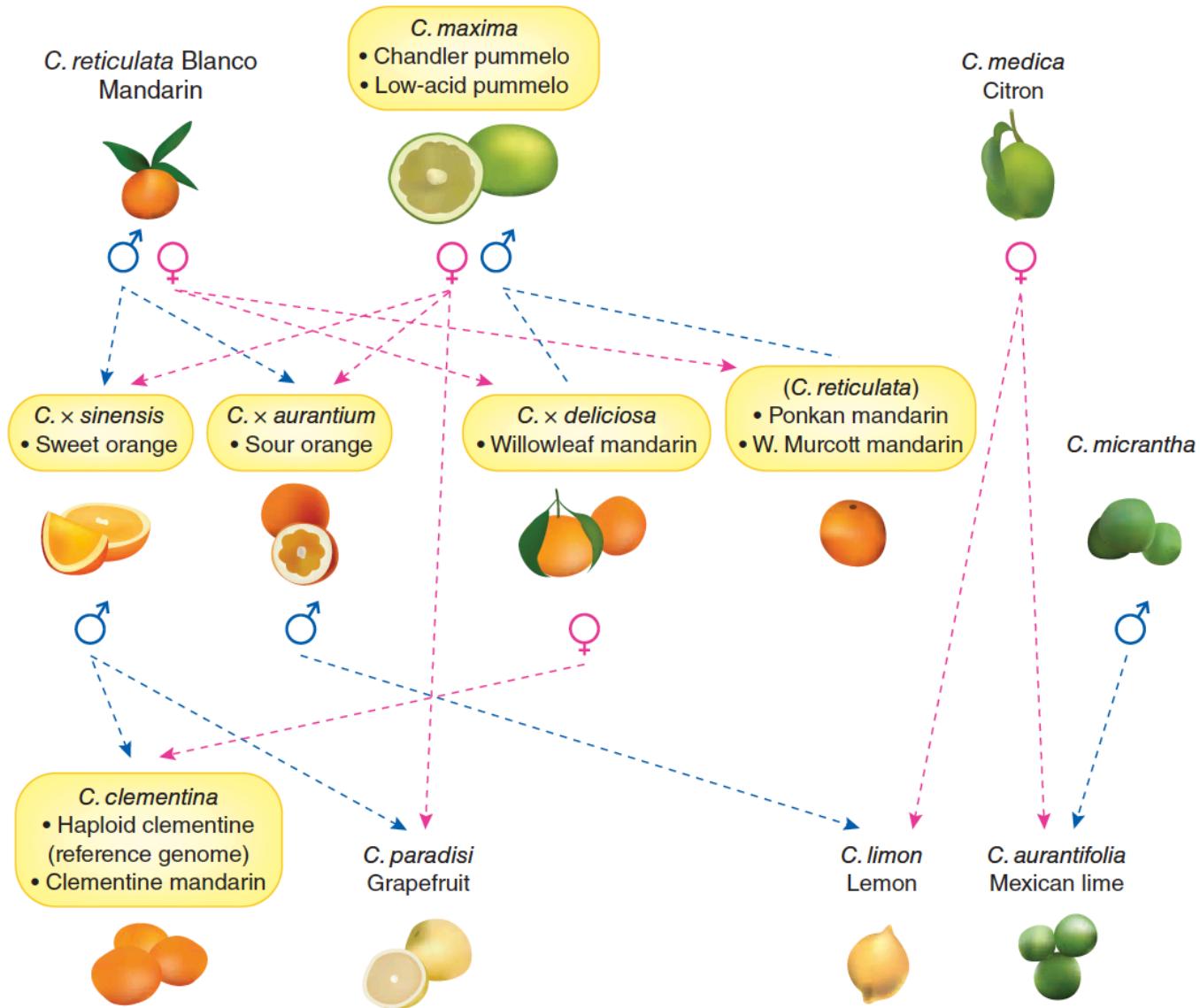
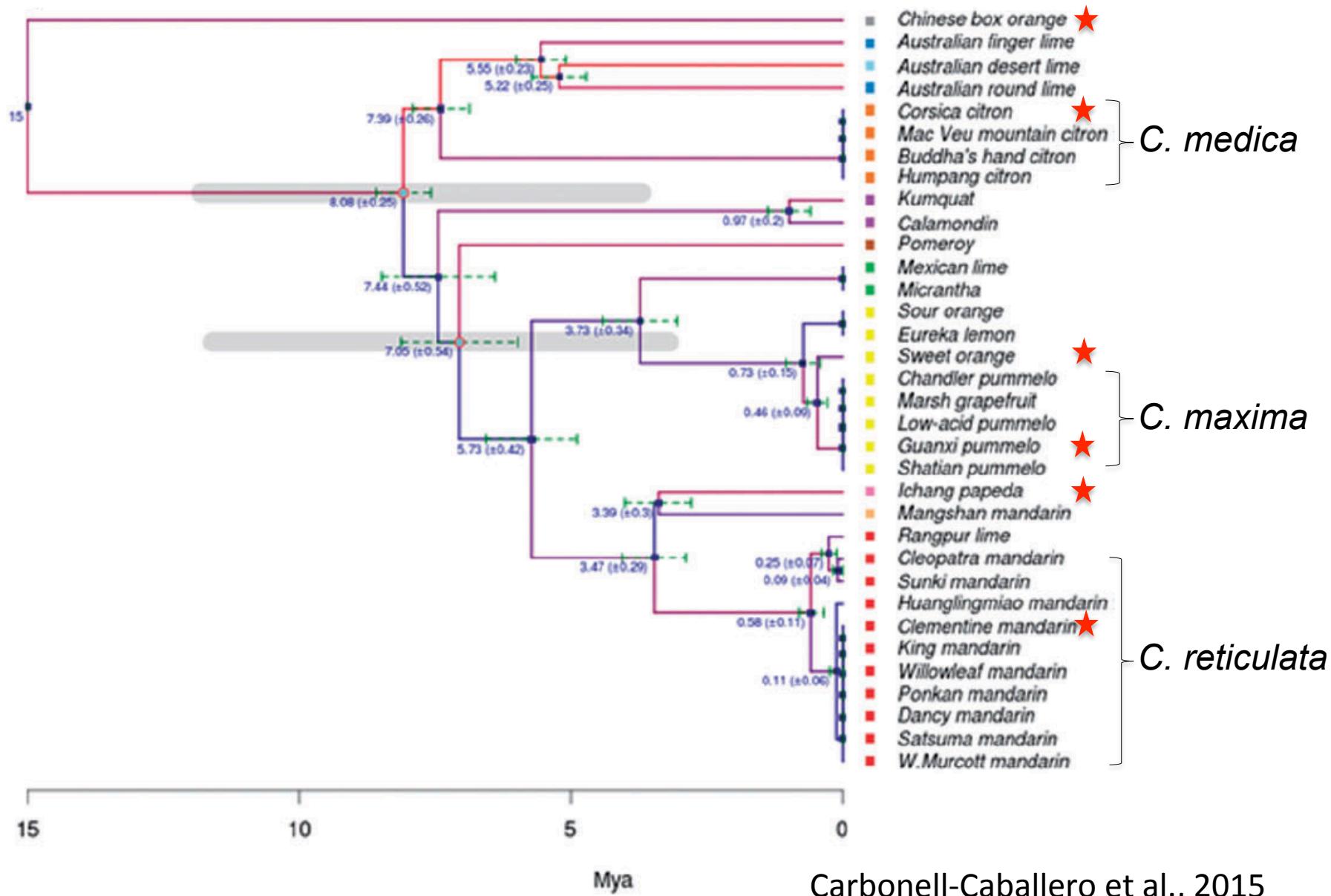


