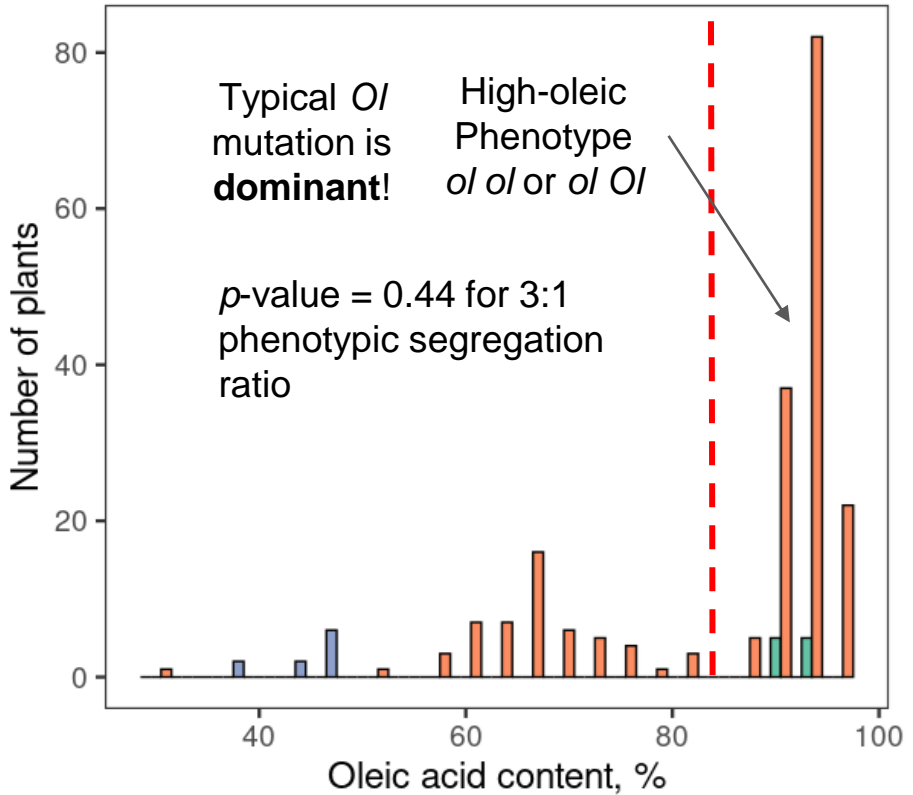
	alpha	alpha/beta	gamma	gamma/delta
<i>Tph1</i>	1	0	1	0
<i>Tph2</i>	1	1	0	0

Oleic acid content - a semi-quantitative case!

VK195xVK303



VK876xVK101

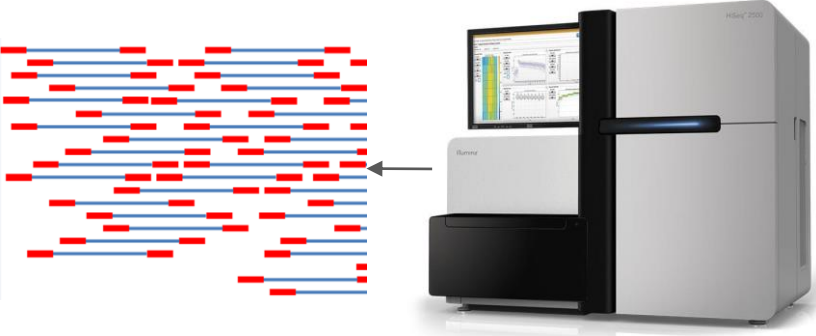


■ wild type parent
 ■ mutant parent
 ■ progeny

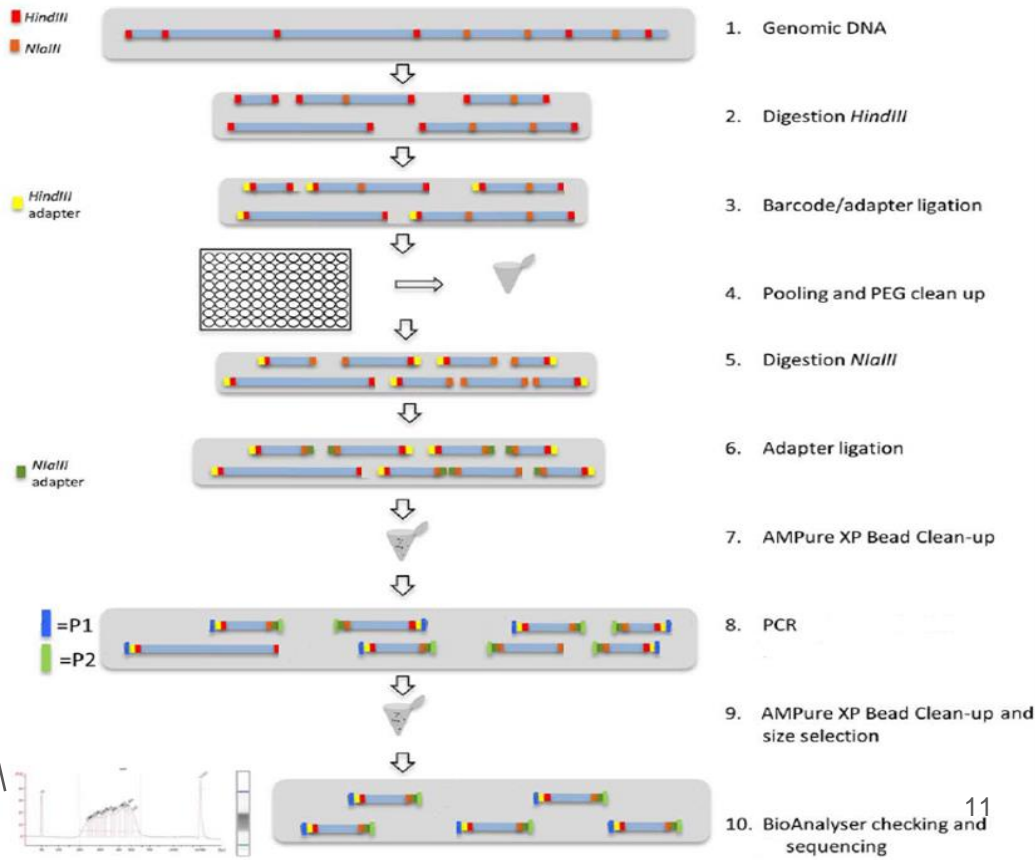
Genotyping strategy for QTL mapping

What was previously used?

- Indel markers
- SSR markers
- SSCP markers



Genotyping-by-sequencing approach



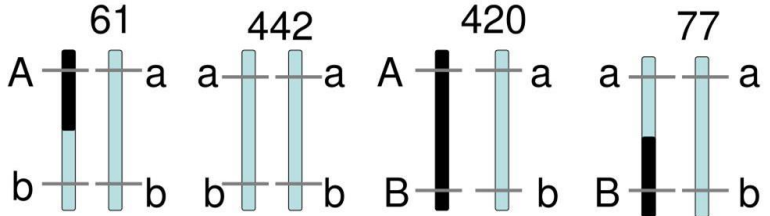
Genetic map construction: LOD and RF

Recombination Fraction

Recombination Fraction $\theta =$

Recombinants / Total =

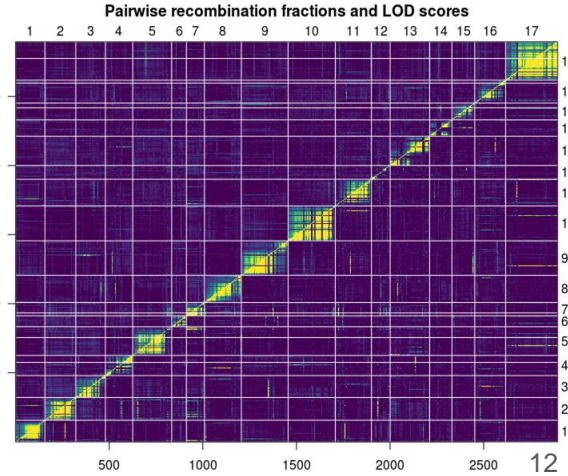
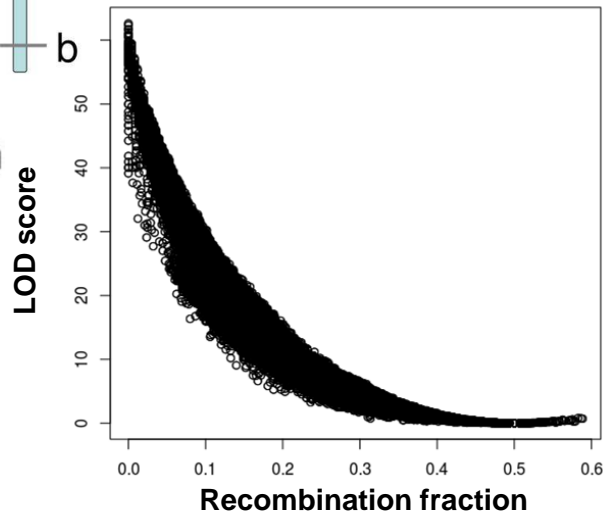
$$61 + 77 / 61 + 77 + 442 + 420 = 138 / 1000 = 13.8\%$$



- LOD score based on recombination

- $LOD(\theta) = \log \left[\frac{(\theta)^R (1 - \theta)^{NR}}{(\theta = 1/2)^{R + NR}} \right]$

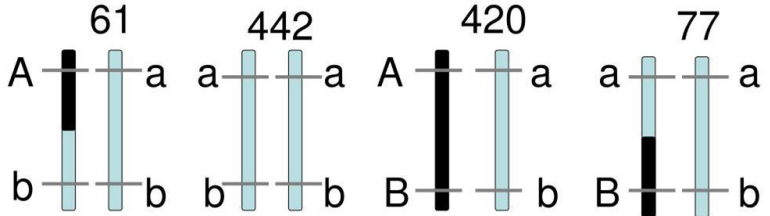
Relationship between LOD and RF



Genetic map construction: LOD and RF

Recombination Fraction

Recombination Fraction $\theta =$
 Recombinants / Total =
 $61 + 77 / 61 + 77 + 442 + 420 = 138 / 1000$
 $= 13.8\%$



- LOD score based on recombination

- $LOD(\theta) = \log \left[\frac{(\theta)^R (1 - \theta)^{NR}}{(\theta = 1/2)^{R + NR}} \right]$

1) Build genetic map *de novo*:

Calculate pairwise LOD and RF

Form linkage groups (chromosomes)

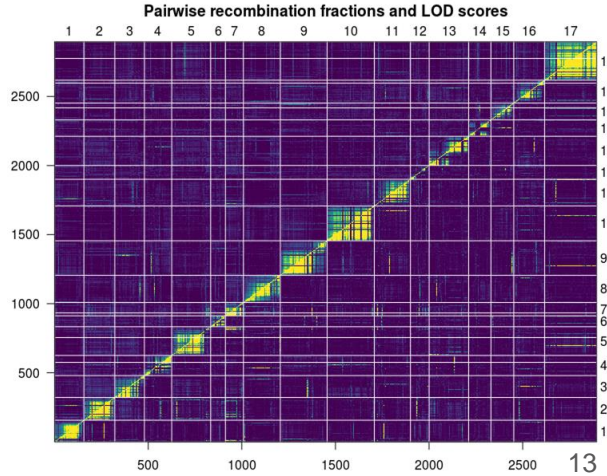
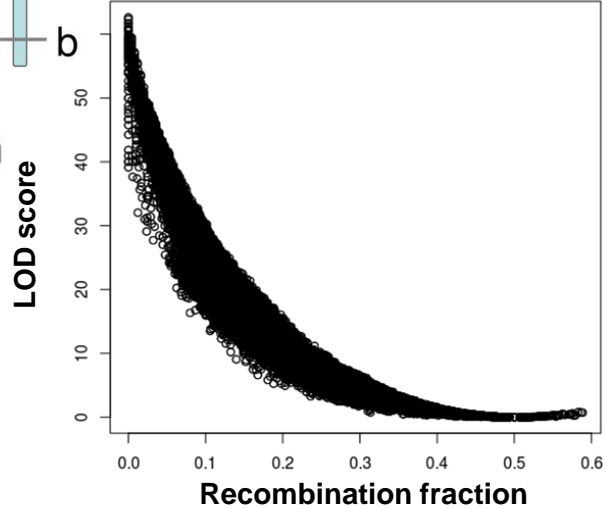
Order markers

2) Reestimate physical map:

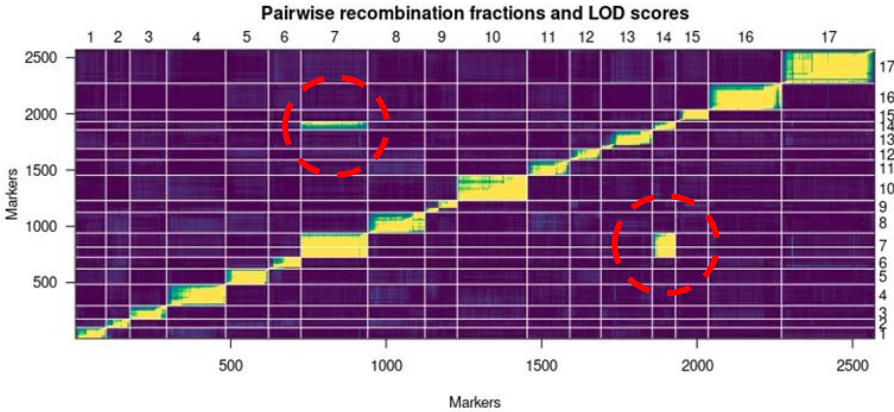
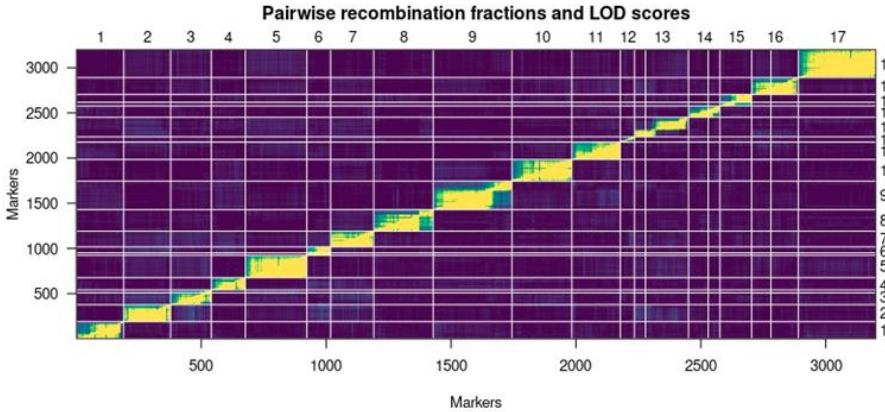
Filtering the markers

Re-ordering markers within linkage groups (chromosomes)

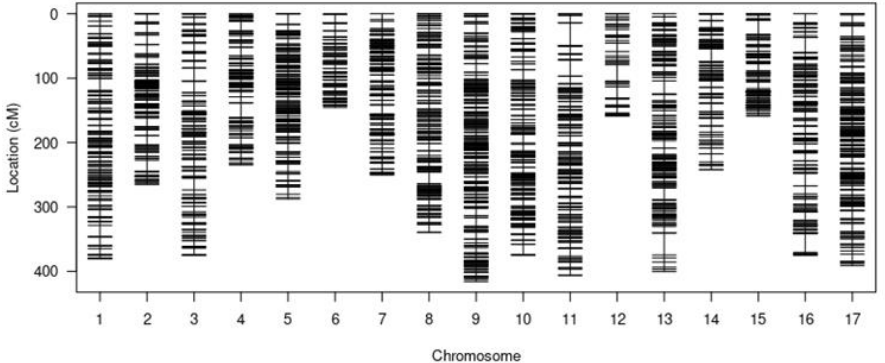
Relationship between LOD and RF



Assembled genetic maps

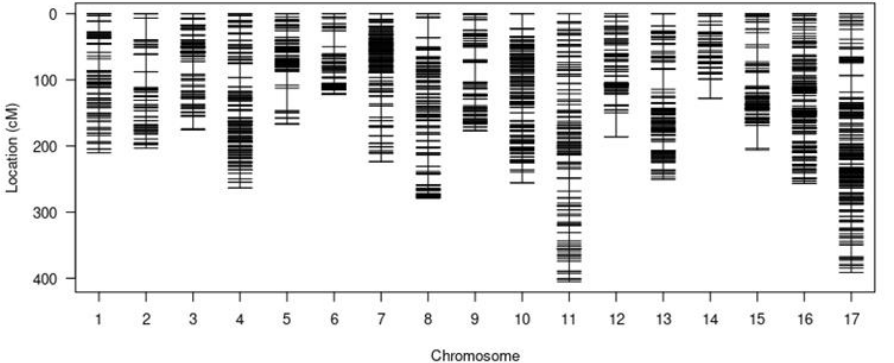


VK195xVK303



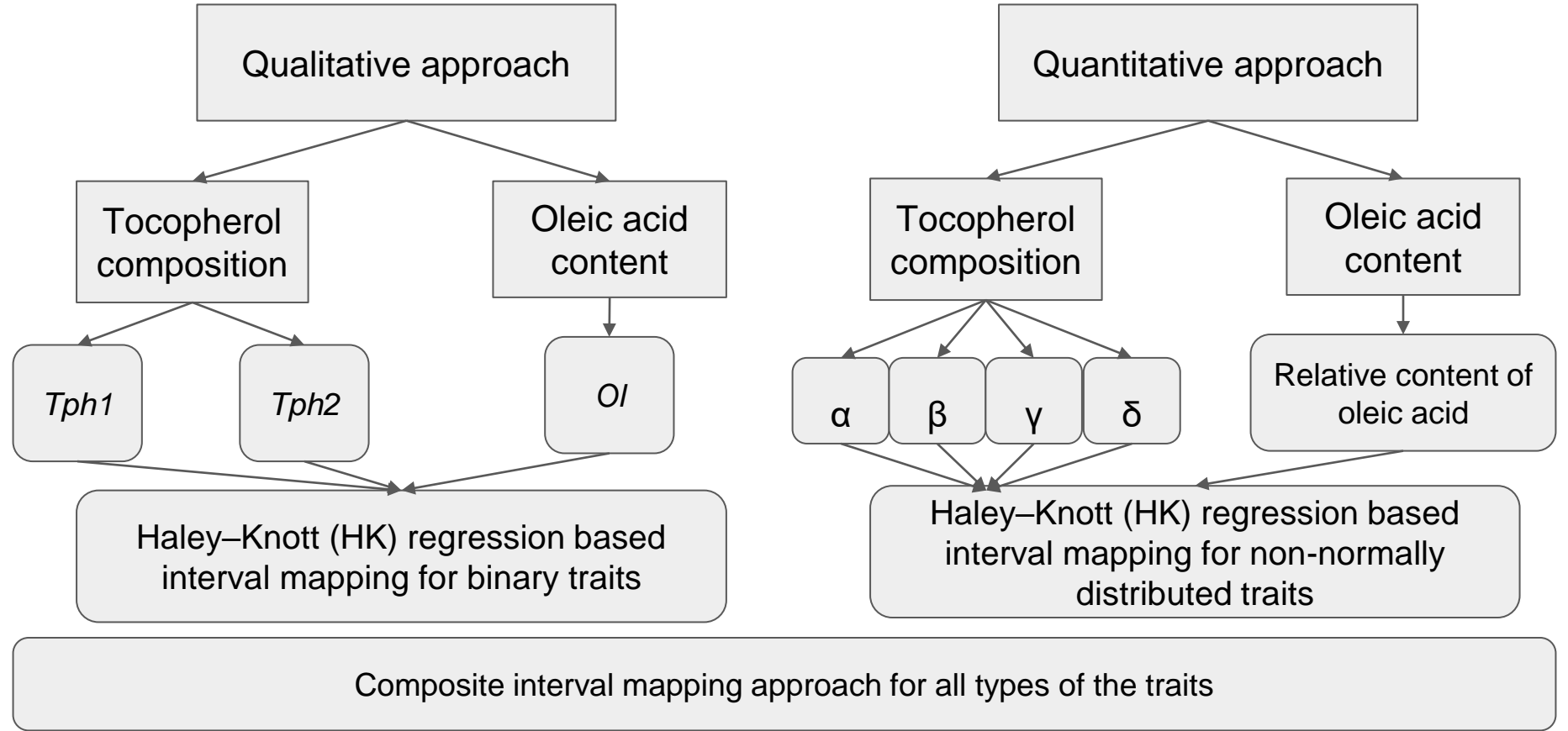
Number of the markers - 3200 ; Length- 5197.7 cM

VK876xVK101



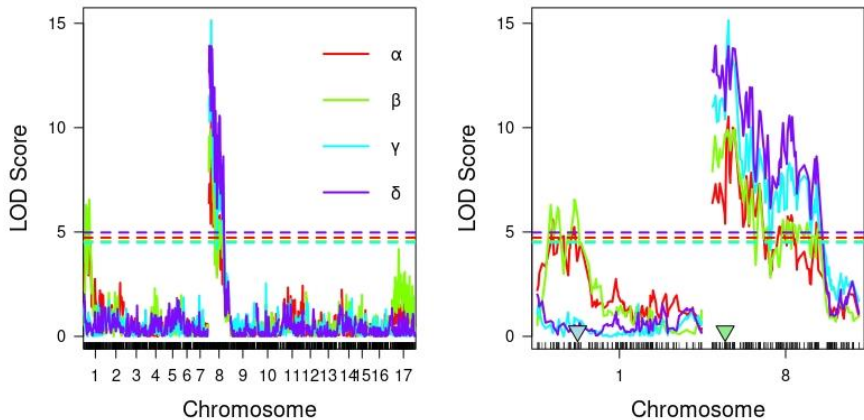
Number of the markers- 2571; Length 3898.8 cM

Mapping strategies



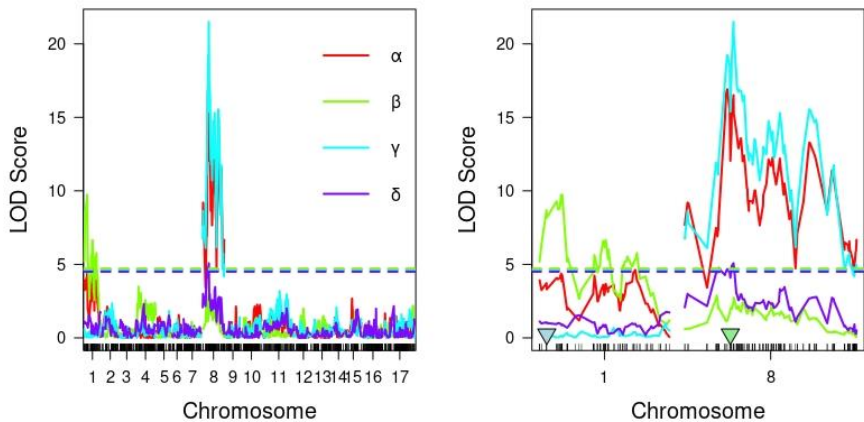
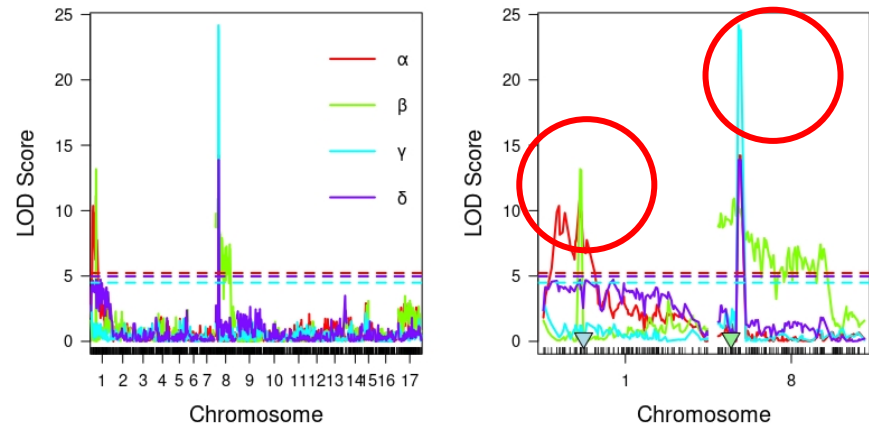
Tocopherol composition mapping (quantitative approach)

Interval mapping (non-parametric)

