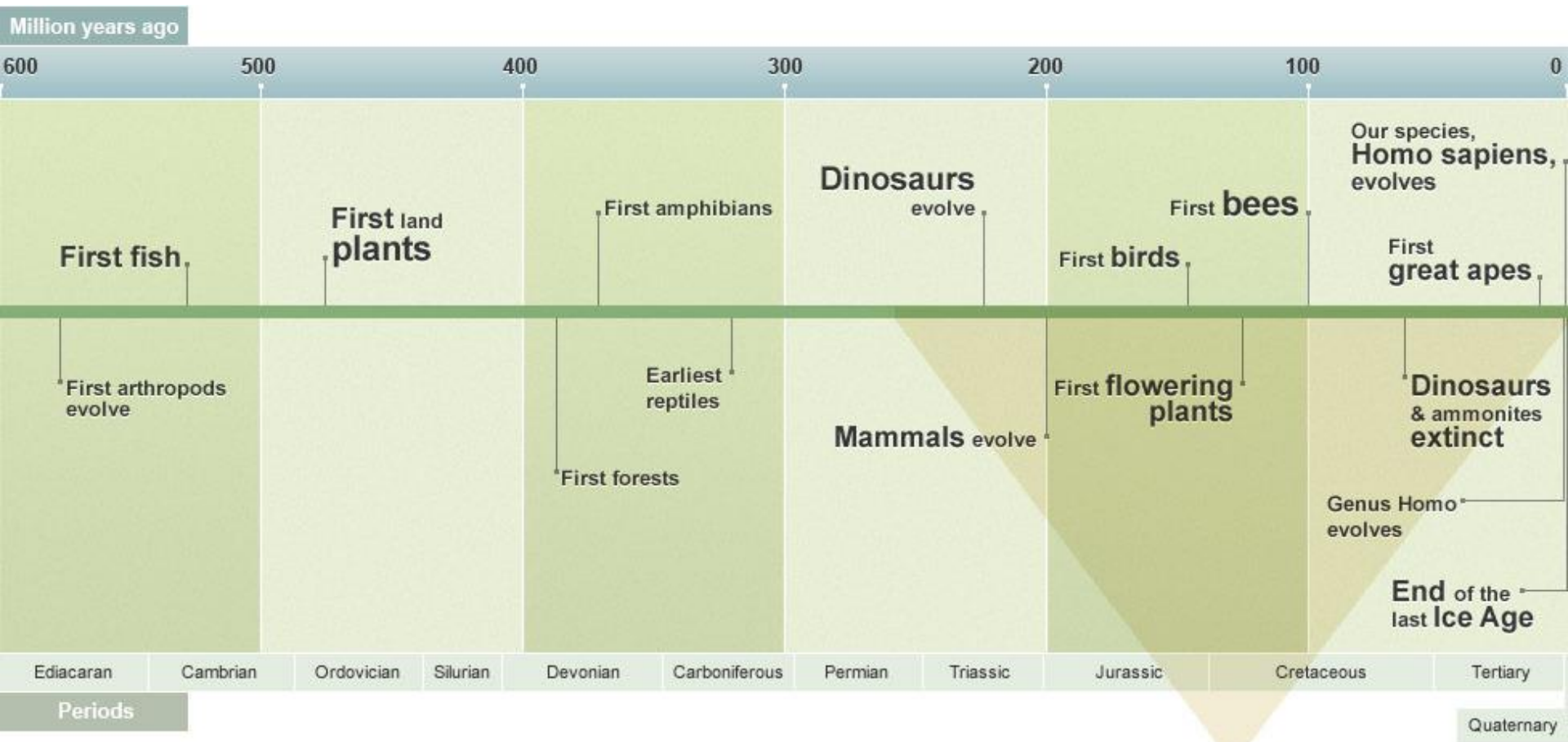


The Amborella genome and the evolution of flowering plants.

Heleri Kirsip

19.11.2014

Õistaimede teke

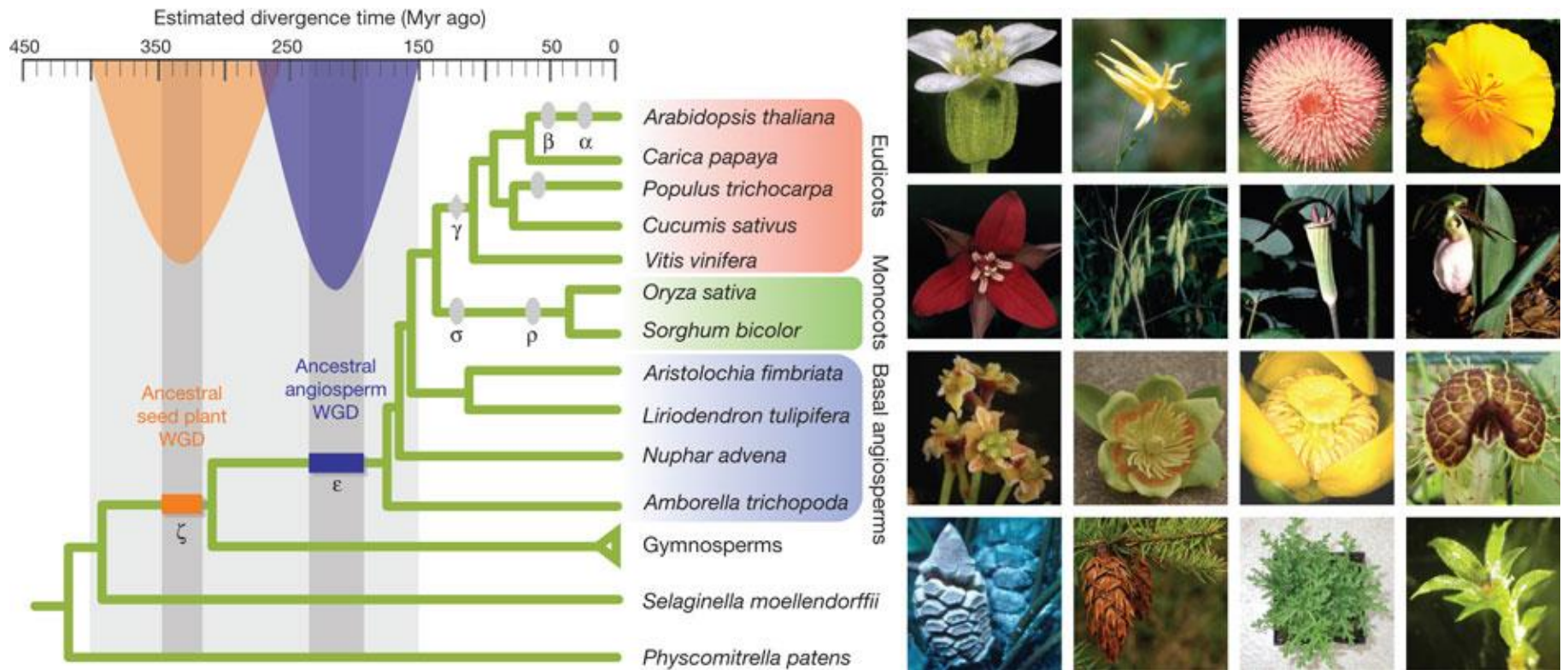


http://www.bbc.co.uk/nature/history_of_the_earth

Õistaimede tähtsus

- Rohkem kui 350 000 liiki.
 - Maismaa
 - Vesikeskkond
- Inimese toidu allikaks.
- Globaalne fotosüntees.
- C sidumine.

Õistaimede süstemaatika



YN Jiao et al. *Nature* 000, 1-4 (2011) doi:10.1038/nature09916

Amborella trichopoda

- *Amborellaceae* sugukonna ainuke esindaja
- Põõsastaim
- New Caledonia, märjad troopilised metsad



http://science.psu.edu/alert/photos/research-photos/biology/dePamphilis_femaleflowersideview_12-2013.jpg

Amborella Genome Project

Eesmärgid:

- *Amborella* genoomi sekveneerimine ja uurimine.
- *Amborella* ja teiste õistaimede genoomide ja nendes sisalduvate elementide võrdlemine ning ansestraalse genoomi/järjestuste leidmine.
- *Amborella* populatsioonigeneetika ning diversiteedi uurimine.