# The origin of citrus species complex



Velasco et al., 2014; Wu et al., 2014

# Phylogenetic tree and divergence time of the citrus species based on Chloroplast genomes



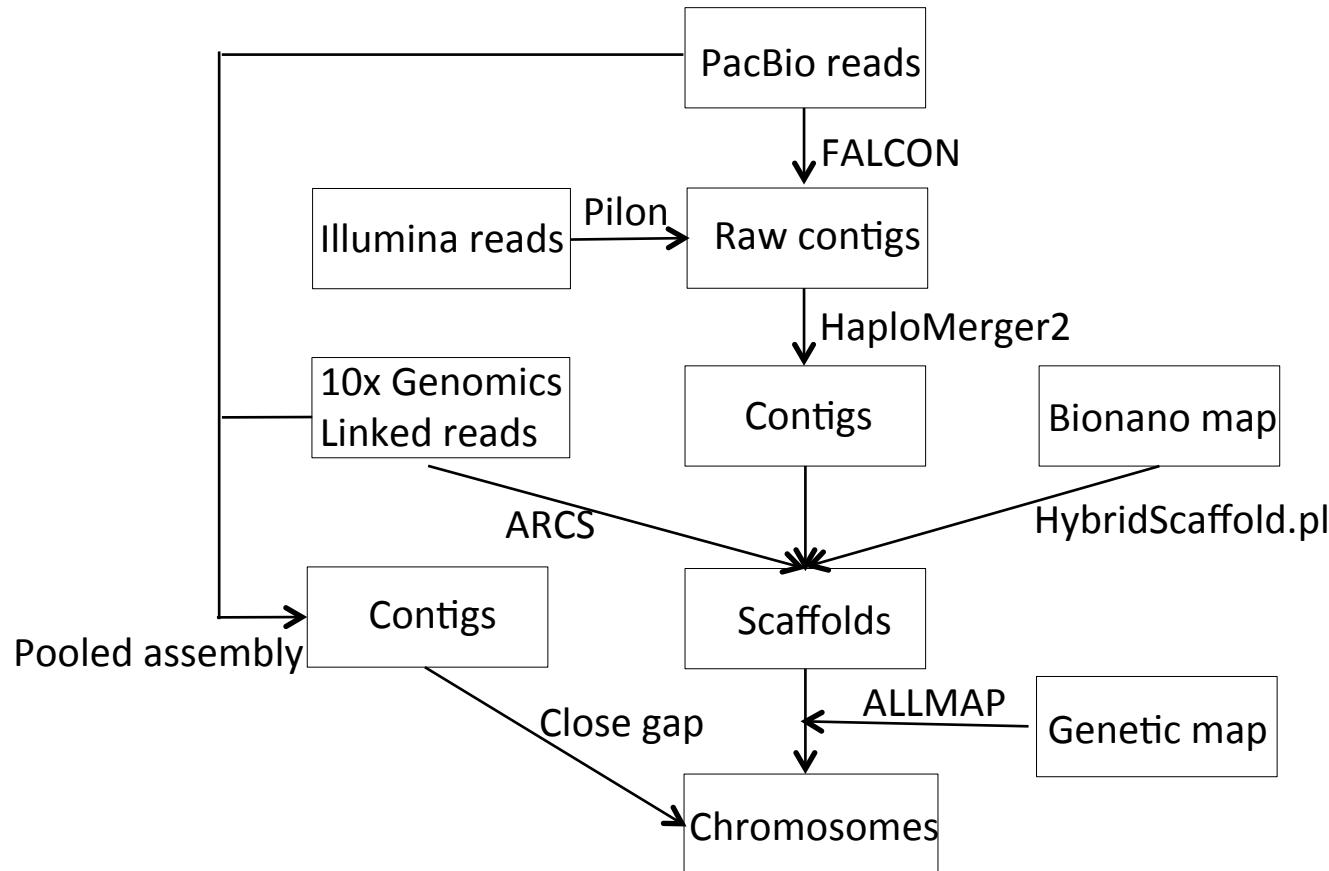
## Challenges in citrus genomics?

- One reference genome is not enough: need multiple reference genomes and a pan-genome that encompasses entire set of citrus genes and functional elements
- Most citrus species are highly heterozygous: accurate genome assembly and phasing of haplotypes can be difficult

## Goals

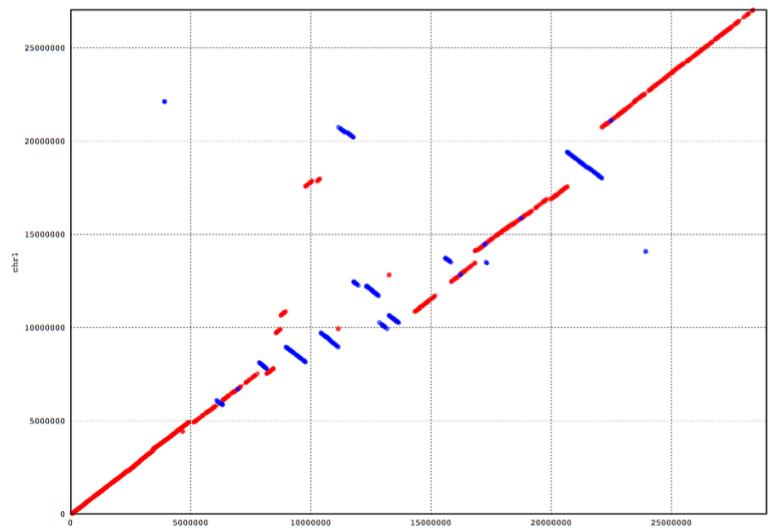
- Assemble a reference level citrus genome from a diploid material
- Phase heterozygous SNPs into haplotypes and construct a phased variation map
- Explore the impact of deleterious mutations on citrus breeding