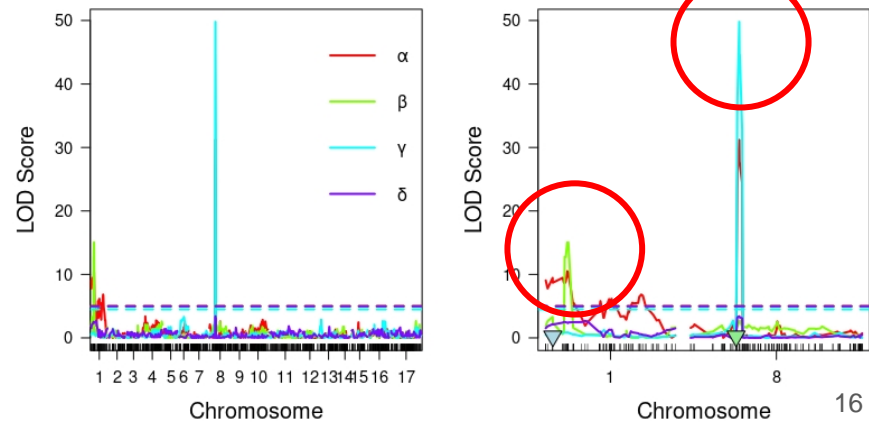


**VK195
x
VK303**

Composite Interval mapping

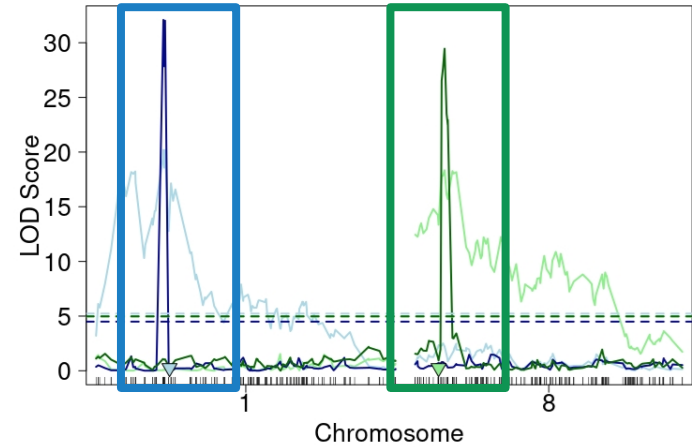
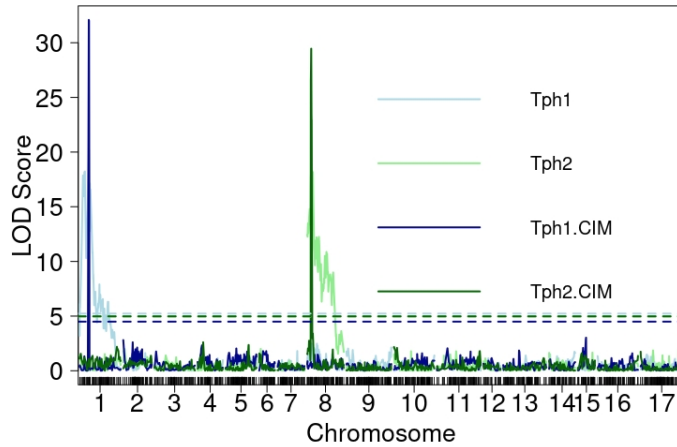


**VK876
x
VK101**

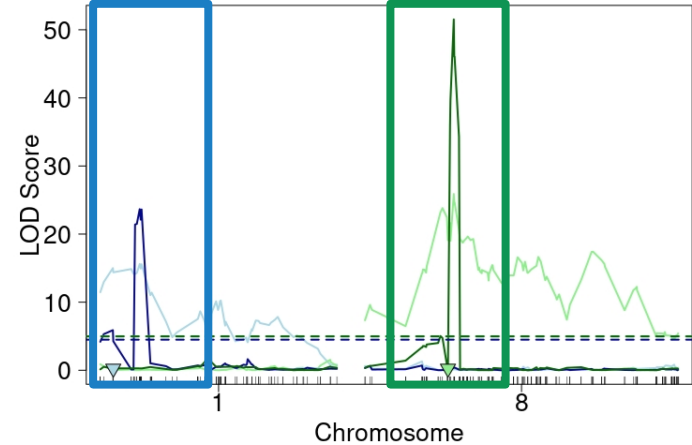
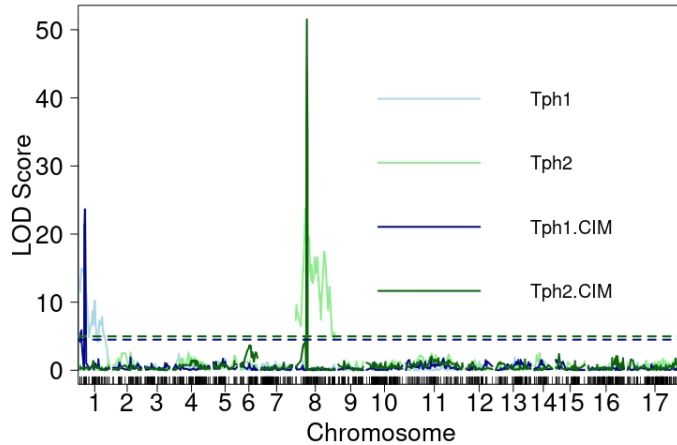


Tocopherol composition mapping (qualitative approach)

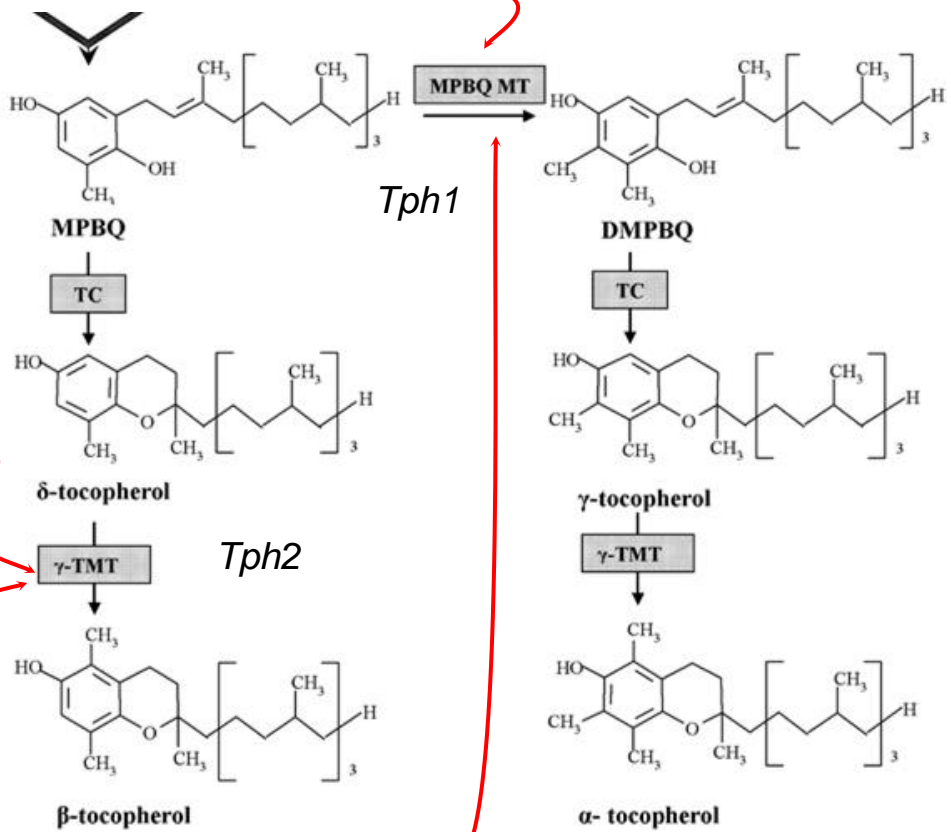
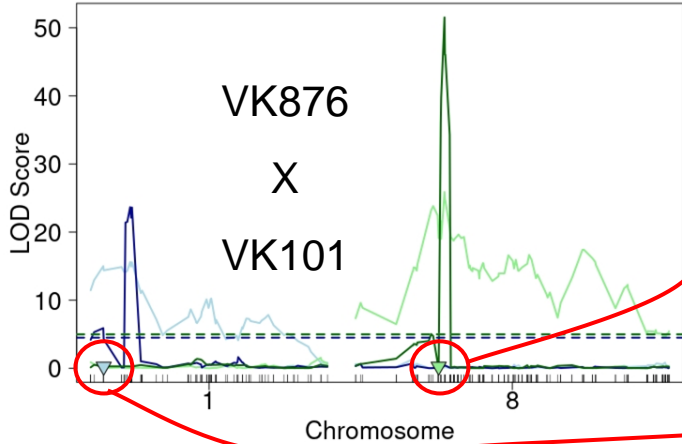
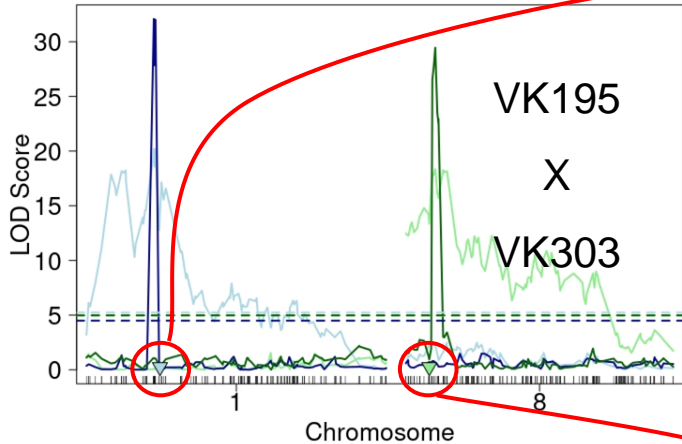
VK195xVK303



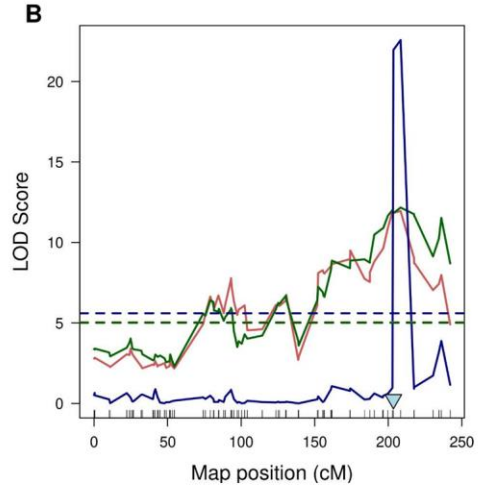
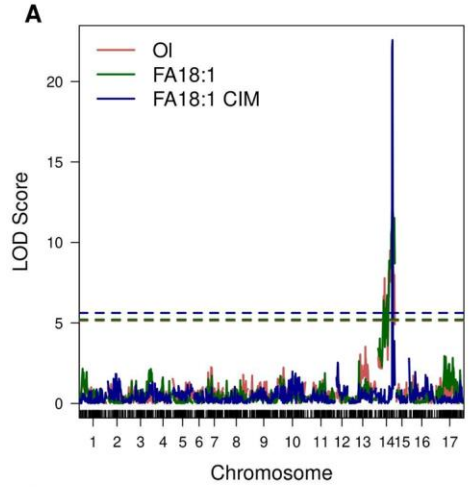
VK876xVK101



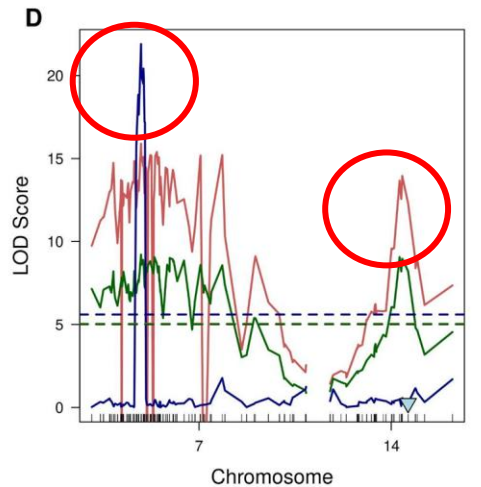
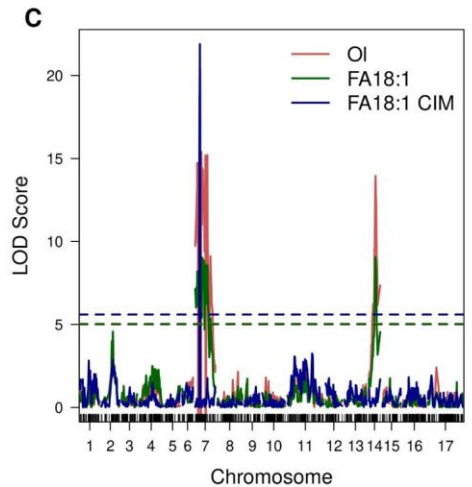
Tocopherol composition mapping (qualitative approach)