DNA sekveneerimine

- NG whole genome shotgun.
 - Roche 454
 - Illumina
 - Sanger BAC-end sequences

Table S1.
Sequenced genome library statistics.

Library Type (Insert Size)	Sequencing Information	Raw Data (reads)		Raw Data Size		Mean Length		Theoretical Depth of Coverage	
		Raw	Clean	Raw	Clean	Raw	Clean	Raw	Clean
454 Titanium	28.5 Plates	38.8 M	27.6 M	14.2 Gb	10.5 Gb	365 bp	381 bp	18.9X	13.9X
454 Titanium plus	17 Plates	22.4 M	18.3 M	12 Gb	10.2 Gb	533 bp	557 bp	16X	13.6X
454 paired-ends (11 Kb)	5 Plates	6.1 M	1.92 M	2.2 Gb	0.699 Gb	348 bp	364 bp	2.9X	0.9X
Illumina mate-pairs (3 Kb)	Full lane HiSeq2000	195 M	17.7 M	19.7 Gb	1.68 Gb	101 bp	95 bp	26.2X	2.2X
Sanger BAC ends (130 Kb)	5.2X Library	69466	63924	48 Mb	44 Mb	695 bp	689 bp	0.06X	0.06X
Total		262 M	66 M	48 Gb	23 Gb			64.1X	30.7X

- Eemaldati kontaminantide (putukad, bakterid, seened, inimeste), kloroplastide ja mitokondri järjestused.
- Lisaks eemaldati PCR vigadest tekkinud duplikatsioonid ja kimäärsed riidid.
- Lisaks eemaldati lühikesed ja väikese kattuvusega järjestused.

Table S2.

Genome library contaminant statistics.

Library Type (Insert Size)	Mitochondrial	Chloroplast	Junction/Inward Facing Reads	Artificial Duplicates	Short	Total
454 Titanium	12.01%	2.86%	N/A	12.94%	5.44%	33.25%
454 Titanium plus	5.02%	.99%	N/A	10.36%	3.77%	20.14%
454 paired-ends (11 Kb)	7.34%	1.88%	N/A	38.94%	7.34%	55.5%
Illumina mate-pairs (3 Kb)	3.18%	0.53%	57.56%	6%	25.26	92.53%
Sanger BAC ends (130 Kb)	1.09%	2.89%	N/A	N/A	N/A	3.98%

Genoomi hinnangud

- Läbivoolu tsütomeetria hindab genoomi suuruseks 870 Mb.
 - Praegu suutsid kokku panna 82% *Amborella* genoomist.
- K-mer´ide sageduste hindamisel saaks *Amborella* genoomi suuruseks 793 Mb.
- Erinevate meetodite baasil hinnatakse *Amborella* genoomi suuruseks 748 Mb.
 - Kõrge-kvaliteetsed järjestused (keskmise 31x kattuvusega) katavad üle 94% genoomist.

Assambleerimine

- Assambleerimiseks kasutati alla 50% algsest DNA kogusest.
 - 5745 scaffolds
 - Keskmiseks pikkuseks 123 kb
- Kokku 706 Mb (94%) assambleeritud järjestusi.
 - Pikim 16 Mb
 - 155 scaffoldi pikemad kui 1.1 Mb (N90)
 - N50 4.9 Mb

Table S5.Contig and Scaffold assembly metrics for the *Amborella* assembly version 1.

	Version 1.0		
	Contigs	Scaffolds	After GB Super-scaffolding
Number	43,234	5,745	5,635
Total Size (bp)	668,257,121	706,332,648	712,382,144
Genome Covered (%)	88.9	93.9	94.7
Largest Size (bp)	287,935	15,980,527	23,365,040
Mean Size (bp)	15,456	122,947	126,409
N10 Size (bp)	85,018	9,414,115	17,768,939
N10 Count	579	7	4
N25 Size (bp)	52,723	6,874,971	14,187,169
N25 Count	2,129	20	11
N50 Size (bp)	29,456	4,927,027	9,308,884
N50 Count	6,448	50	26
N75 Size (bp)	14,812	2,820,768	6,368,207
N75 Count	14,404	97	49
N90 Size (bp)	7,108	1,154,593	2,935,351
N90 Count	23,962	155	70

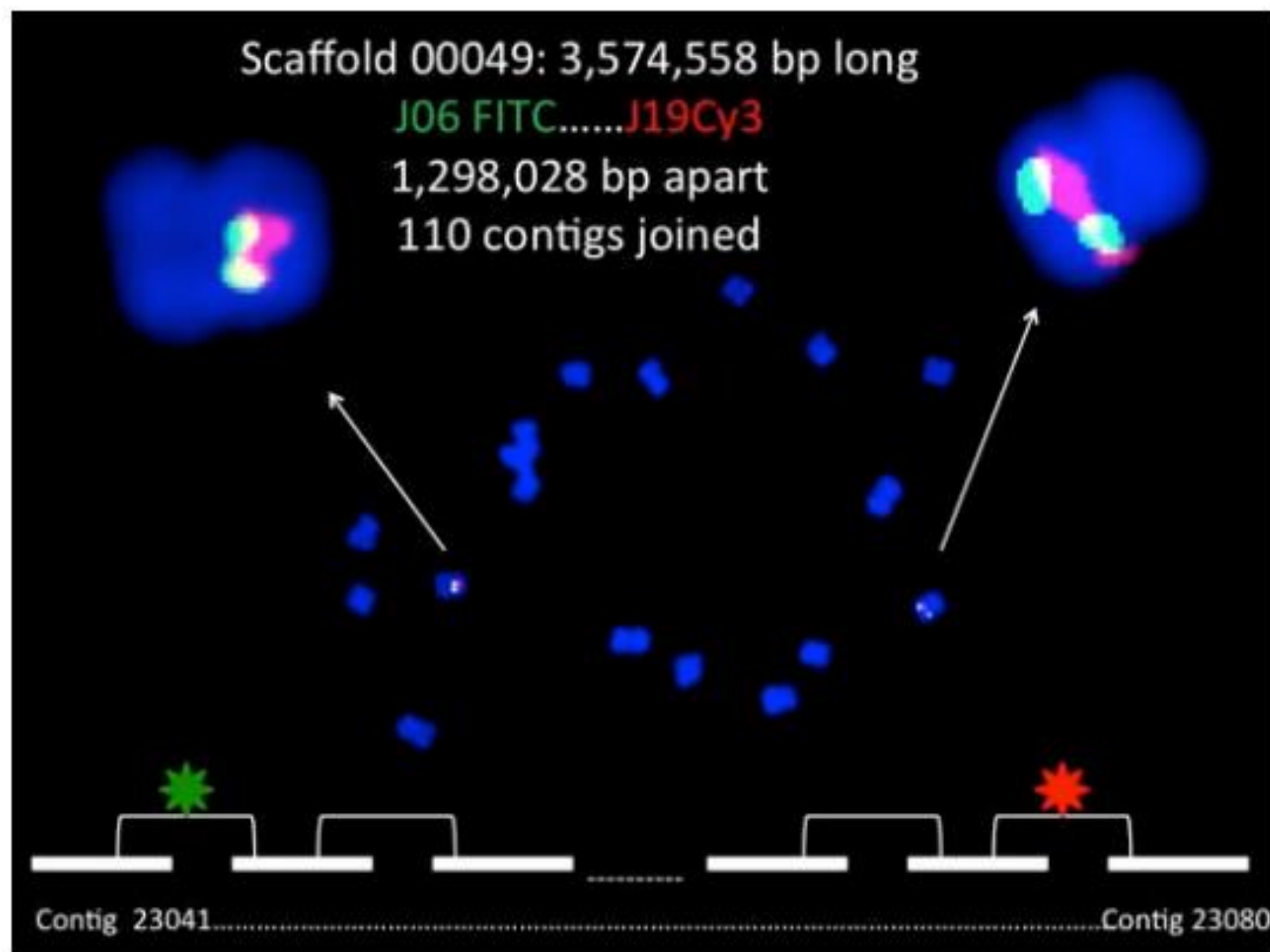


Figure S1. Assembly validation by FISH to verify co-localization of BAC-based probes that map to large scaffolds.