# Workflow of hybrid genome assembly

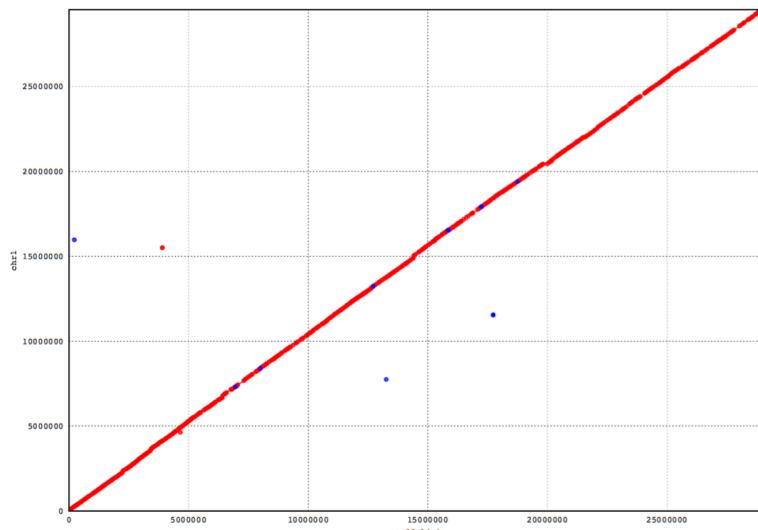


# Pacbio long reads are not sufficient for resolving genome complexity

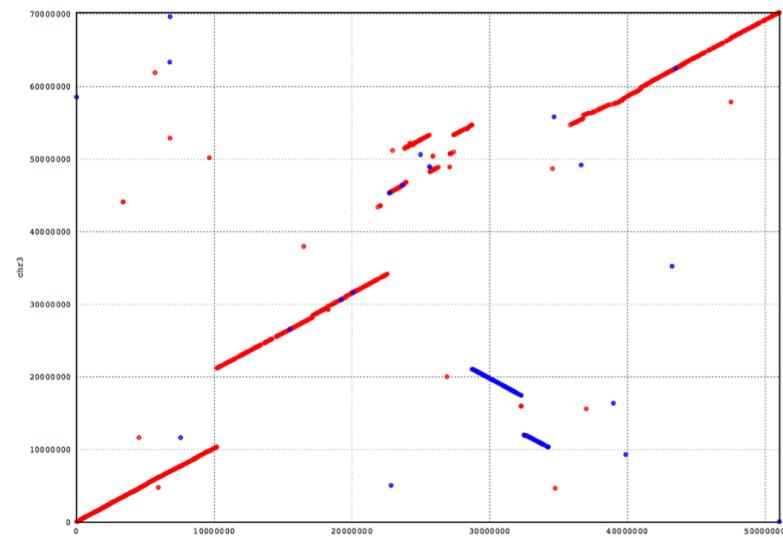
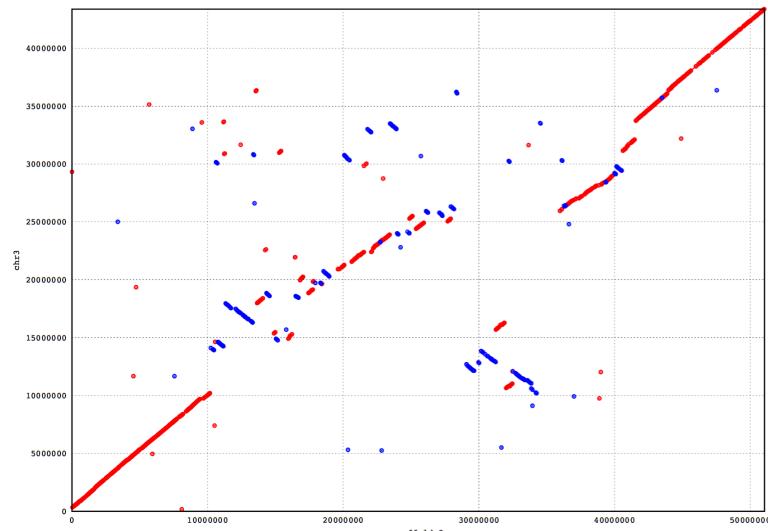
Pacbio only



Pacbio + bionano hybrid



Clementine scaffold 1



Clementine scaffold 3

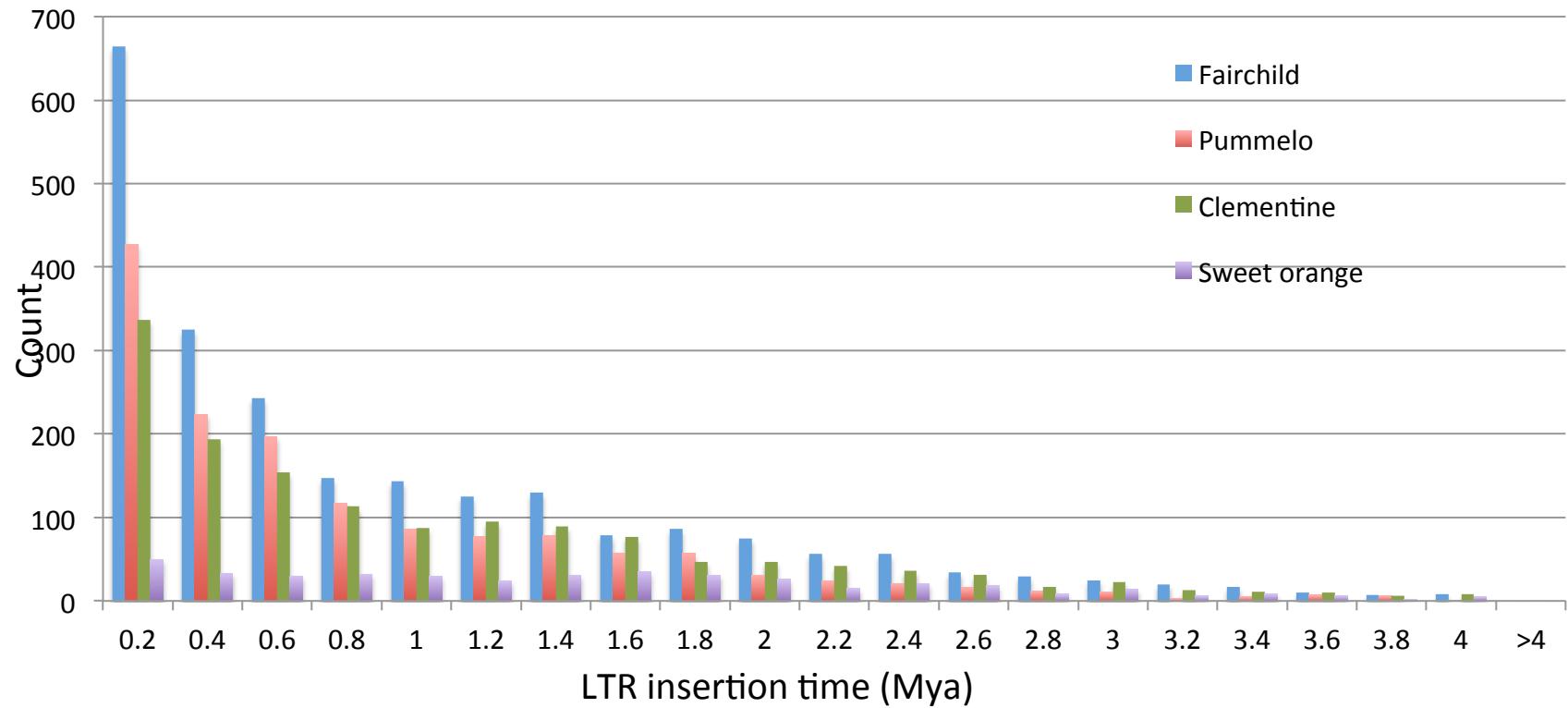
# Hybrid assembly of Fairchild reference genome