Mapping oleic acid content

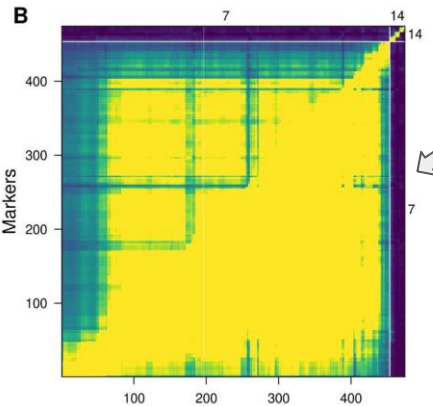
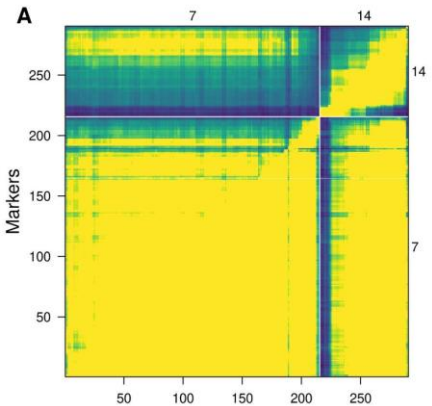
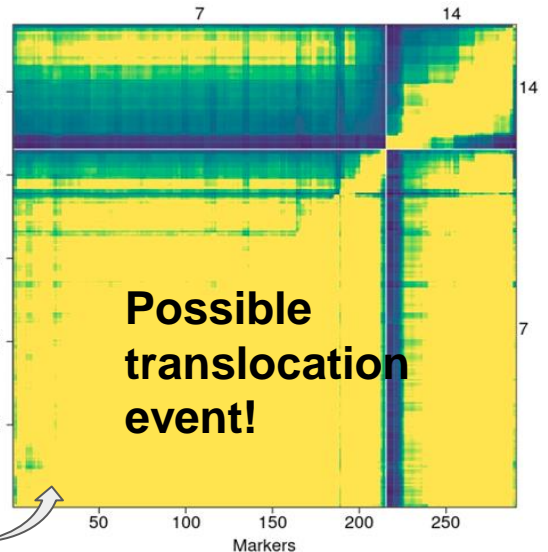
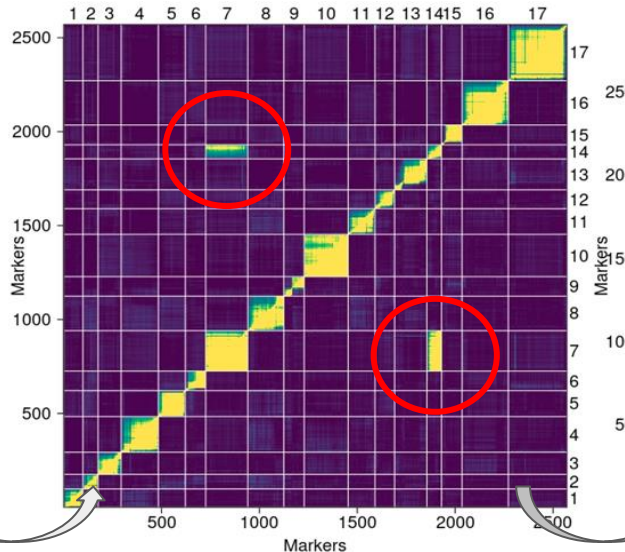
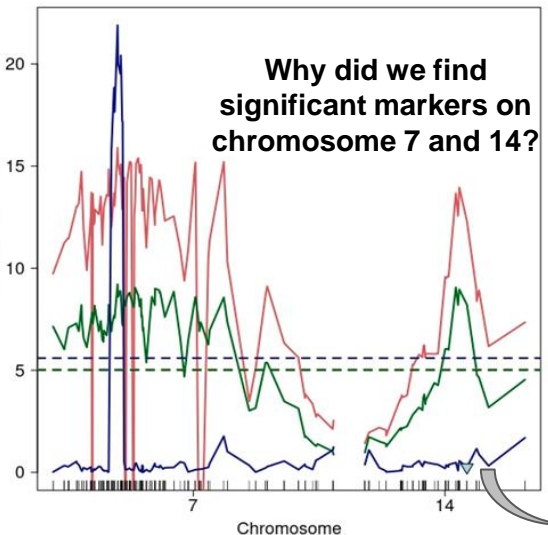


VK195xVK303



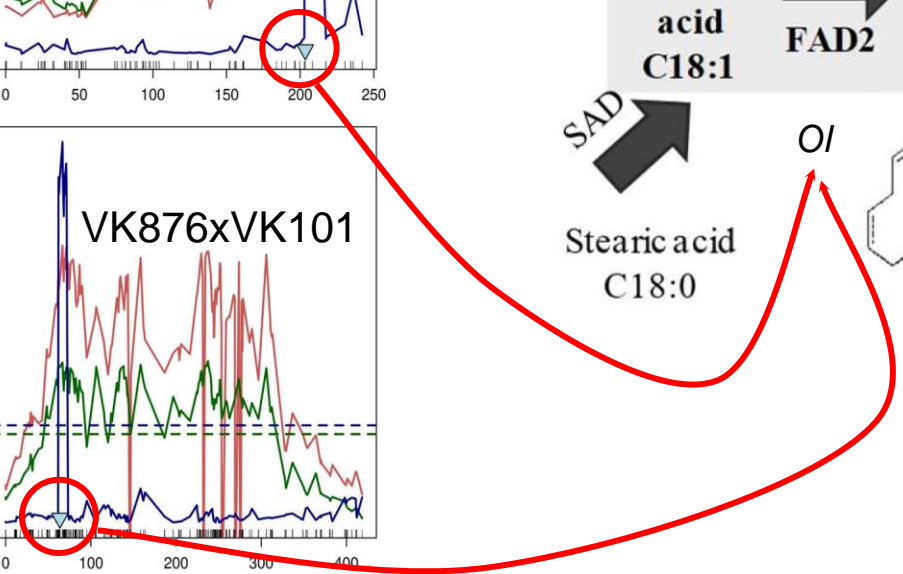
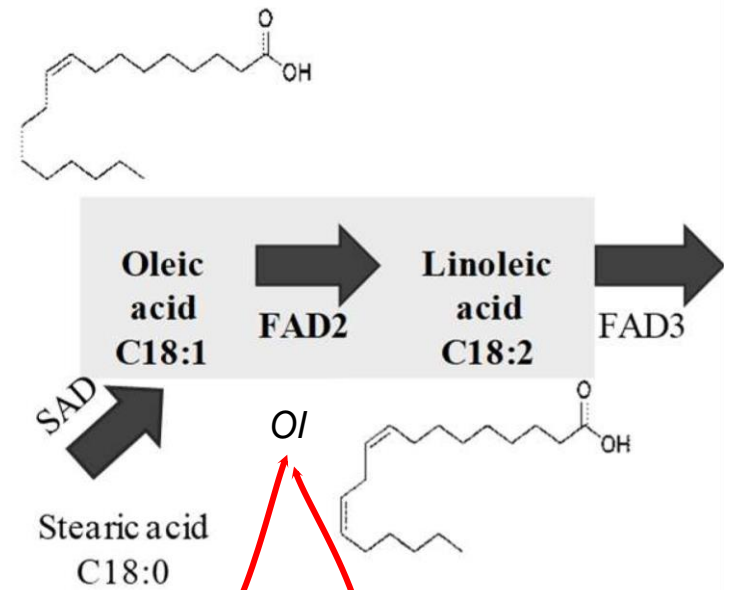
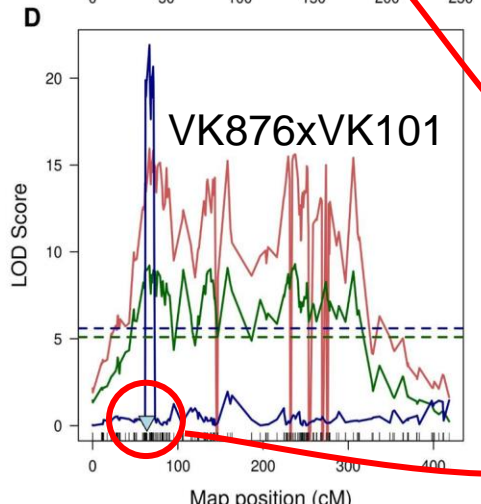
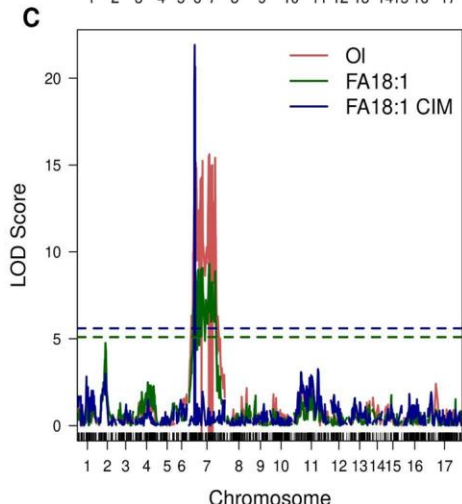
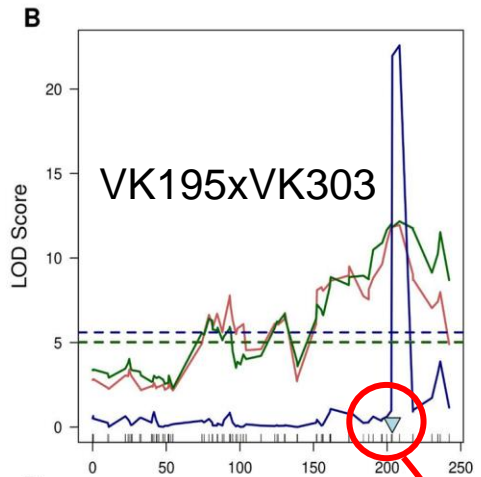
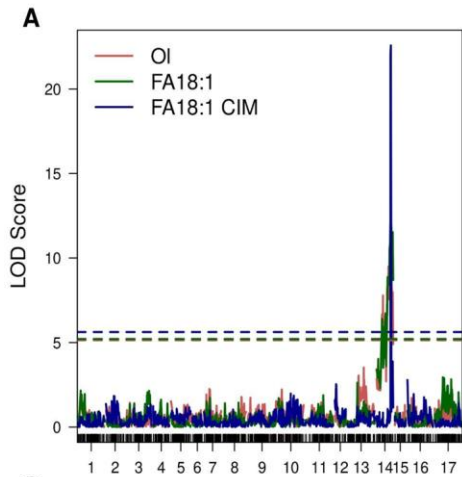
VK876xVK101

Issue with chromosomes 7 and 14, map reconstruction



Genetic map reconstruction for chromosomes 7 and 14

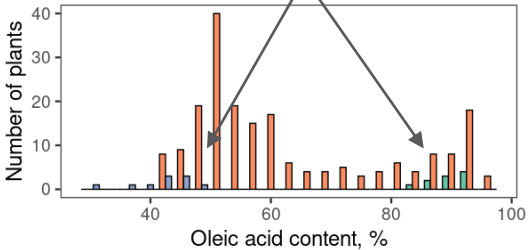
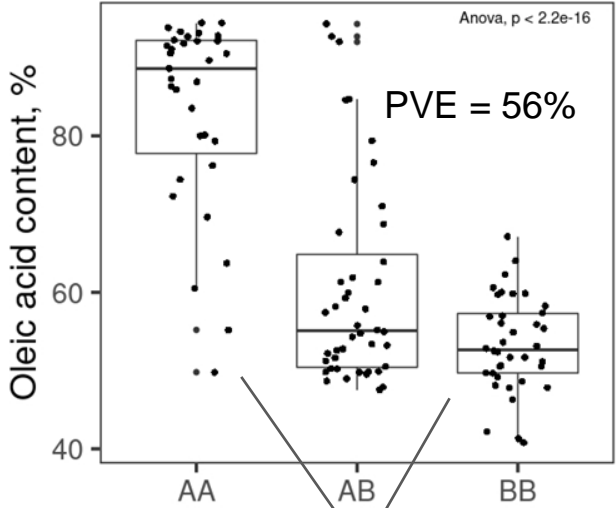
Mapping oleic acid content (reconstructed map)