71 BACs anchor 35 of the larger scaffolds

Establishes a physical map for *Amborella*
 ~190 Mb of assembly
 ~20% of the genome

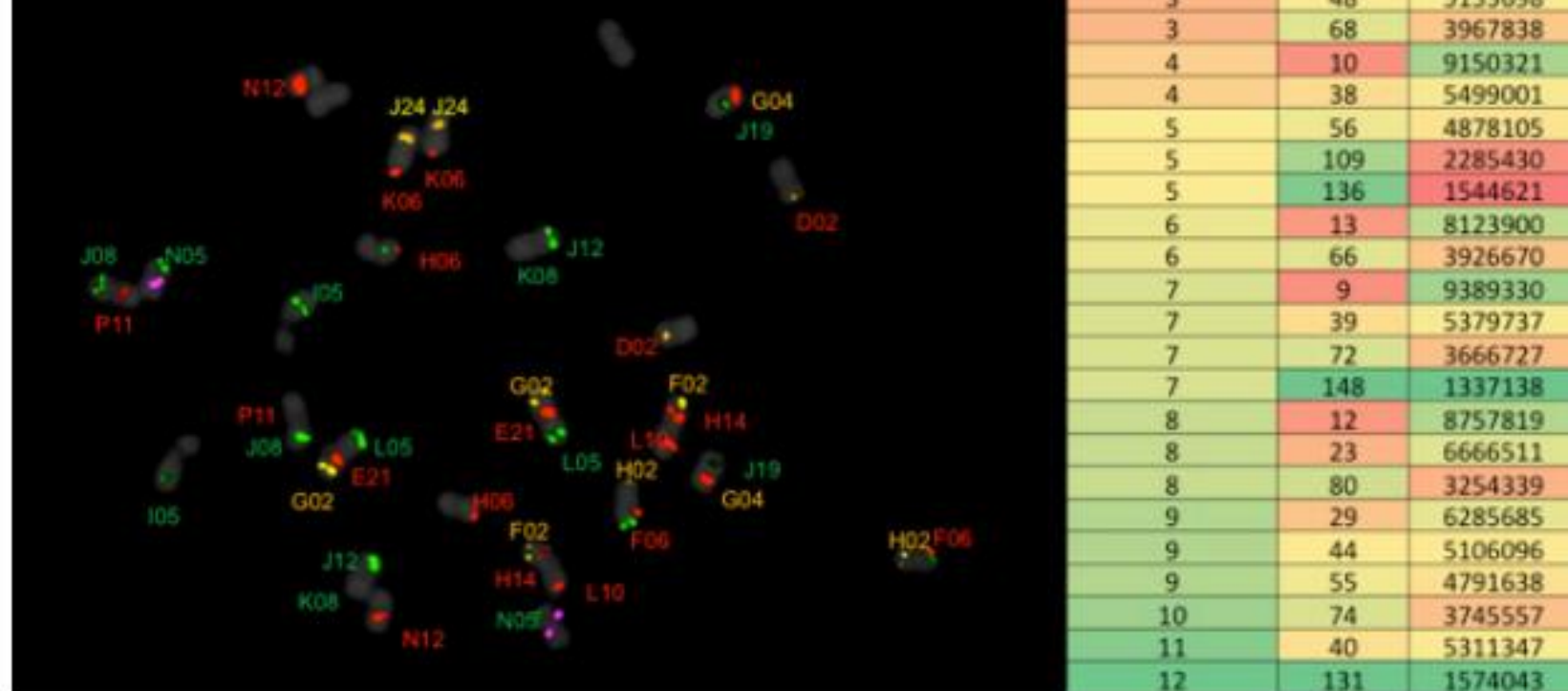


Figure S2. FISH karyotype and the chromosomal assignment of 35 scaffolds in the genome assembly.

Ülegenoomsed duplikatsioonid angiospermides

Kui *Amborella* WGD oleks
liinispetsiifiline, siis oleks
võrdlus viinamarja genoomiga
suhtes 2:3.

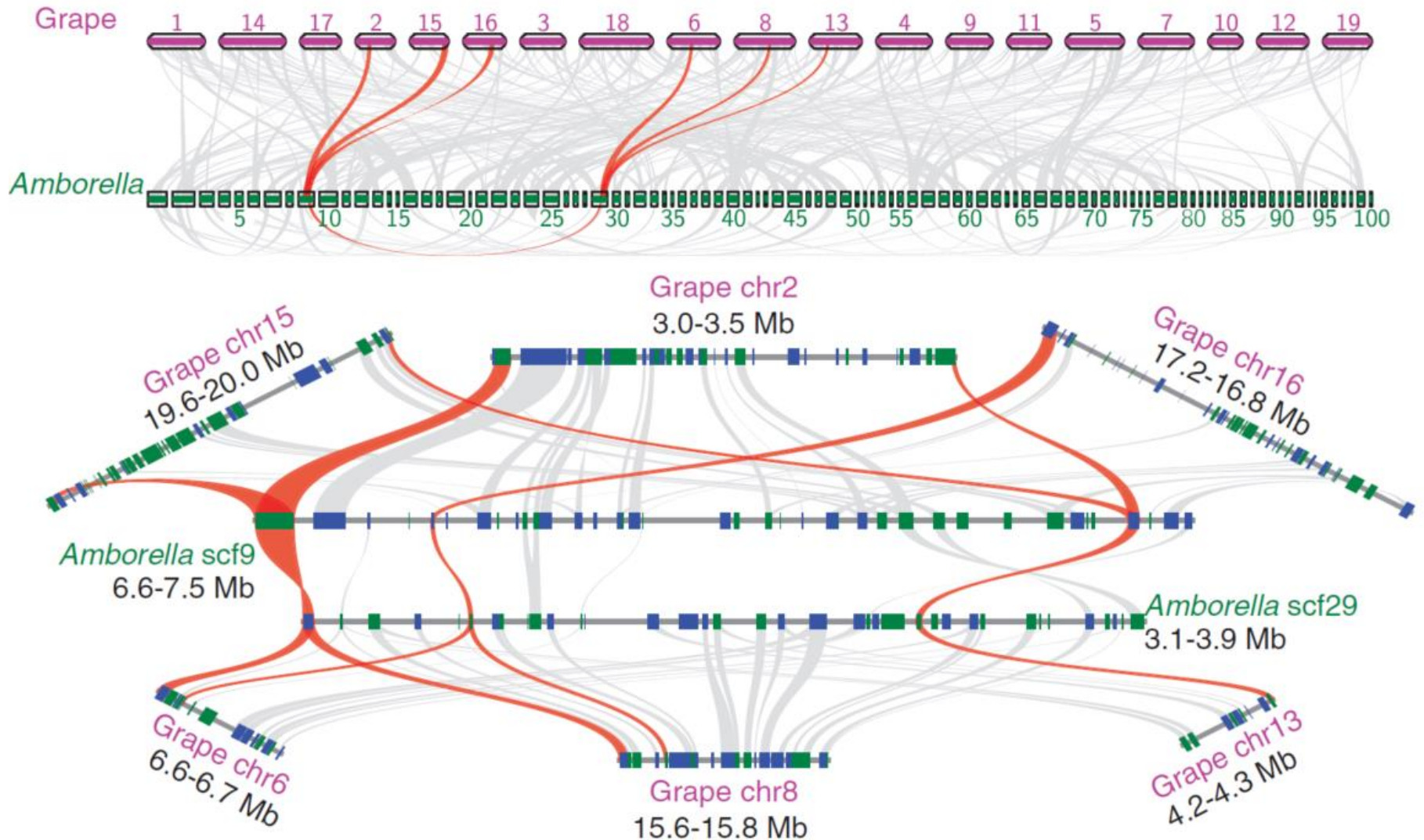
Suhe 1:3 näitab aga seda, et
Amborella WGD ei ole mitte
liinispetsiifiline, vaid on
tõenäoliselt toimunud nende
kahe organismi eellases.

Tõendab *Amborella* eristumist
enne *gamma* sündmust.