	Sweet Orange	Clementine	Pummelo	Fairchild
Genome type	Haploid	Haploid	Haploid	Diploid
Assembled size (Mb)	327	301	345	366
Gap-free seq (Mb)	301	295	344	358
Longest scaf (kb)	8 393	5 105	53 007	64 121
Longest ctg (kb)	323	1 235	10 624	22 448
Scaf/ctg N50 (kb)	1778/50	6800/119	40 395 / 2182	42 136 / 9 996
Assembler	SOAPdenovo	ARACHNE	HGAP	Hybrid scaffold
Sequence	214X Illumina	7X Sanger	57X Pacbio	120X Pacbio + bionano + 10x
Gene completeness (CEGMA)	91.5% (227/248)	91.9% (228/248)	92.7% (230/248)	91.9% (228/248)
Gene completeness (BUSCO)	95.2% (1371/1440)	94.2% (1356/1440)	95.2% (1373/1440)	94.2% (1357/1440)
Intergenic completeness (LAI)	3	11	12	15

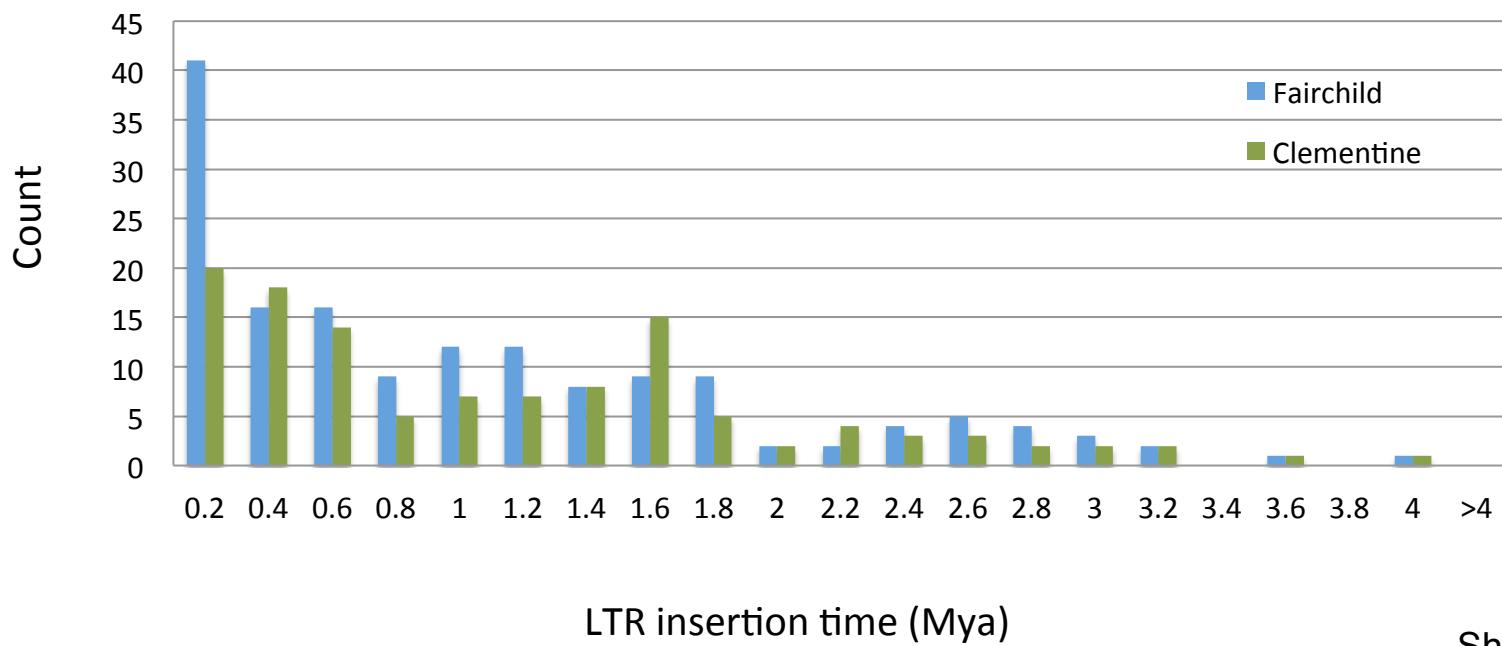
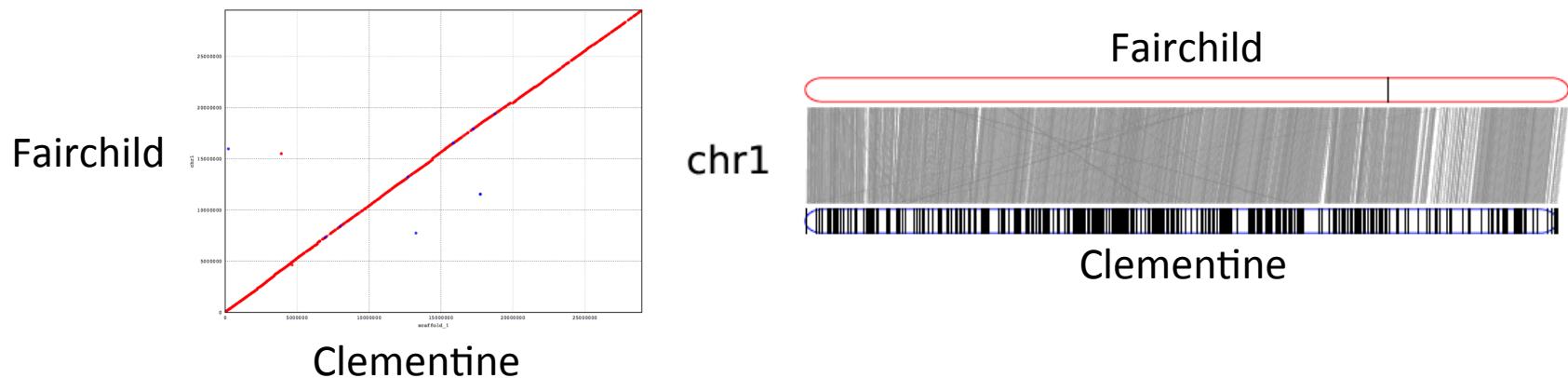
## More TEs are assembled in Fairchild genome?