Genotype effects for oleic acid content

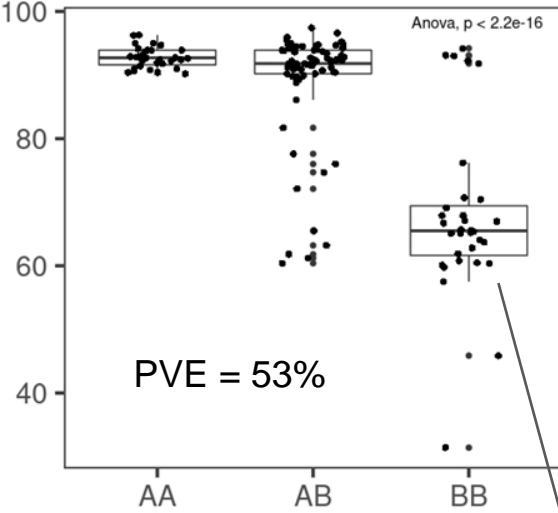
VK195xVK303

S14_137821336

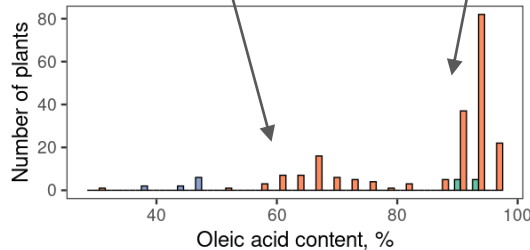
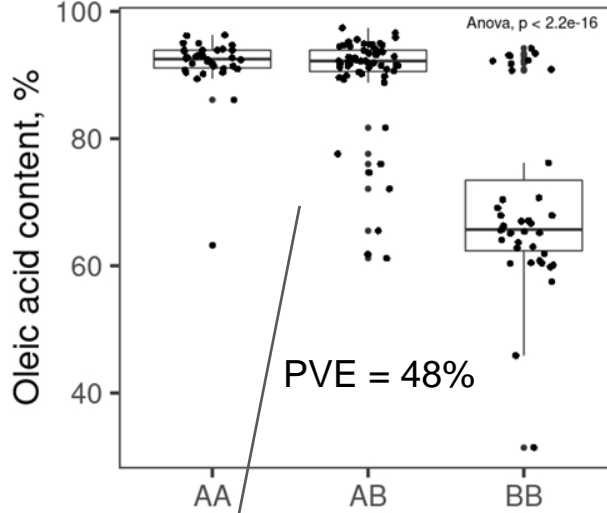


VK876xVK101

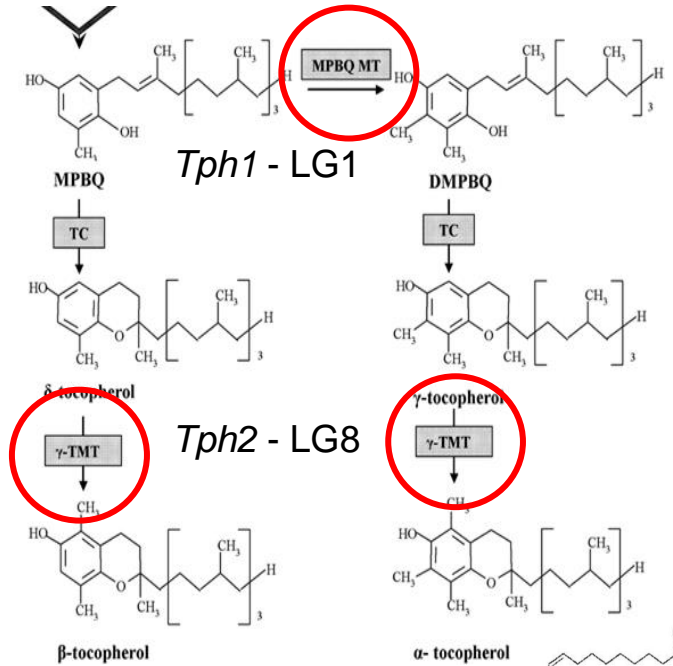
S7_37821562



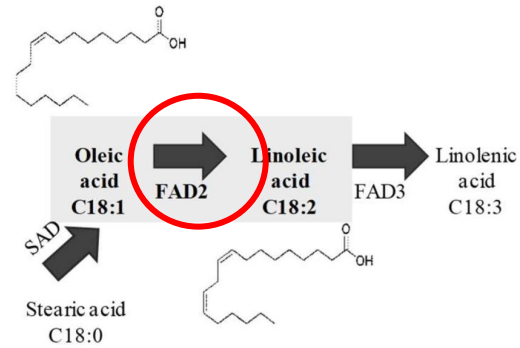
S14_78886306



Conclusions



OI - LG14 or LG7



- Two loci of major effects were identified on chromosomes 1 and 8 controlling tocopherol composition
- One locus controlling oleic acid content was identified on chromosome 14 and 7
- No minor effect loci were found for studied traits



G3, 2022, 12(4), jka036
<https://doi.org/10.1093/g3journal/jka036>
 Advance Access Publication Date: 12 February 2022
 Investigation

Genetic mapping of loci involved in oil tocopherol composition control in Russian sunflower (*Helianthus annuus* L.) lines

Rim Gubaev ^{1,2*}, Stepan Boldyrev ^{1,2}, Elena Martynova ¹, Alina Chernova ^{1,2}, Tatyana Kovalenko ^{1,3}, Tatyana Peretyagina ², Svetlana Goryunova ^{1,4,5}, Denis Goryunov ^{1,4,5}, Zhanna Mukhina ^{1,6}, Cecile Ben ^{1,7}, Laurent Gentzittel ¹, Philipp Khaitovich ¹, Yakov Demurin ^{1,3*}

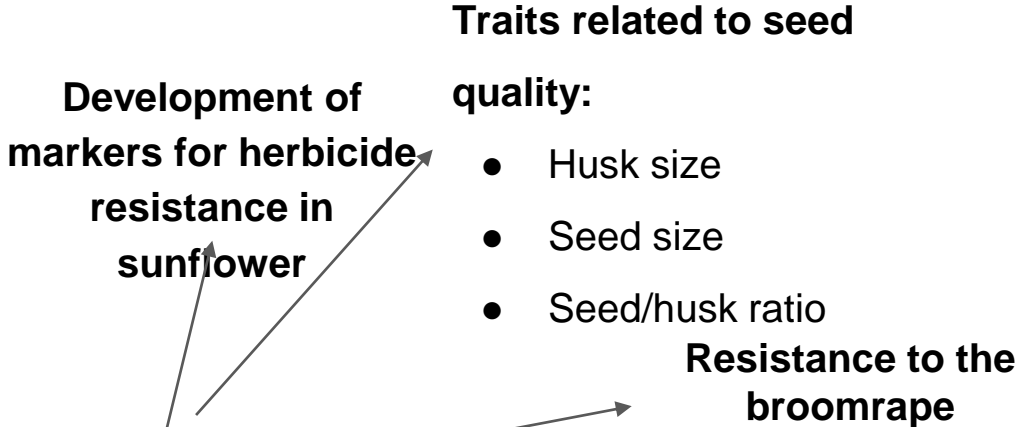
¹Skolkovo Institute of Science and Technology, Moscow 121205, Russia,
²LLC "Oil Gene", Moscow 121205, Russia,
³Pustovoit All-Russia Research Institute of Oil Crops, Krasnodar 350038, Russia,
⁴Russian Potato Research Center, Kraskovo 140051, Russia,
⁵Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow 119333, Russia,
⁶Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 119992, Russia,
⁷All-Russia Rice Research Institute, Krasnodar 350921, Russia.

*Corresponding authors: Digital Agriculture Laboratory, Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, bld. 1, Moscow 121205, Russia.
 Email: rim.gubaev@skoltech.ru; Laboratory of Genetics, Pustovoit All-Russia Research Institute of Oil Crops (VNIMK), Filatova St. 17, Krasnodar 350038, Russia.
 Email: yakdemurin@yandex.ru.

Abstract

Tocopherols are antioxidants that preserve oil lipids against oxidation and serve as a natural source of vitamin E in the human diet. Compared with other major oilseeds like rapeseed and soybean, sunflower (*Helianthus annuus* L.) exhibits low phenotypic diversity of tocopherol composition, both in wild and cultivated accessions from germplasm collections. Two major mutations that alter tocopherol composition were identified in genetic collections, and several studies suggested additional loci controlling tocopherol composition, with their expression possibly depending on the genetic background. In the present study, we performed QTL mapping of tocopherol composition in two independent F2 crosses between lines with contrasting tocopherol composition from the Pustovoit All-Russia Research Institute of Oil Crops (VNIMK) collection. We used genotyping-by-sequencing (GBS) to construct single nucleotide polymorphism-based genetic maps, and performed QTL mapping using quantitative and qualitative encoding for phenotypic traits. Our results support the notion that the tocopherol composition in the assessed crosses is controlled by two loci. We additionally selected and validated two single nucleotide polymorphism markers for each cross which could be used for marker-assisted selection.

Russian sunflower improvement initiative - a bigger picture



Core: **Skoltech**
Skolkovo Institute of Science and Technology



OilGene

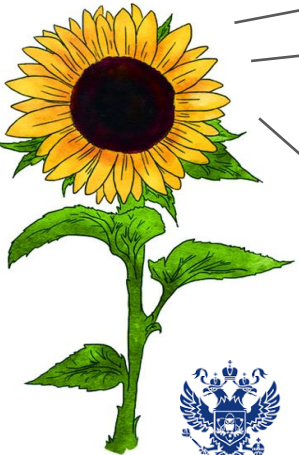


ВНИИМК

Collaborators:



AGROPLAZMA
BREEDING AND SEED PRODUCTION COMPANY



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ



USC University of Southern California

Core team and mentors



**Alina
Chernova**

Sequencing



**Elena
Martynova**

Molecular
biology



**Stepan
Boldyrev**

Plant /
molecular
genetics



**Rim
Gubaev**

Data analysis /
Bioinformatics



Prof. Philipp
Khaitovich, Skoltech



Prof. Yakov
Demurin, VNIIMK



Prof. Sergey
Nuzhdin,
USC

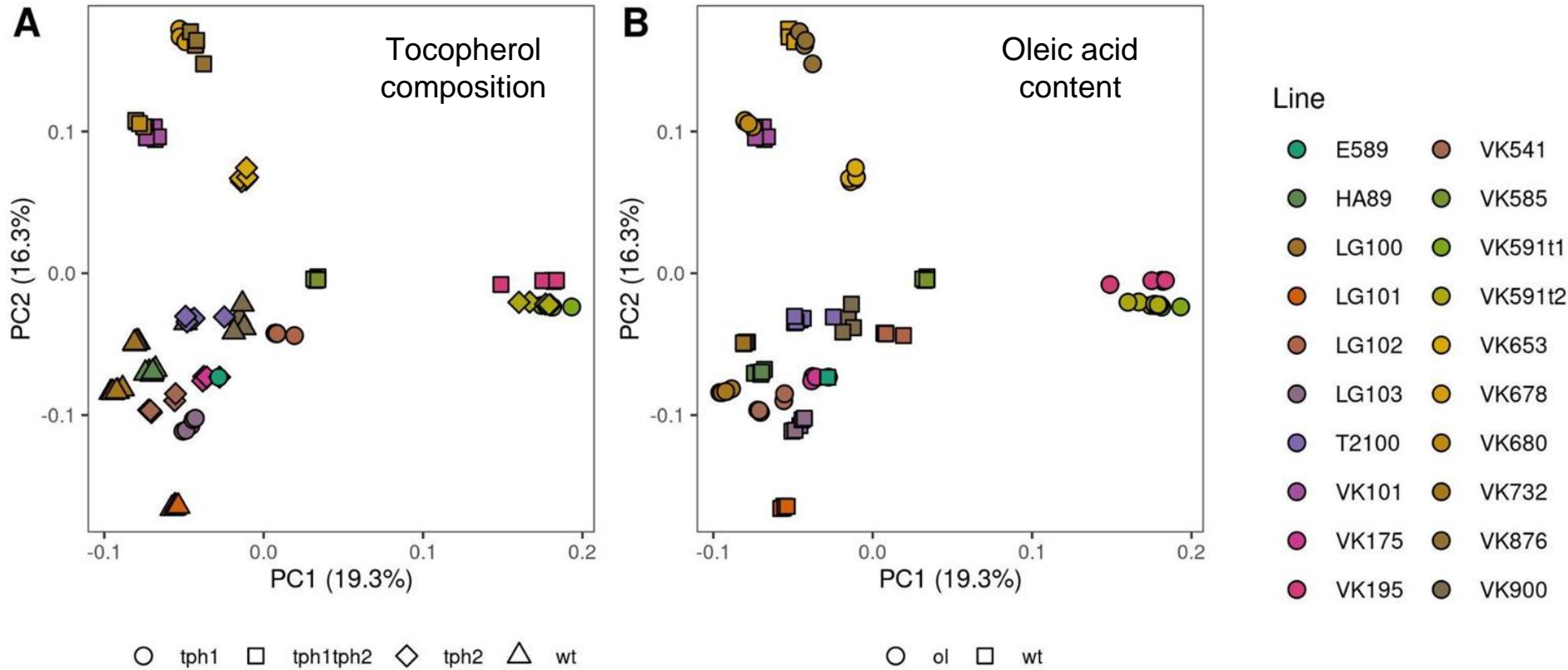


Prof. Laurent
Gentzbittel,
Skoltech



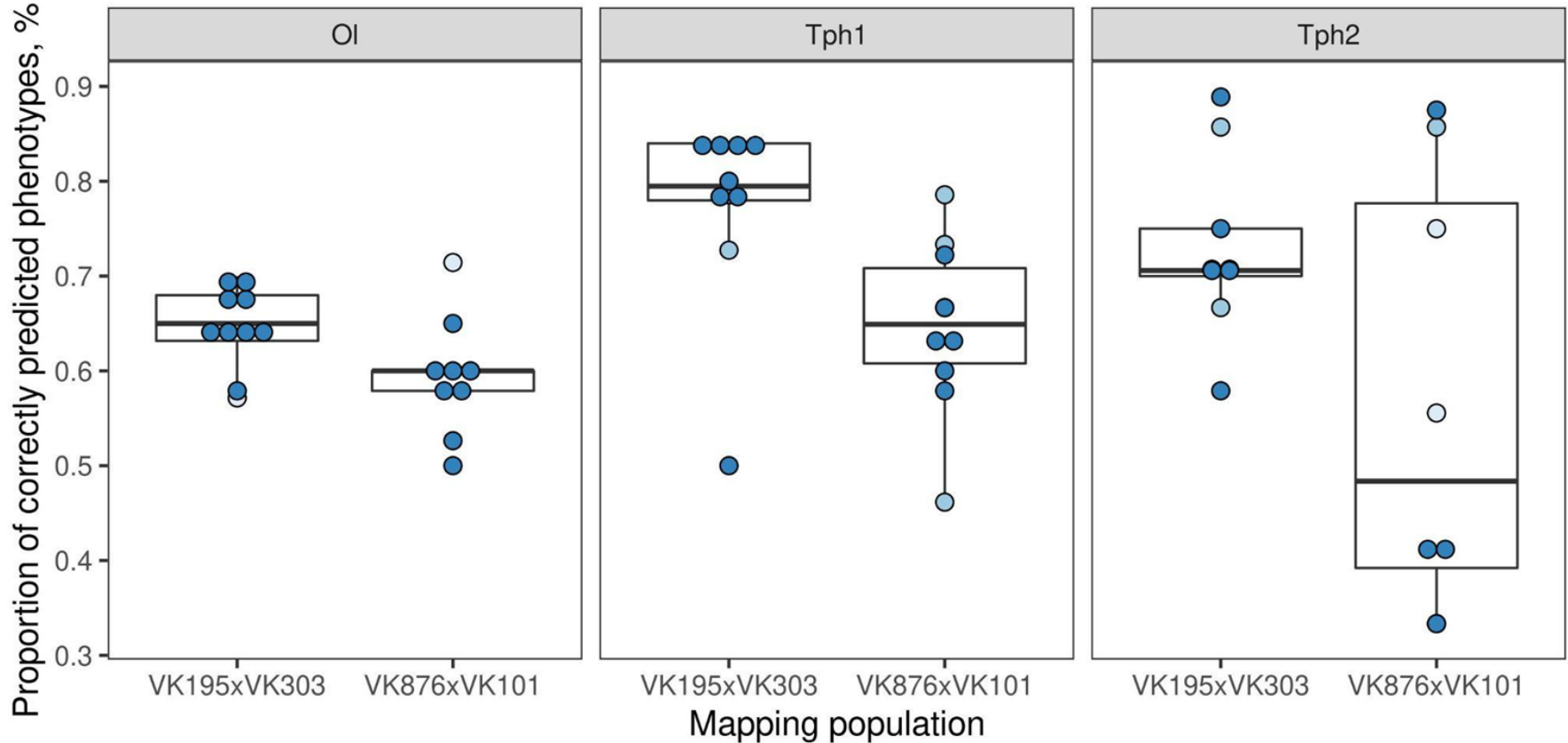
Assoc. Prof.
Cecile Ben,
Skoltech

Marker using independent sample

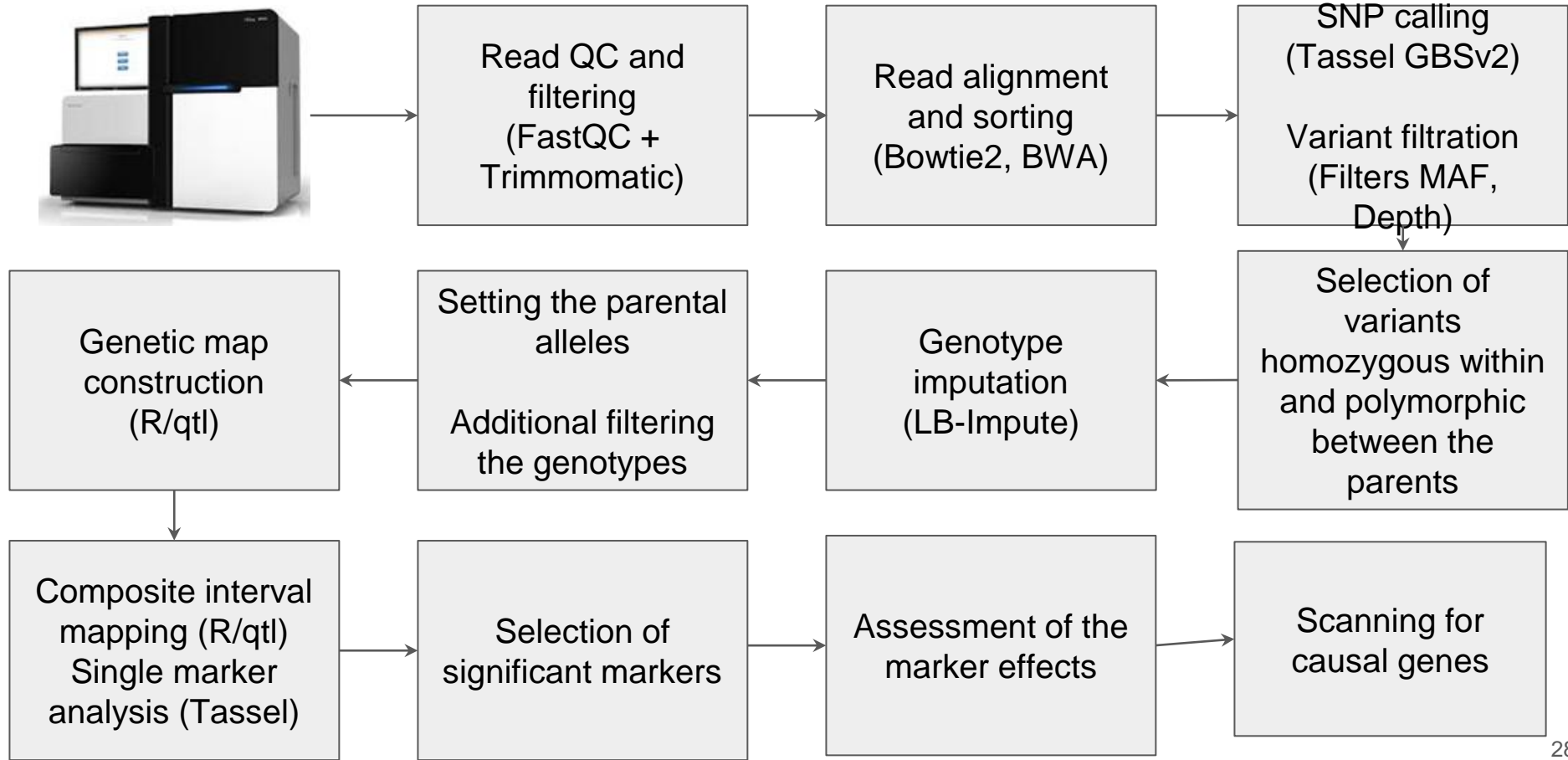


Population structure and respective phenotypes of sample. Color indicates line, shape corresponds to the phenotype.