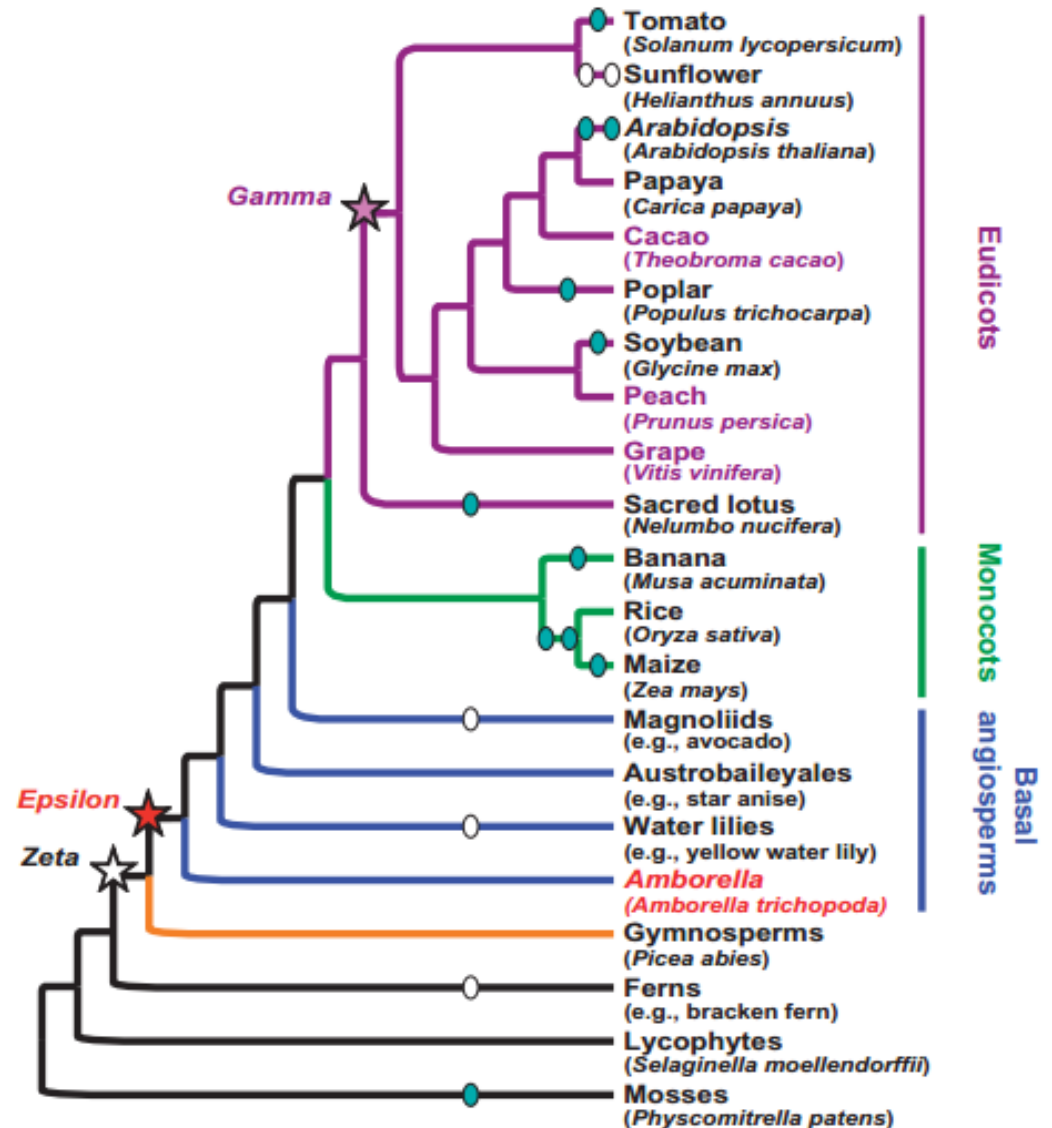


Figure S6. Syntenic path dotplot (130) of *Amborella* (x-axis) versus grape (y-axis). Note the 1:3 syntenic depth relationship between grape and *Amborella*.

Süntensus

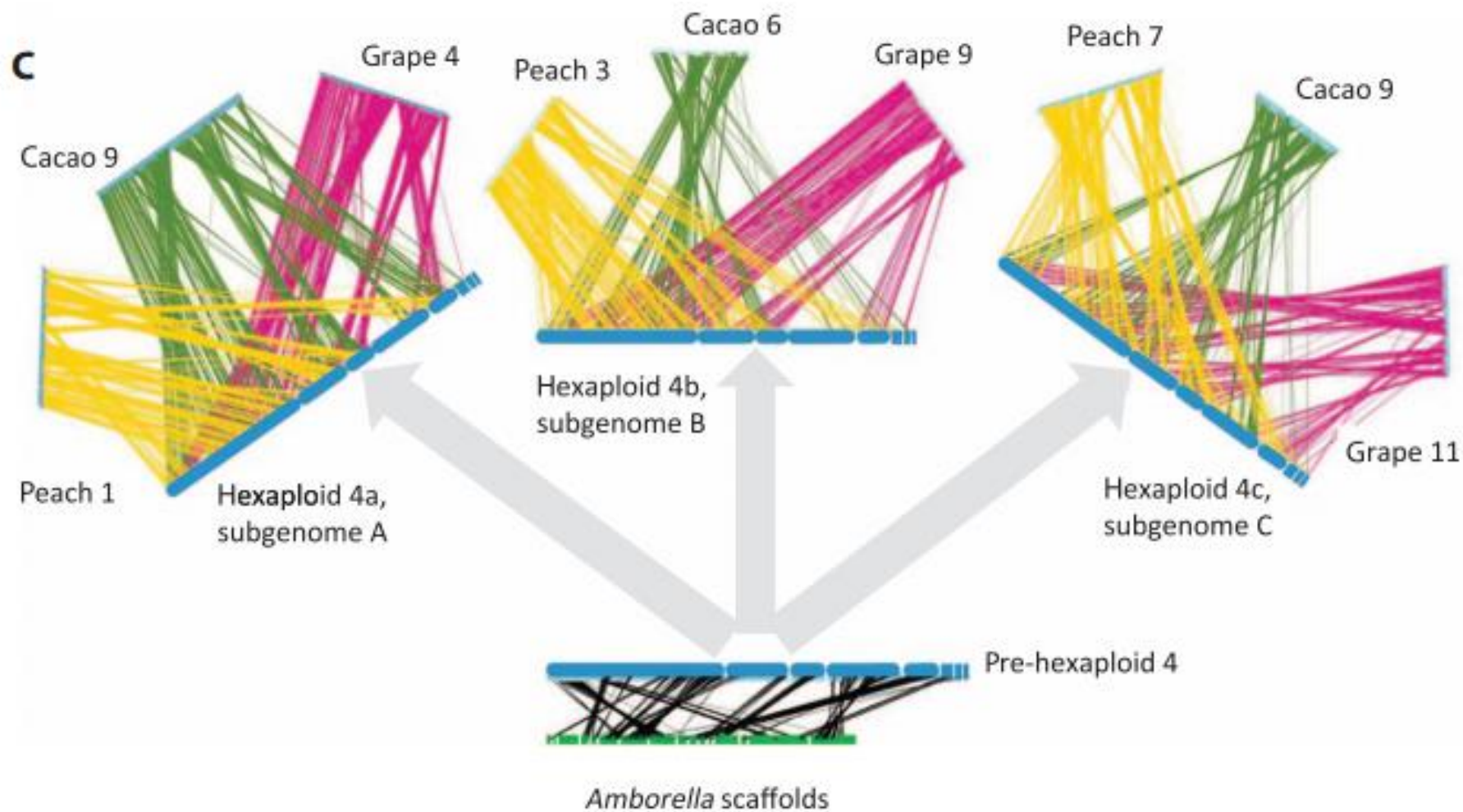


- *Epsilon* – ansestraalse angiospermide WGD.
- *Zeta* – seemnetaimede WGD. Raske tõestada – suuremahuline geenide kaotsimine ja ümberkorraldused genoomis.
- Ellipsid - polüploidistumised