	Sweet Orange (Mb)	Clementine (Mb)	Pummelo (Mb)	Fairchild (Mb)
Retrotransposon	54.81	63.97	113.34	130.50
LTR/Copia	23.61	23.76	54.05	26.86
LTR/Gypsy	29.41	36.19	39.44	65.27
LINE	1.20	3.74	5.70	9.81
Other Retro	0.59	0.28	14.14	28.57
DNA Transposon	6.86	8.11	25.39	26.46
<i>hAT</i>	1.09	2.87	7.34	6.91
<i>Mutator</i>	4.57	2.63	9.02	8.26
<i>PIF-Harbinger</i>	0.32	0.29	2.10	3.01
<i>Tc1-Mariner</i>	NA	0.45	1.33	1.12
CACTA	0.63	1.70	5.60	5.86
Other DNA	0.22	0.17	NA	1.31
<i>Helitron</i>	0.03	0.00	1.36	7.58
Unknown	NA	56.13	10.68	20.78
Total Transposon	61.67	128.21	149.40	185.33
Contig size (Mb)	301	295	344	358

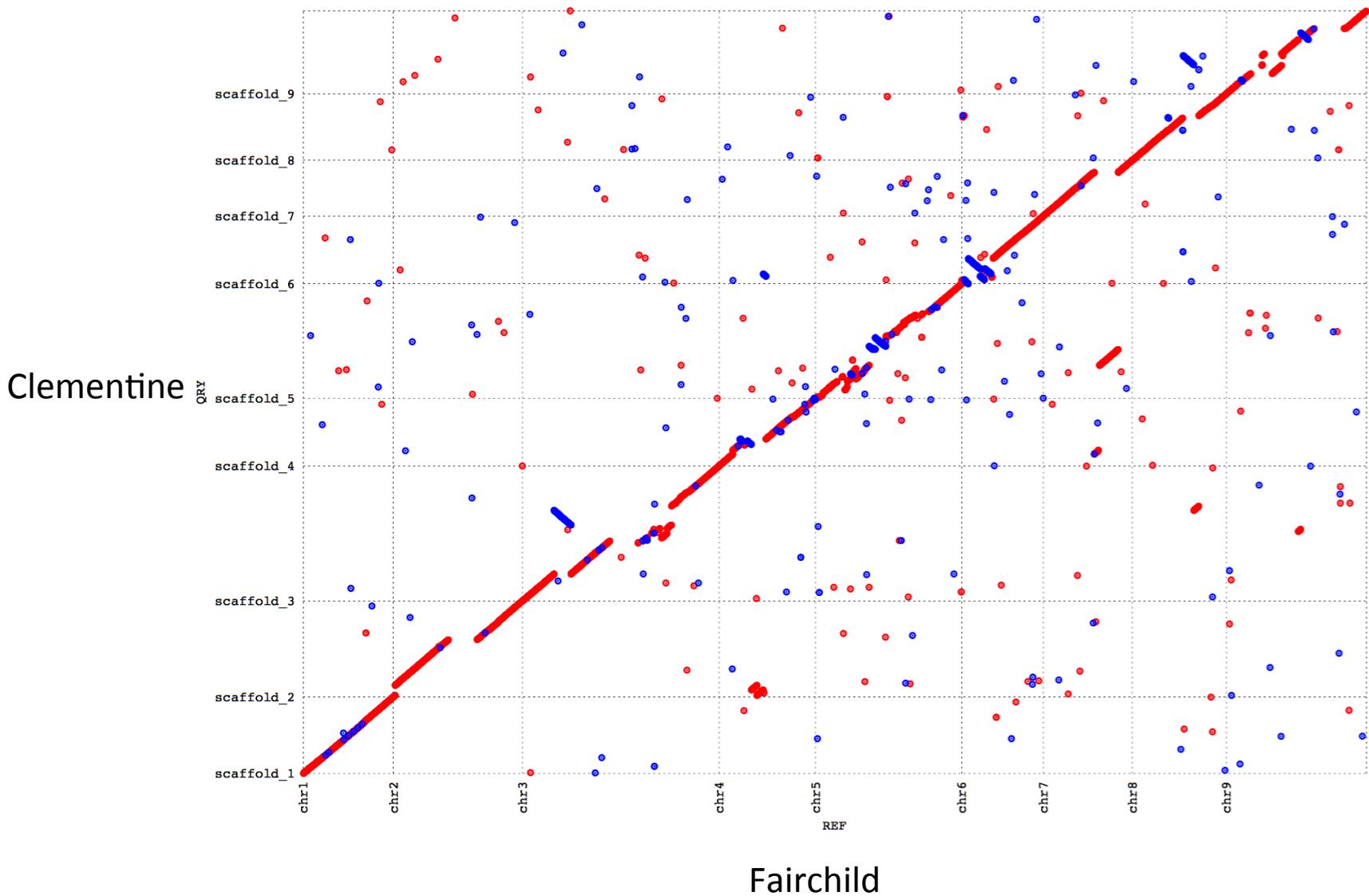
# More intact LTR retrotransposon are assembled in Fairchild genome?