Marker validation on the independent plant sample

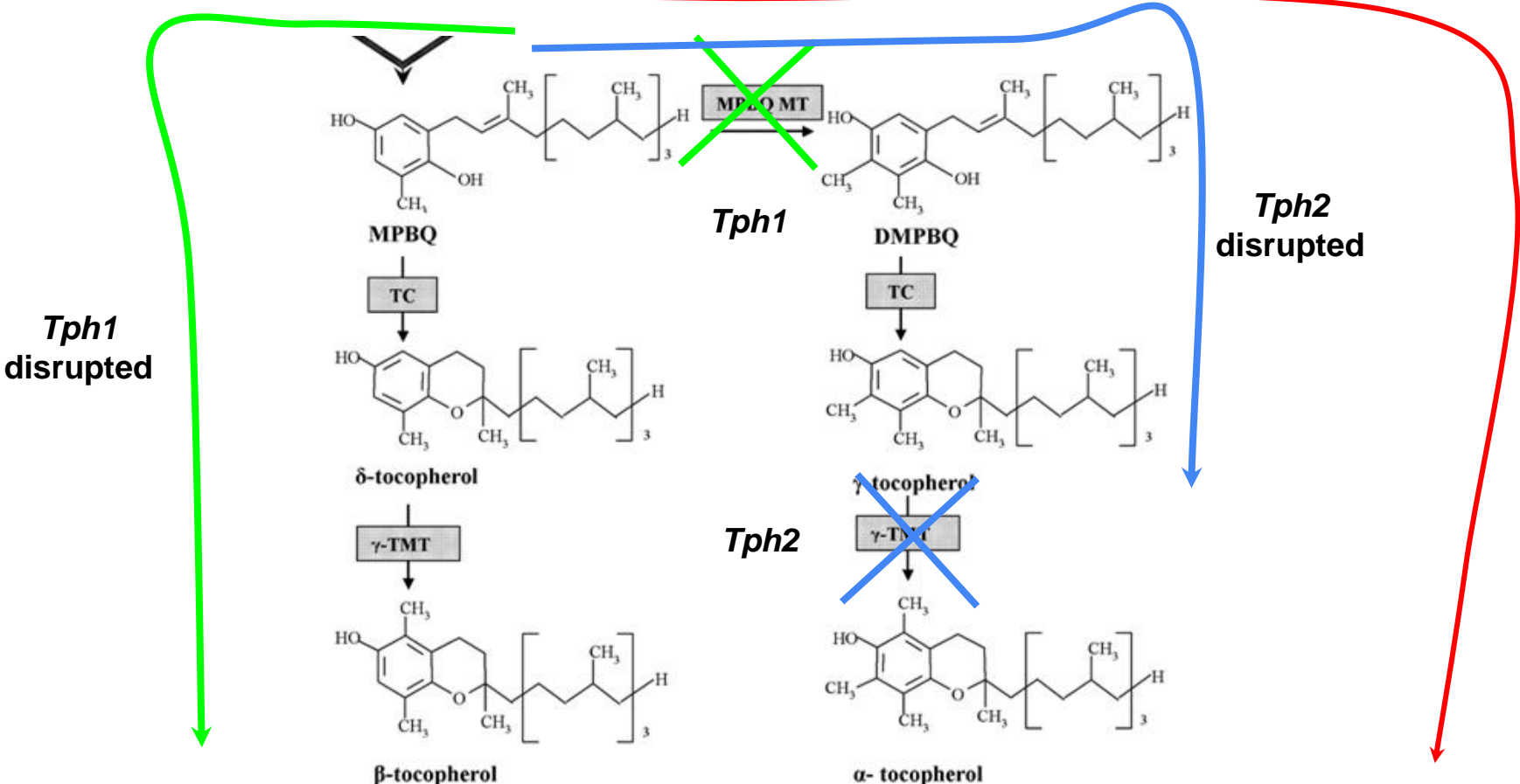


Analysis pipeline

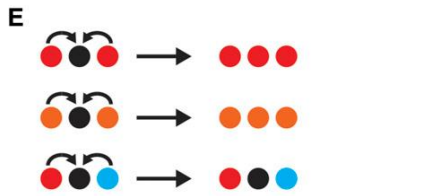
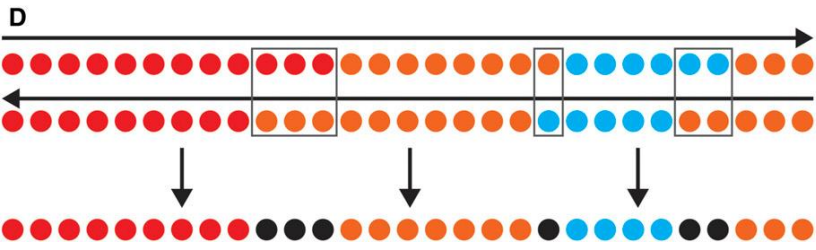
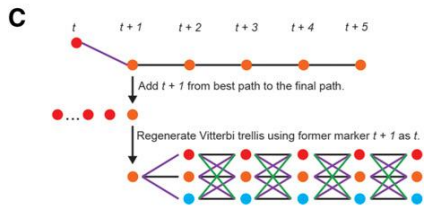
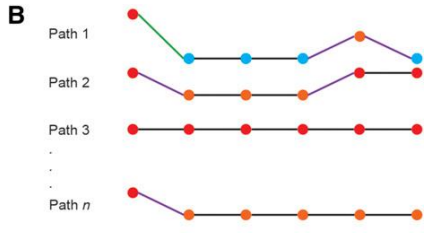
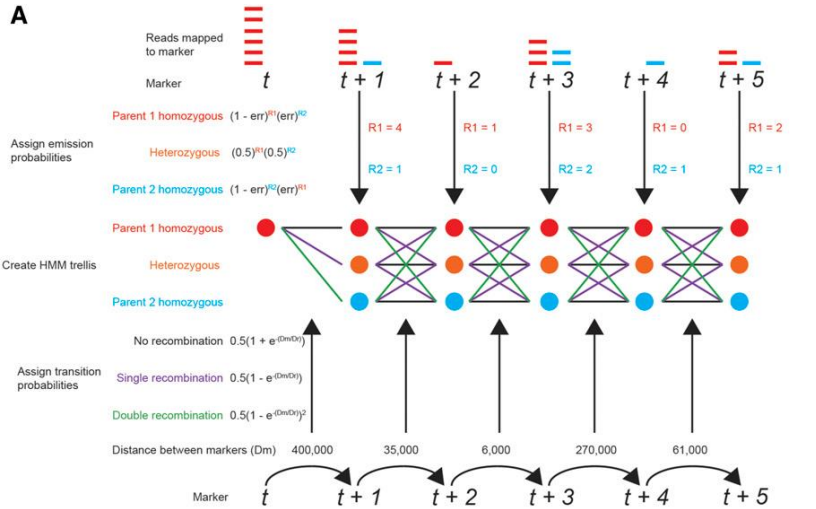


Effect of *Tph1* and *Tph2* on beta and delta tocopherol content

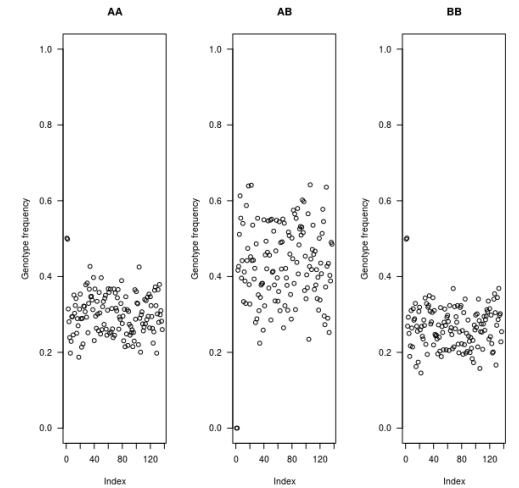
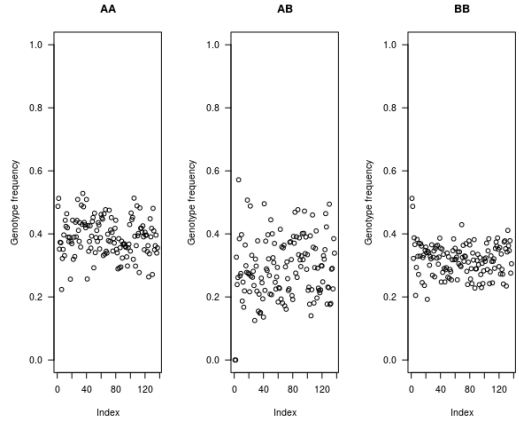
Normal path



Imputation with LB-Impute



Imputation results

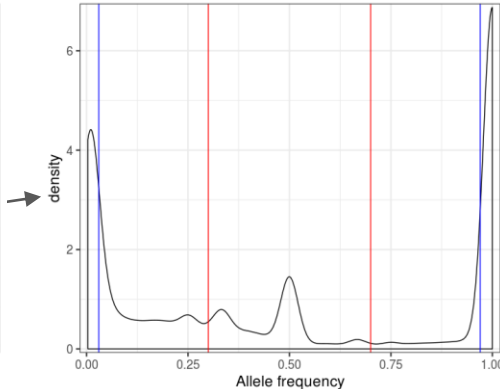
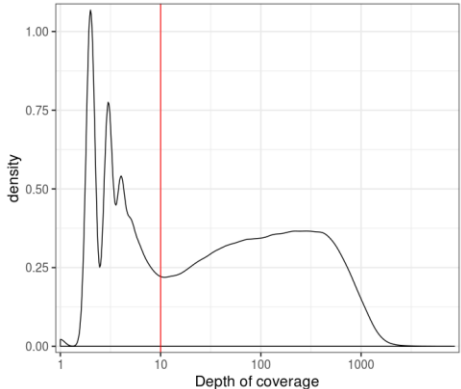


Genotyping data filtering

Selection of variants homozygous within and polymorphic between the parents

Variant depth filter

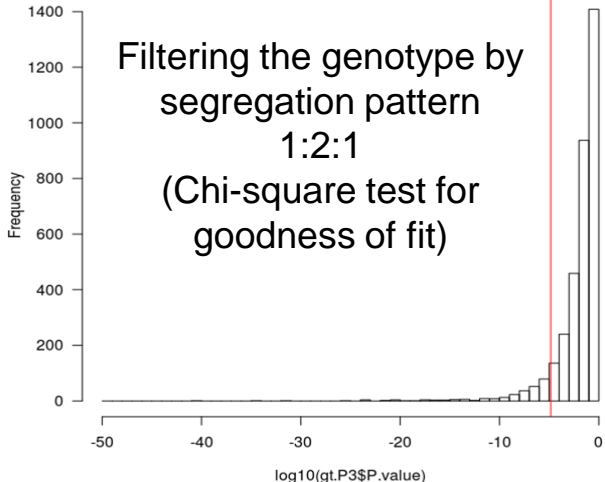
Minor allele frequency



	9897: 68232109	9898: 68232115	9899: 70677281	9900: 70677324	9901: 70716718	9902: 70813899	9903: 70813996	9904: 70991646	9905: 70991697	9906: 70991703	9907: 71056732	9908: 72216866	9909: 72216882	9910: 73103379	9911: 73103411	9912: 73103551	9913: 73103567	9914: 73103582	9915: 73103609	9916: 74956914	9917: 74956919	9918: 75709679	9919: 75709692	9920: 75709723	9921: 76204476	9922: 76204485	9923: 76204495	9924: 76204557	9925: 76284728	9926: 76284736	9927: 76284741	9928: 76284758	9929: 76284817	9930: 76563131							
BK195	N	N	N	N	A	N	N	N	N	N	N	N	N	N	N	N	N	N	N	G	C	T	G	G	T	A	G	N	N	N	N	N	N	N	N	N	N	N			
BK303	N	N	N	N	T	N	N	N	N	N	N	N	N	N	T	N	N	N	N	G	C	T	G	G	T	A	G	N	N	N	N	N	N	N	N	N	N	N			
P3 sts10	N	N	N	N	N	N	N	N	N	N	N	N	N	A	C	N	N	N	N	G	C	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N			
P3 sts100	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N		
P3 sts101	N	N	N	N	T	C	C	C	G	A	C	G	T	N	N	N	N	N	N	N	N	N	N	N	N	N	T	A	G	C	N	N	N	N	N	N	N	N	N		
P3 sts107	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	G	T	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N		
P3 sts108	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	C	A	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N		
P3 sts109	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	T	A	G	C	N	N	N	N	N	N	N	N	N	N	
P3 sts112	N	N	N	T	T	A	T	T	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	T	G	G	T	A	G	C	N	N	N	N	N	N	N	N	
P3 sts113	T	A	N	N	T	C	C	T	A	T	T	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	
P3 sts114	N	N	N	N	N	N	C	C	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
P3 sts115	N	N	T	T	A	N	N	C	G	A	C	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
P3 sts116	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Number of markers	Before filtering	After filtering
VK195 × VK303	119,964	3566
VK876 × VK101	182,749	5979

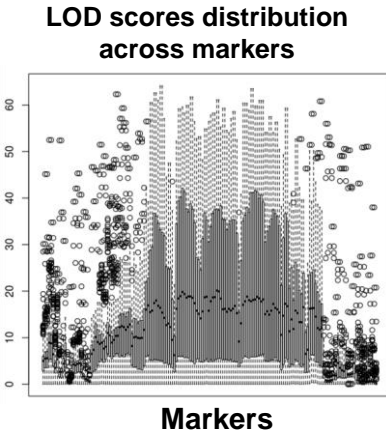
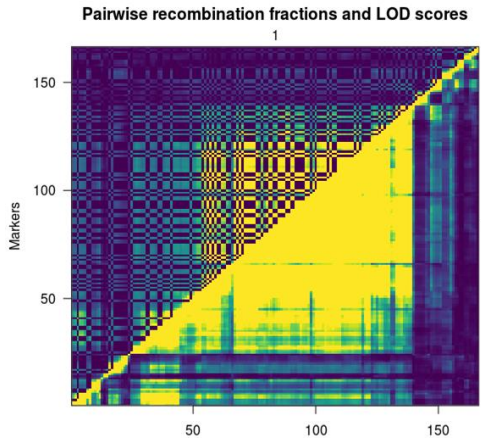
Filtering the genotype by segregation pattern 1:2:1 (Chi-square test for goodness of fit)



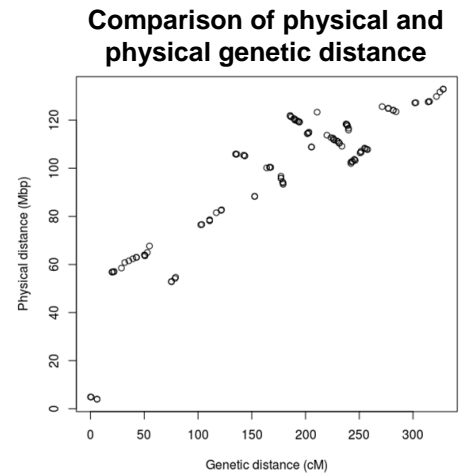
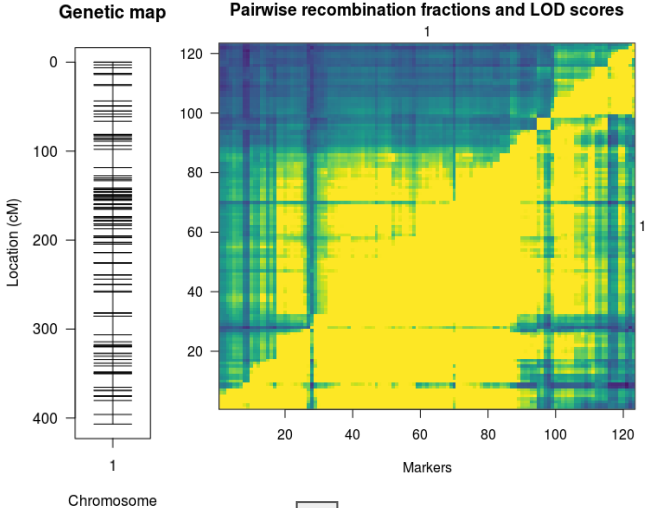
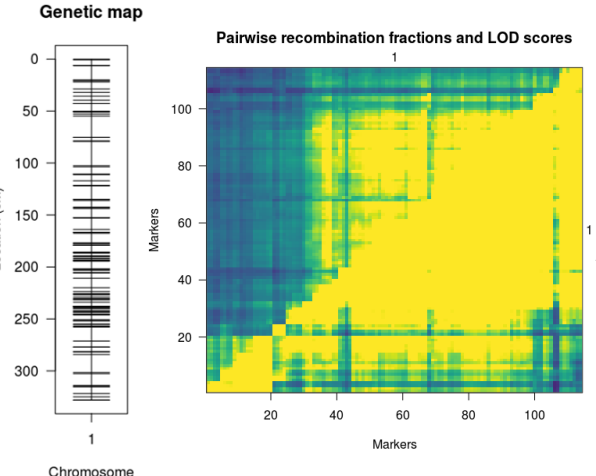
Individuals with missing genotypes (> 30%)