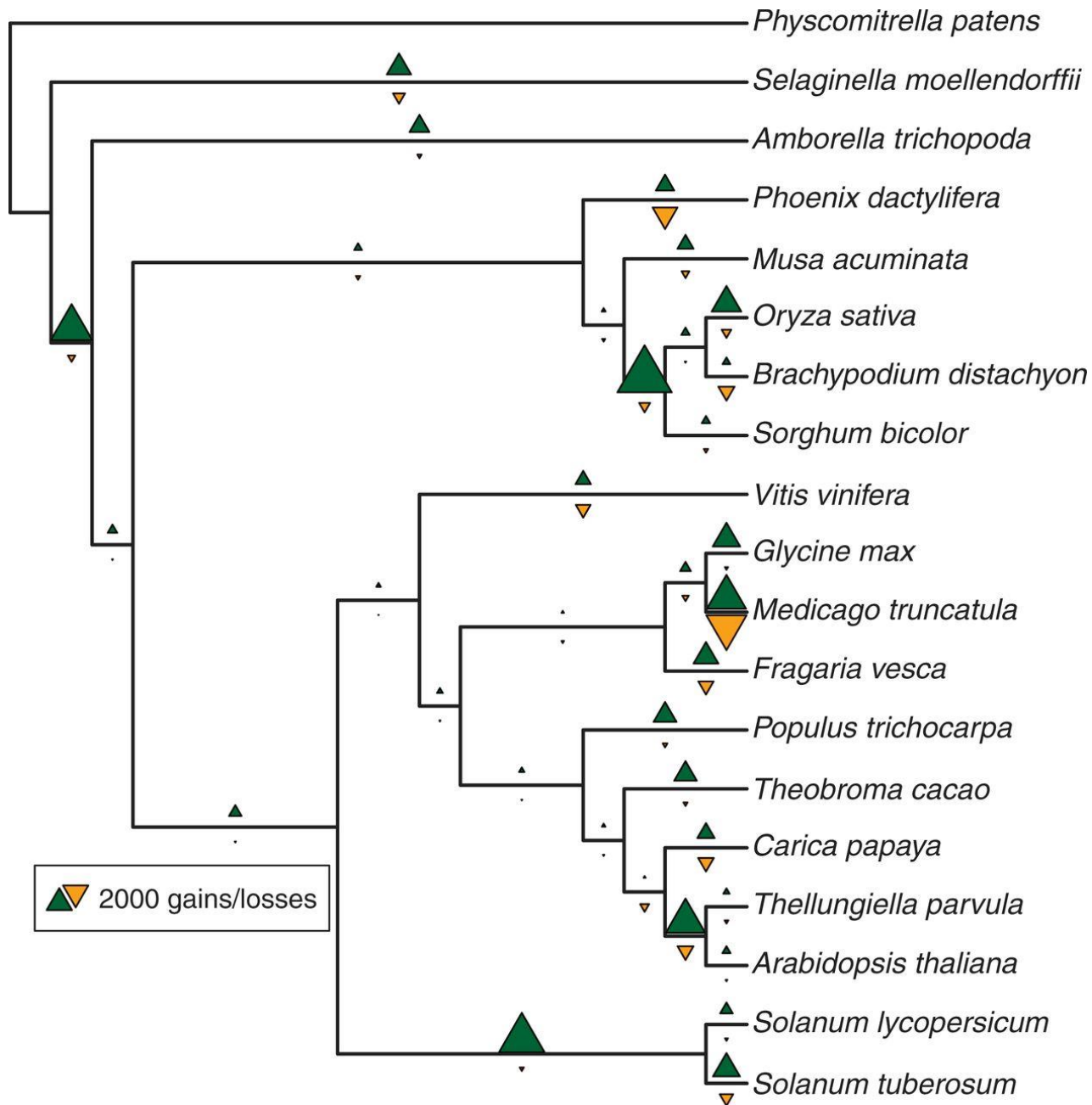
Kaheiduleheliste taimede eellasjärjestus

- Kombineeriti *Amborella* scaffold'i informatsioon kaheiduleheliste taimede (virsik, kakao, viinamari) kromosoomide informatsiooniga.
- 7 kromosoomi.
- Välisrühmana kasutati *Amborella*'t.



Angiospermide eellase genoomi iseloomustus

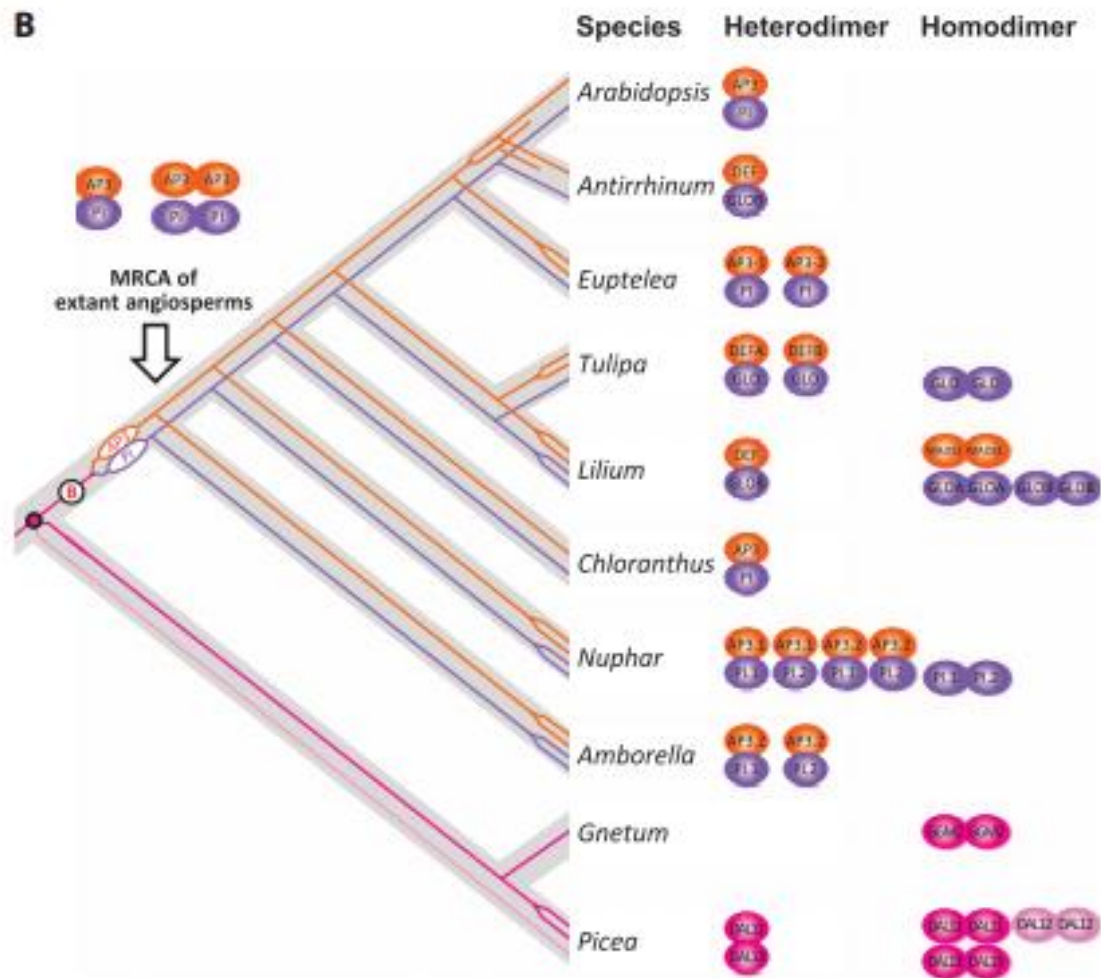
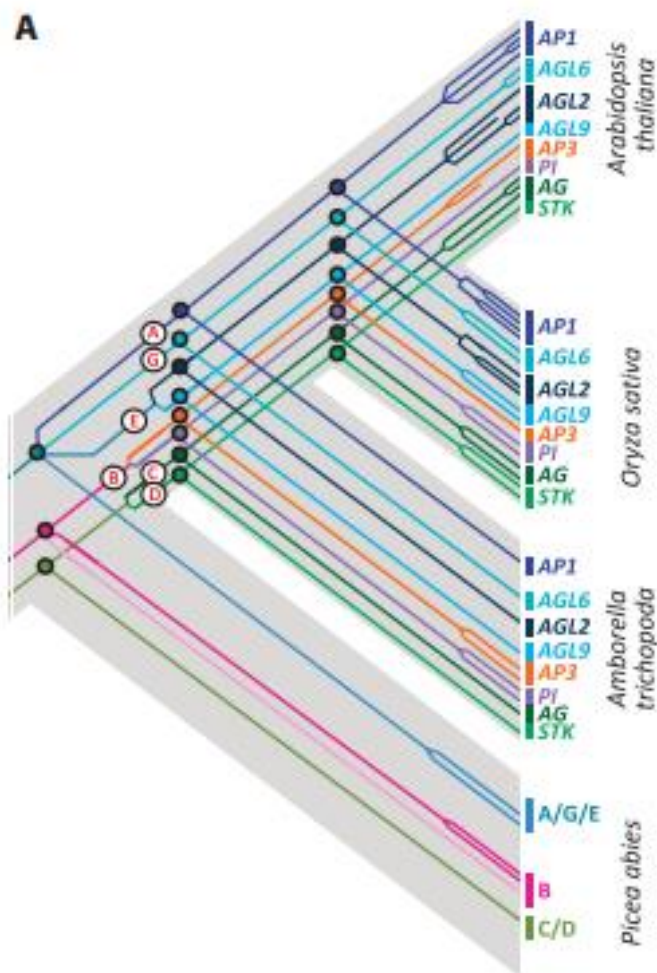
- 22 maismaataime, moodustati ortogruppe ja super-ortogruppe.
- Leiti suur hulk ortogruppe, mis arvatavasti tekkisid koos õistaimedega.
 - Paljunemine (sh MADS-Box)
 - Õie teke (*Arabidopsis*)
 - Regulaatoorsed
 - Transkriptsioon
 - Geeniekspressiooni
 - Biokeemilised protsessid
 - Metaboolsed protsessid
 - Areng
 - Viljalehed
 - Endospermid
 - Meristeemid
 - Õitsemise aeg