# Intact LTR on chromosome 1 between Fairchild and Clementine

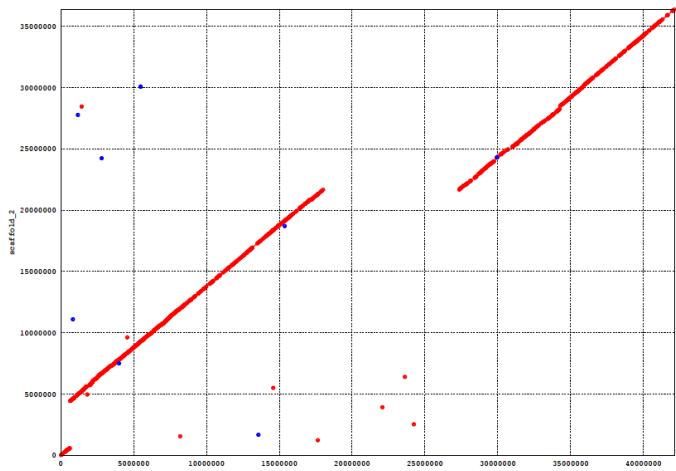


# Sequence comparisons reveal major rearrangements between Clementine and Fairchild mandarin

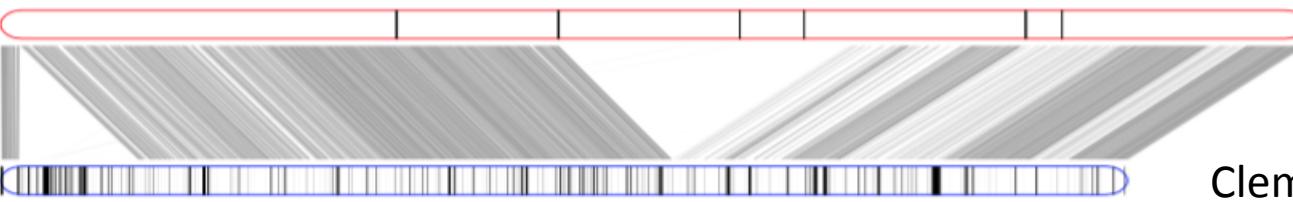


# Sequence comparison on chromosome 2

Clementine



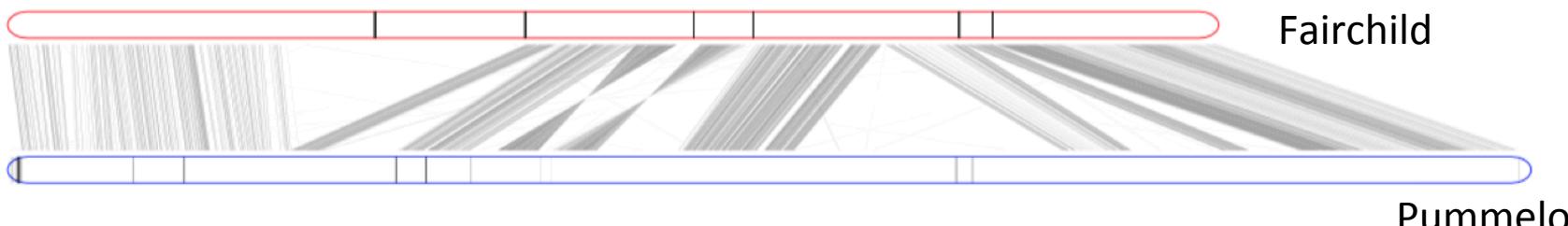
chr2



Fairchild

Clementine

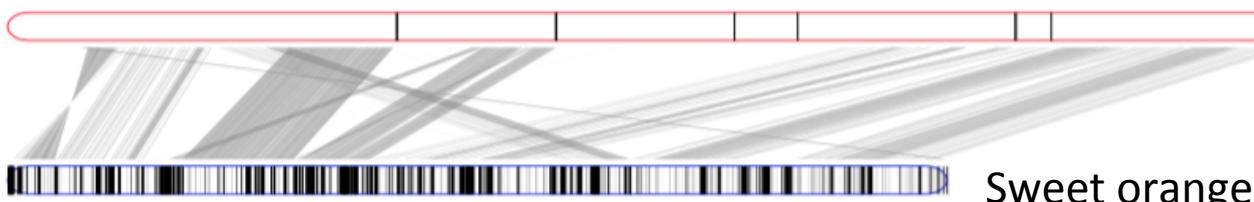
chr2



Fairchild

Pummelo

chr2

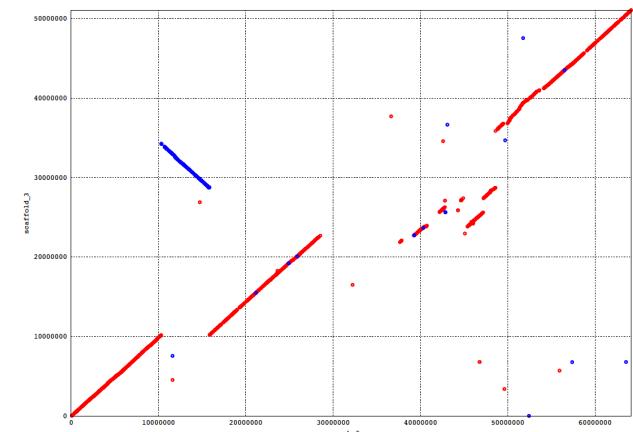


Fairchild

Sweet orange