Markers that are genotyped in less than 70% of individuals

Genetic map construction

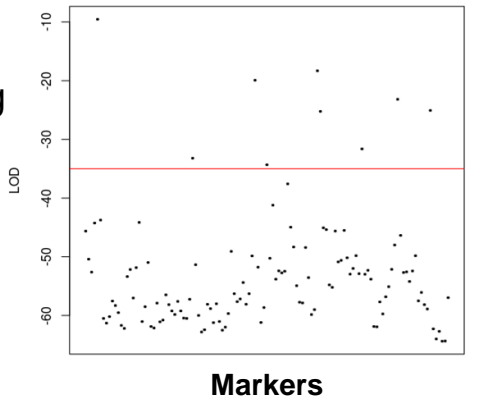


Ordering

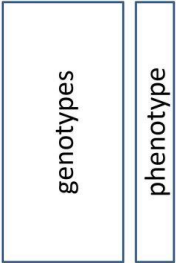
Reordering

Dropping single markers



LOD Threshold calculation

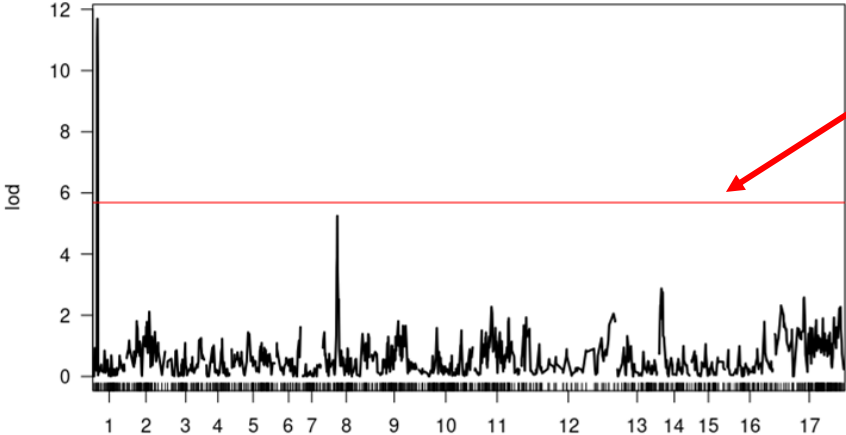
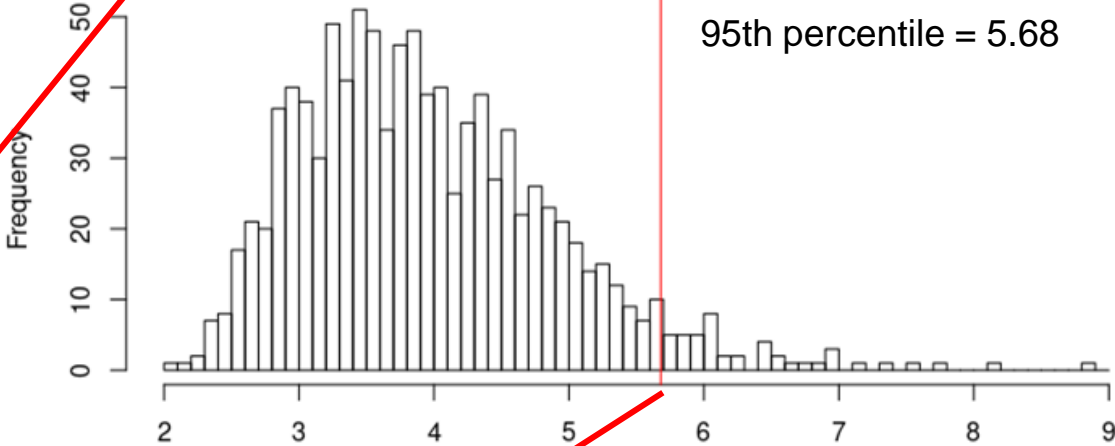
Single trait permutation schema



→ LOD over genome → max LOD

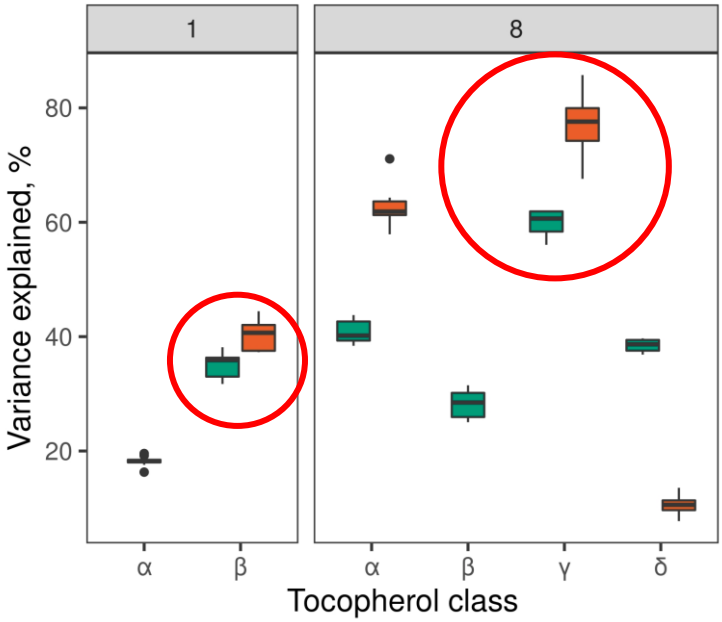
- 1. shuffle phenotypes to break QTL
- 2. repeat 1000 times and summarize

The distribution of max LOD scores across permutations

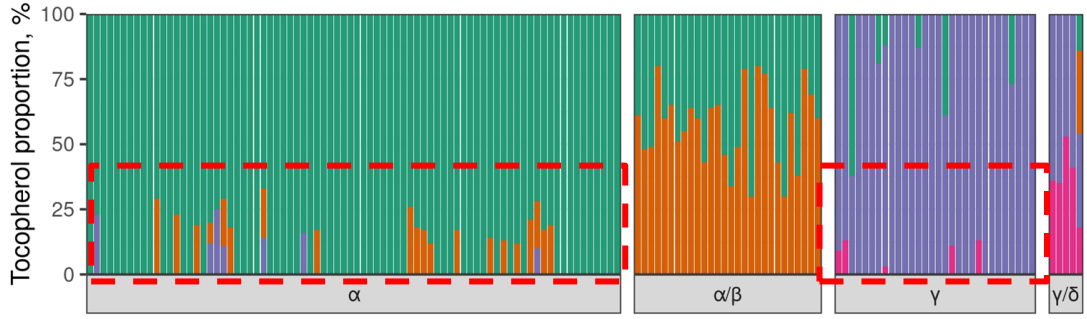
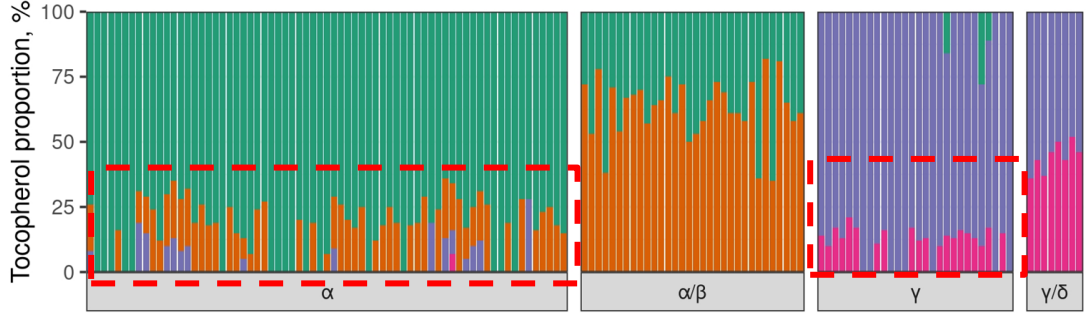


Variance explained by top 10 significant markers

Chromosomes

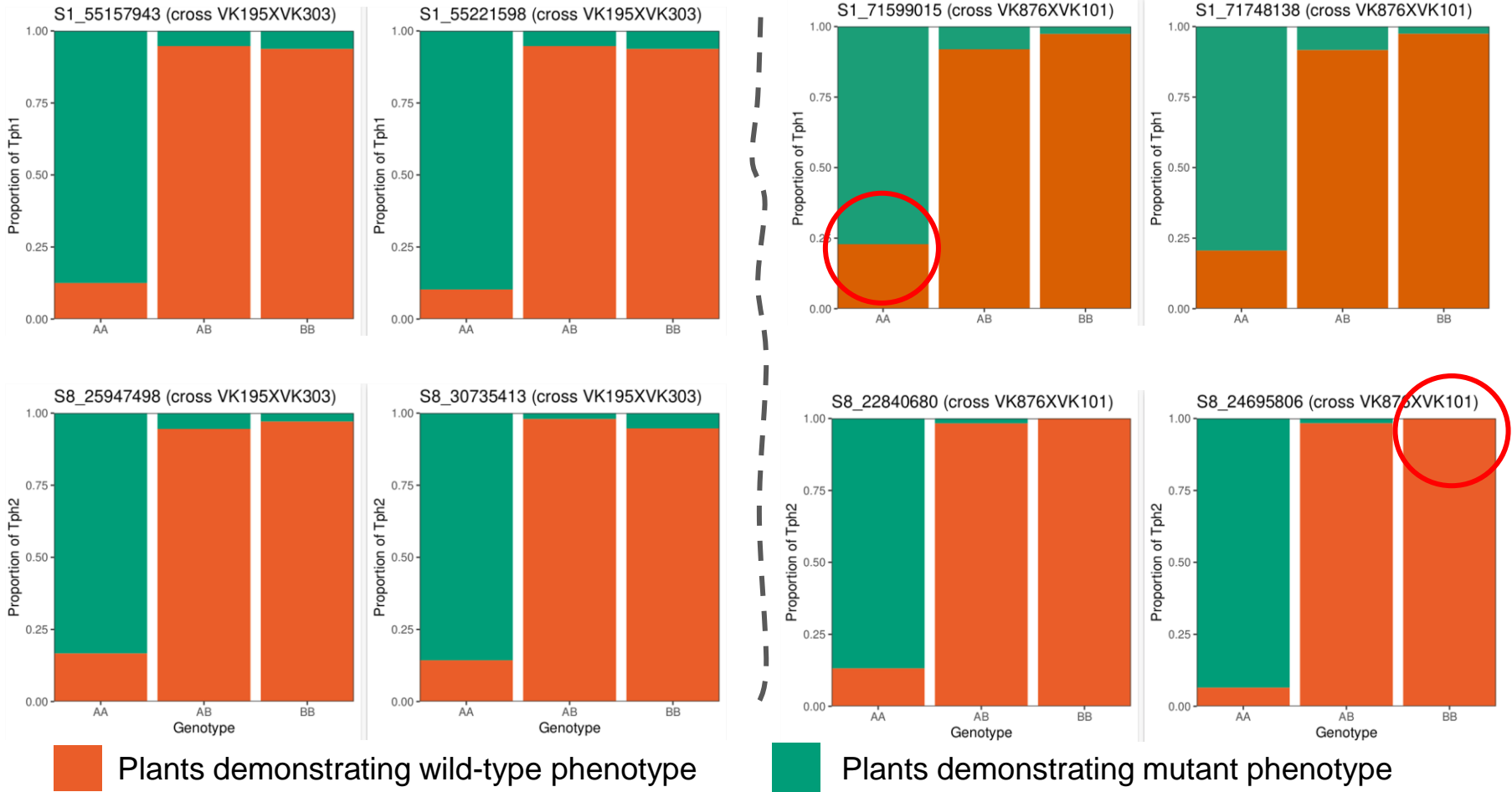


Cross █ BK195XBK303 █ BK876XBK101



Tocopherol class █ alpha █ beta █ gamma █ delta

Genotype effects (qualitative approach)



Effects of the markers for cross VK876xVK101