- Samas paljud õistaimedele iseloomulikud õie-geenid olid enne angiosperme olemas (nt õielehtede spetsiifilised geenid).
- Rohkem kui 70% geene, mis on seotud õitsemisega, olid olemas seemnetaimede eelastel.
 - Õitsemise alustamine
 - Meristeemi identsusgeenid
 - Õie struktuurigeenid.
- Õistaimede 4 peamist regulaatorgeenide gruppi on veel vanemad (esinevad soontaimedel).

MADS-BOX geenid

- Õie arenguks ühed kõige vajalikumad transkriptsioonifaktorid.
- *Amborella* genoom kodeerib 36 MADS-box geeni.
- Koonduvad 21 klaadi, milles igaühes on vähemalt üks geen mõnes teisest angiospermide liinist.
- 8 liinispetsiifilist geeniperekonda olid angiospermide eellastel, arvatavasti pärinesid need 3 seemnetaime eellase liinidest.

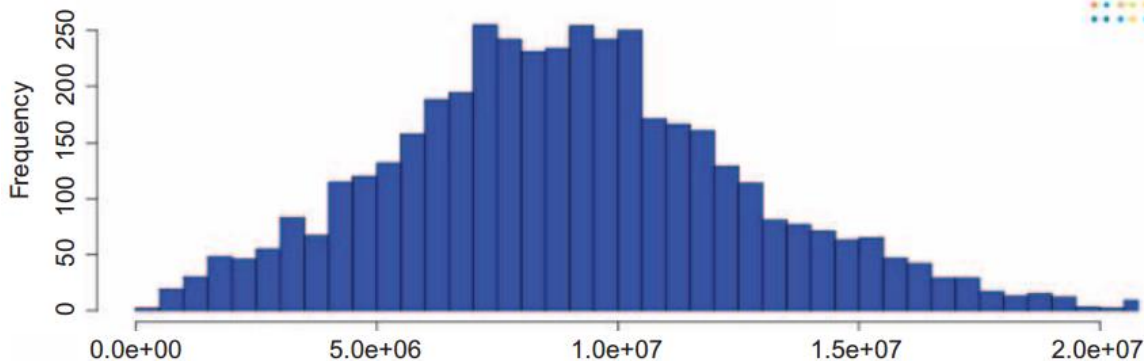


- *Amborellal* on, võrreldes teiste angiospermidega, üks GSK3 geen rohkem.
- *Amborella* seemne varu-globuliinid sarnanevad nii ühe- kui ka kaheiduleheliste taimede liinidele.
- *Amborellal* puuduvad detekteeritavad lenduvad ühendid, kuigi temas leidub terpeeni süntaase.
- Peamised angiospermide rakukesta muutused on toimunud pärast *Amborella* lahknemist.

Transponeeruvad elemendid

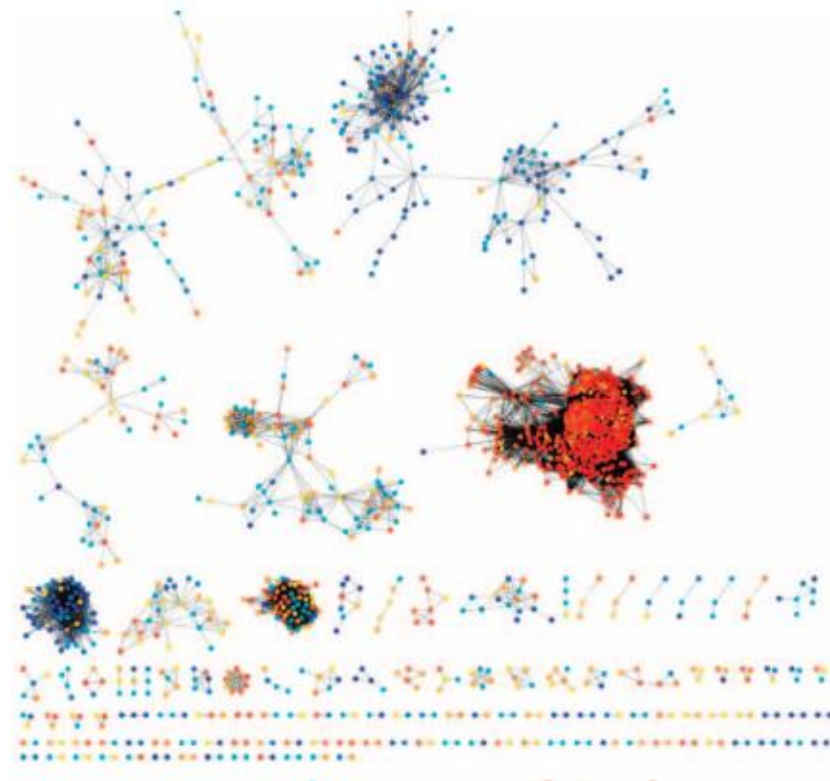
LTR retrotransposoonid

- 10% *Amborella* LTR'idel suurem divergentsus kui 80% (mille alusel jagatakse nad erinevatesse perekondadesse).
- Klasterdamise meetod.
- Insertiooniajad 4.0 – 17.6 Ma.



A

Gypsy



EPRVd

- 2.4% *Amborella* genoomist.
- TE'd ja EPRVd moodustavad *Amborella* kindlaksmääratud genoomist 57.2%.

DNA transposoonid

- 4 superklassi on esindatud.
- Üldiselt enamus kõrgalt degradeeritud
- Kõrge divergentsus ja puuduvad ITR'id.