# Sequence comparison on chromosome 3

Clementine



Fairchild

Fairchild

Clementine

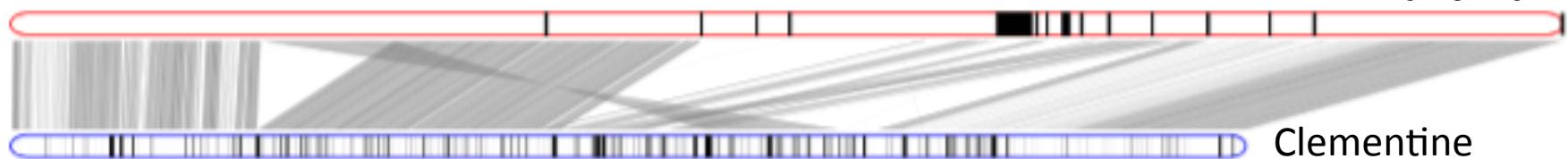
Fairchild

Pummelo

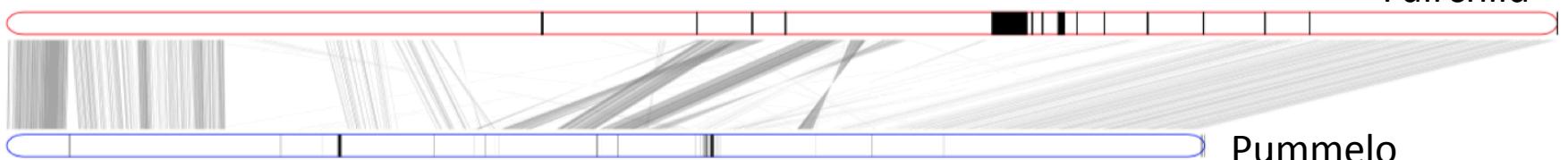
Fairchild

Sweet orange

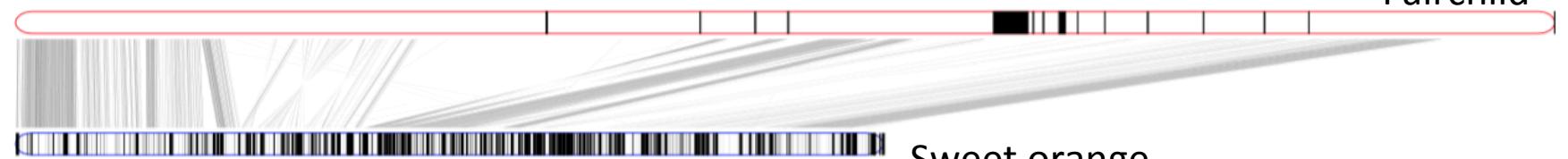
chr3



chr3

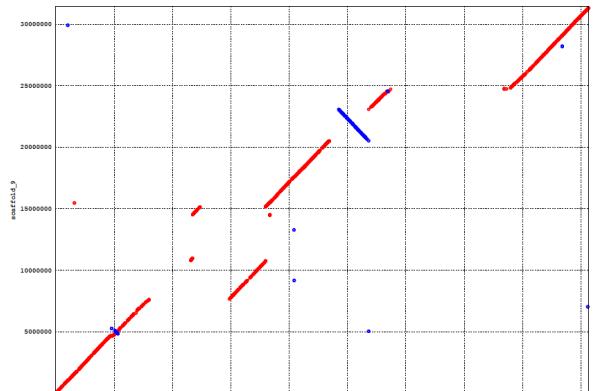


chr3



# Sequence comparison on chromosome 9

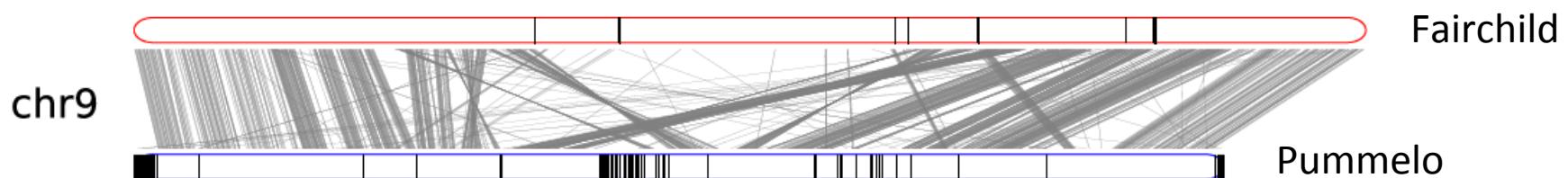
Clementine



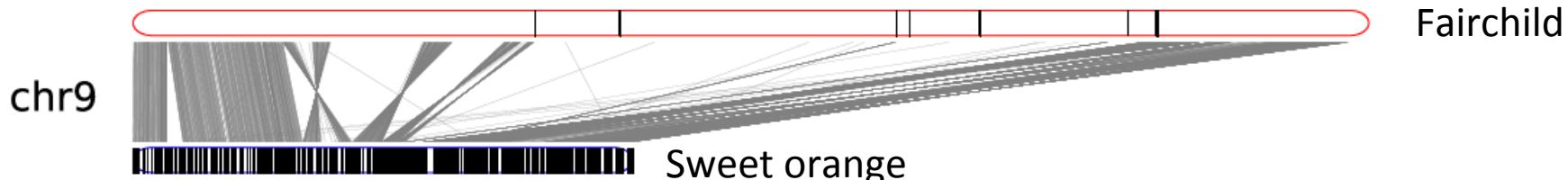
Fairchild



Clementine