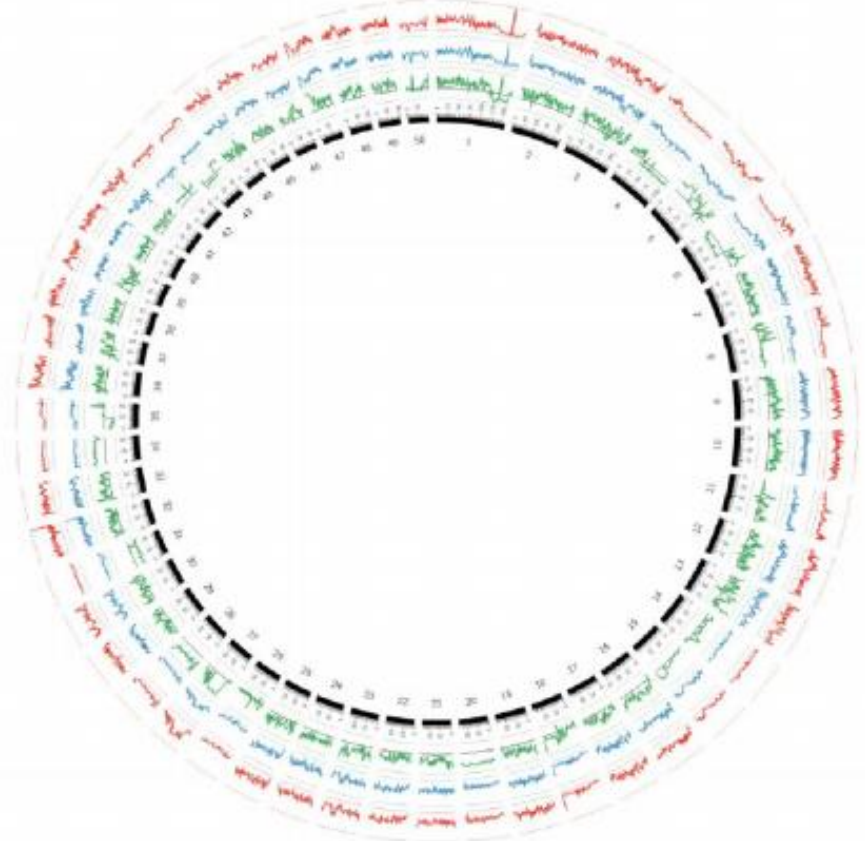
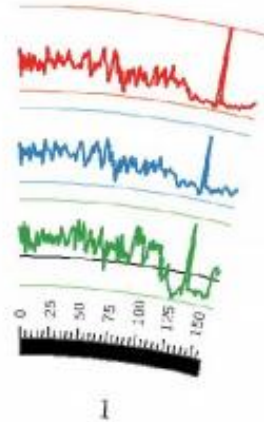
***Amborellas* puuduv transposoonne aktiivsus võib viidata kas väga efektiivsele vaigistamisele või aktiivsete transposaaside kaotamisele.**

Väikesed RNA'd

- Leiti 56 000 lookust, mis peaksid tootma reguleerivaid sRNA'sid pikkustega 20 – 24 nt.
 - siRNA
 - 124 MIRNA lookust
 - 27 perekonda (sh 5 uut) kuulusid tõenäoliselt ansestraalse õistaime genoomi.
- miRNA sihtmärgid on konserveerunud läbi õistaimede arengu.
- 78% liinispetsiifilisi miRNA'sid (63 perekonda) olid 23-24 nt pikad. Pikemaid miRNA'sid on *Amborella* genoomis üle 2 korra rohkem kui teistes maismaa taimedes. Need ei ole konserveerunud

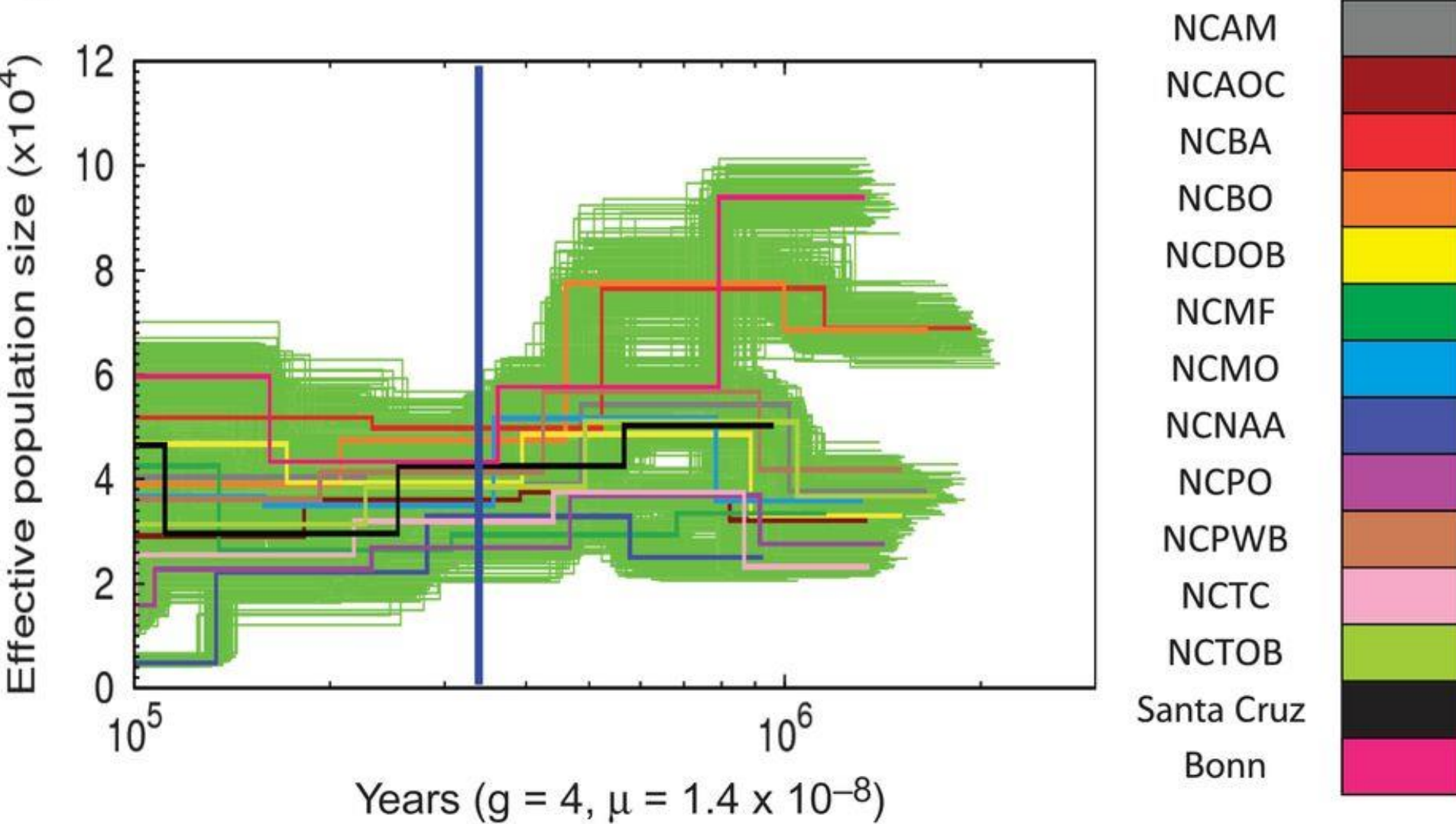
Populatsioonigenoomika

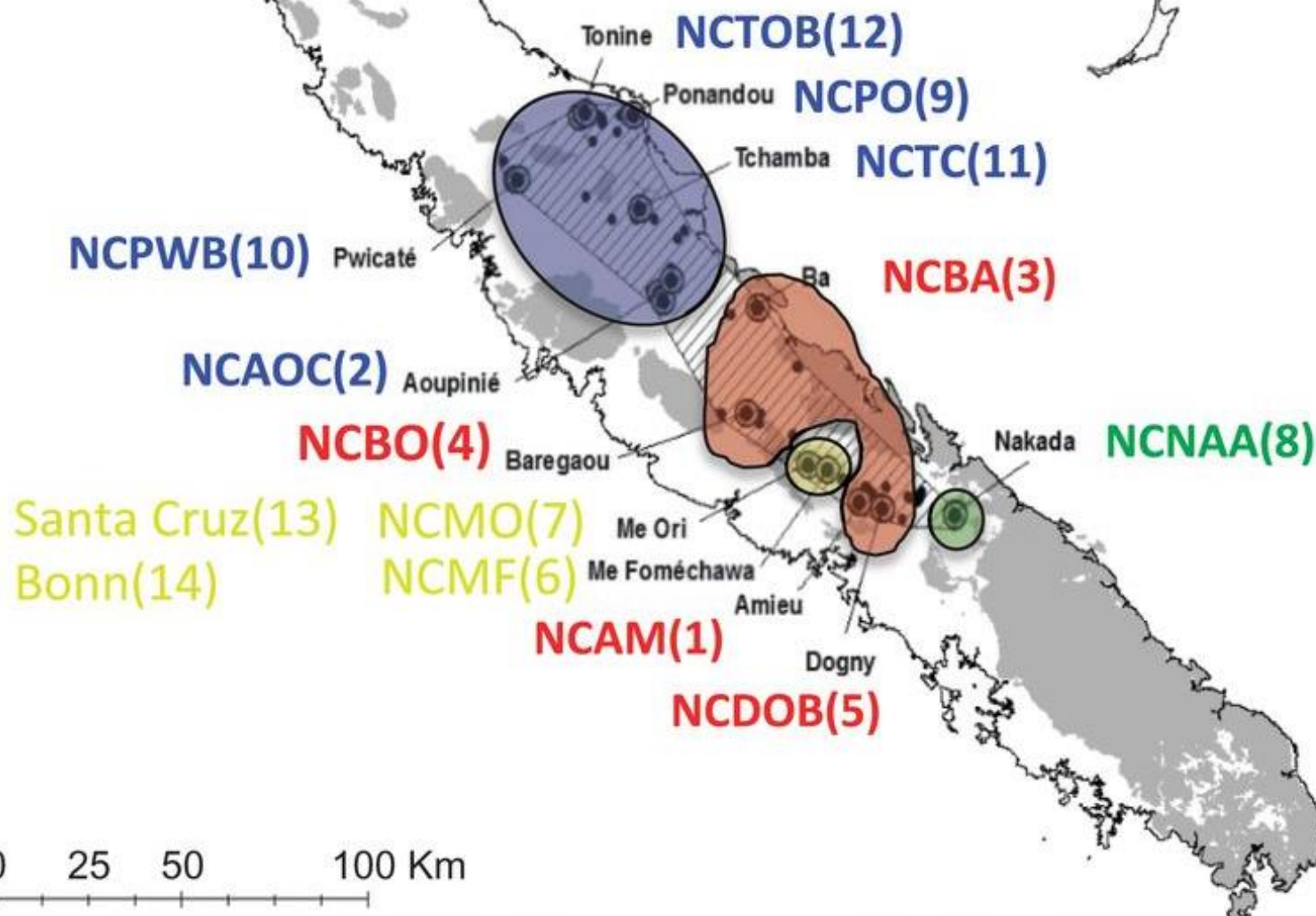
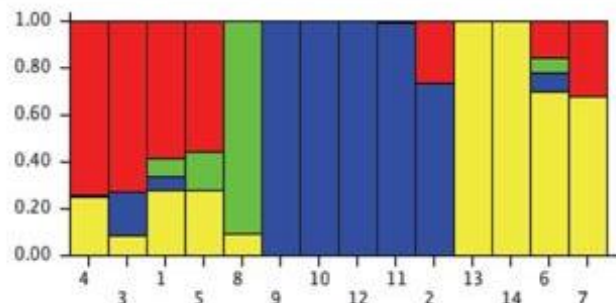
- 14 indiviidi
- Uuringusse kaasati enamuste populatsioonide esindajaid.



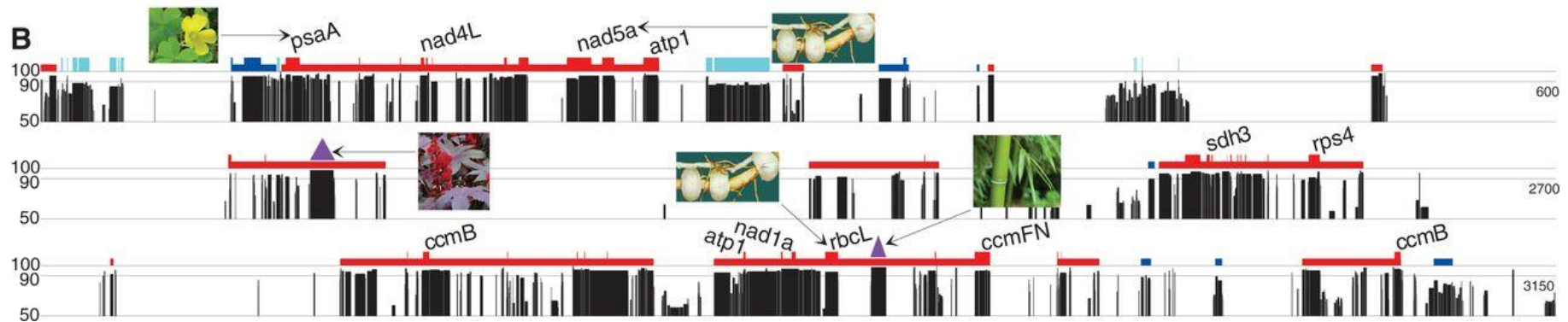
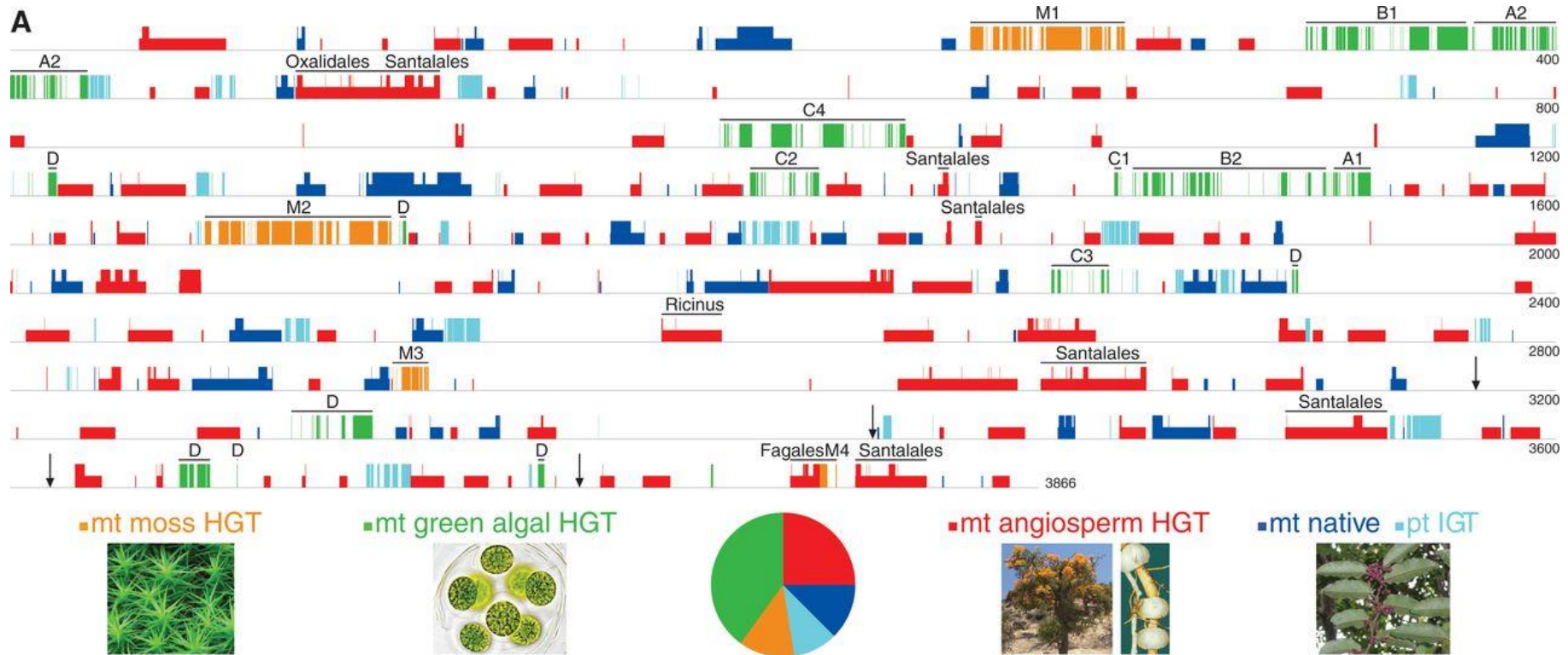
- Väga suur liigisisene diversiteet ($\theta_w = 0.0017, p = 0.0021$).
- Tajima D (üle genoomi) positiivne ($D = 0.8137$)
 - Tasakaalustav valik
 - Hiljutine pudelikaela efekt või migratsioon
- Genoomisisene varieeruvus suurem
 - Kõrge polümorfsuse ja heterogeensusega piirkonnad.
 - Invariantseid piirkonnad negatiivse Tajima D' ga.

Üldine populatsiooni suuruse vähenemine.

A

B

Amborella mitokonder



TÄNAN!