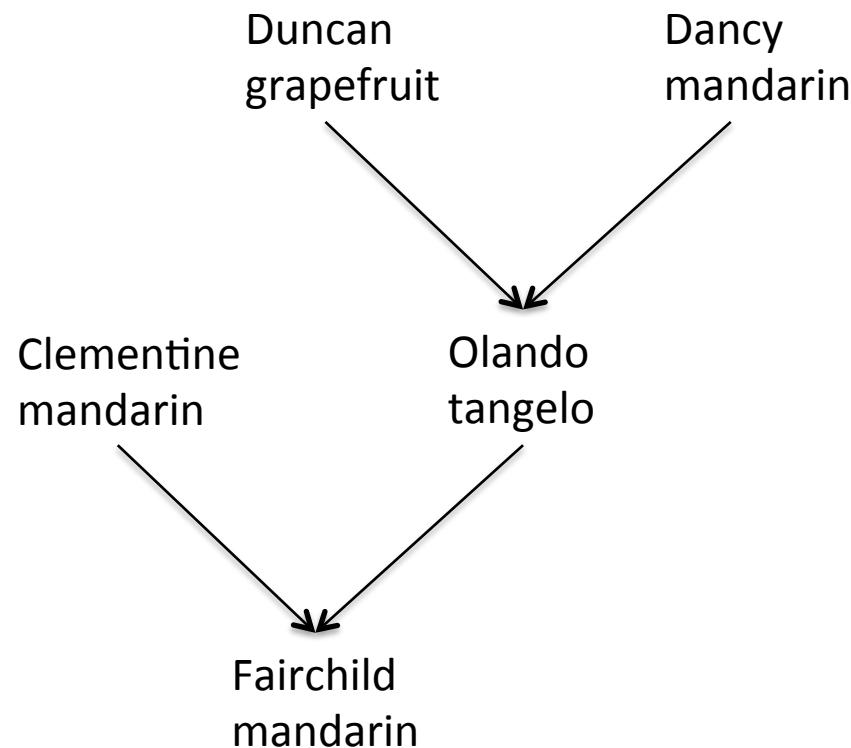


Pummelo



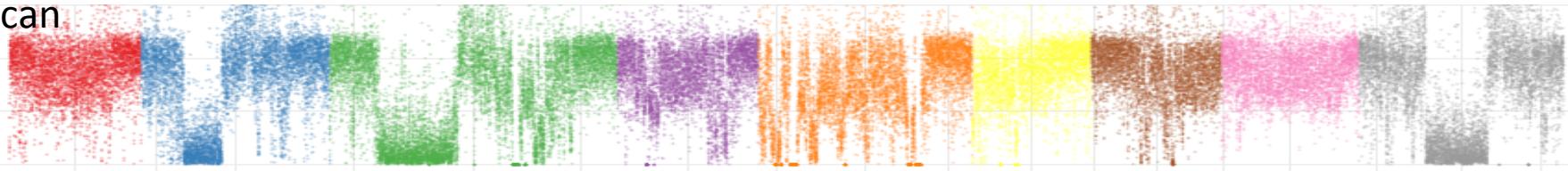
Sweet orange

# Fairchild pedigree

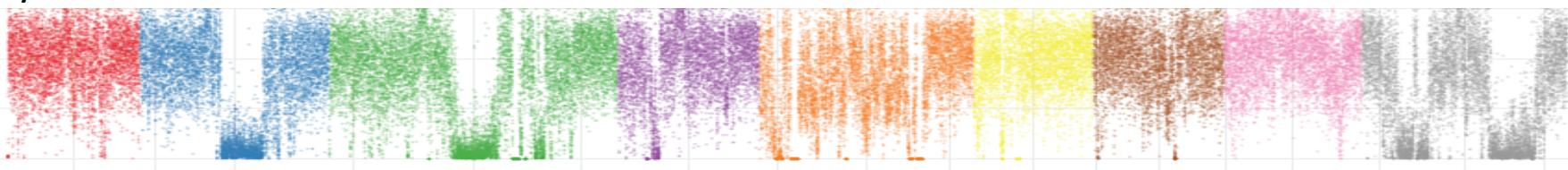


# Three regions in the Fairchild genome contain divergent haplotypes

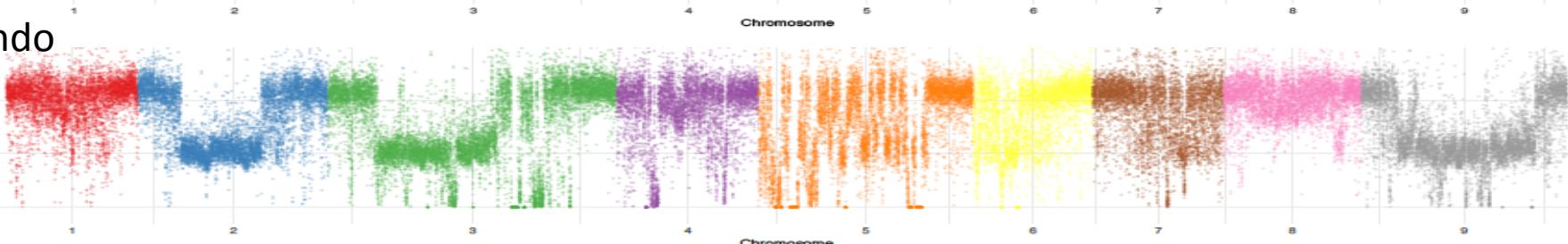
Duncan



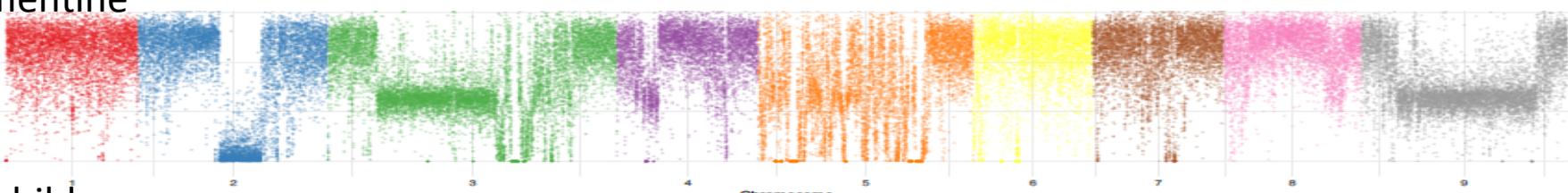
Dancy



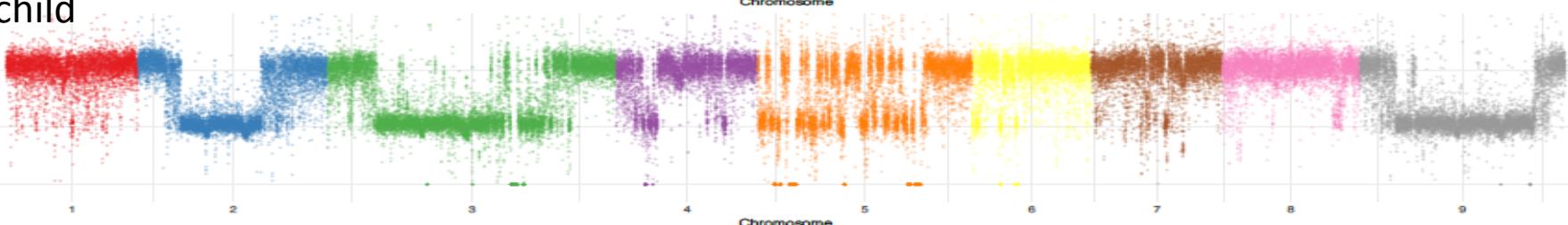
# Olando



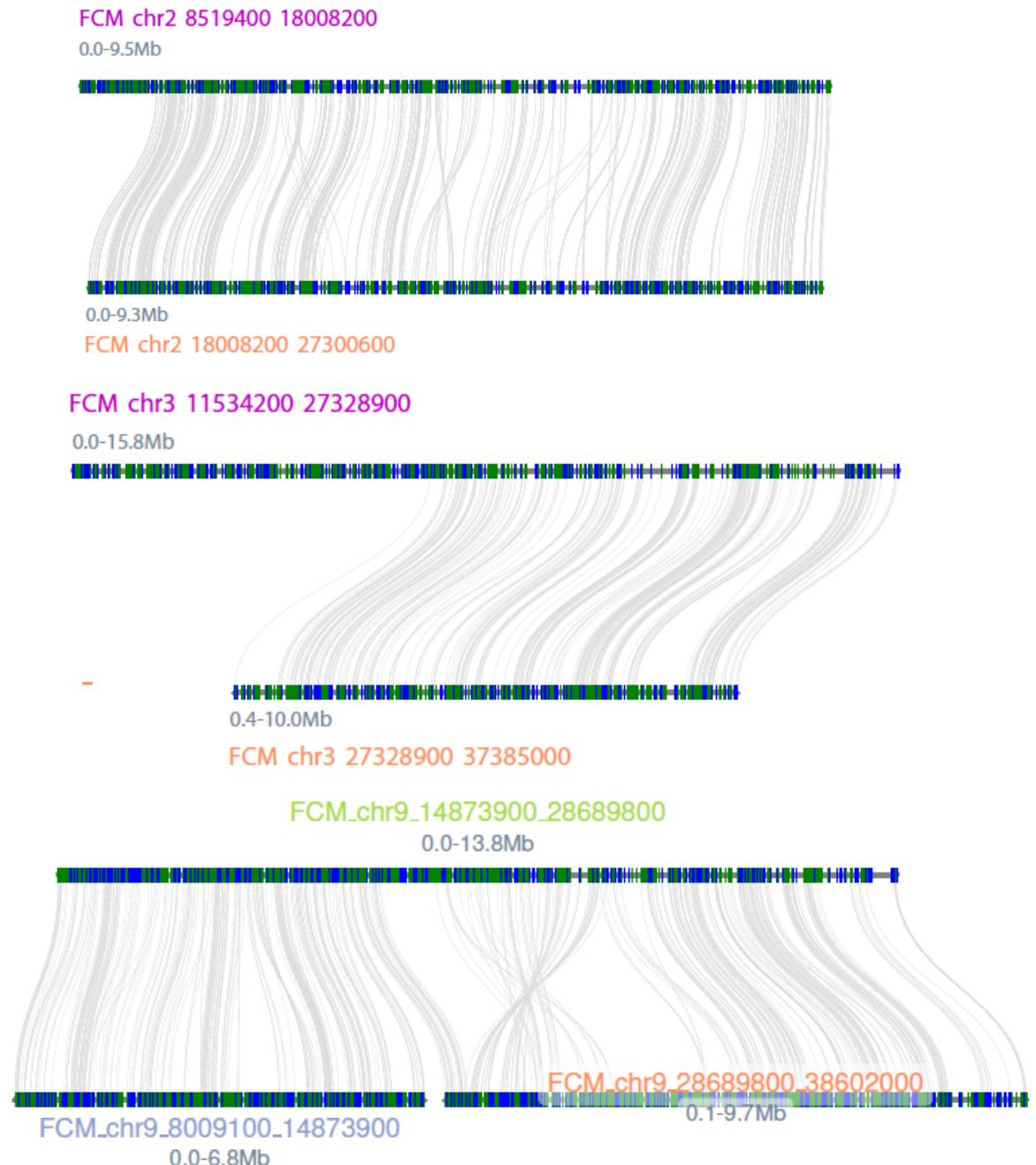
# Clementine



Fairchild



# Gene colinearity between divergent haplotype blocks

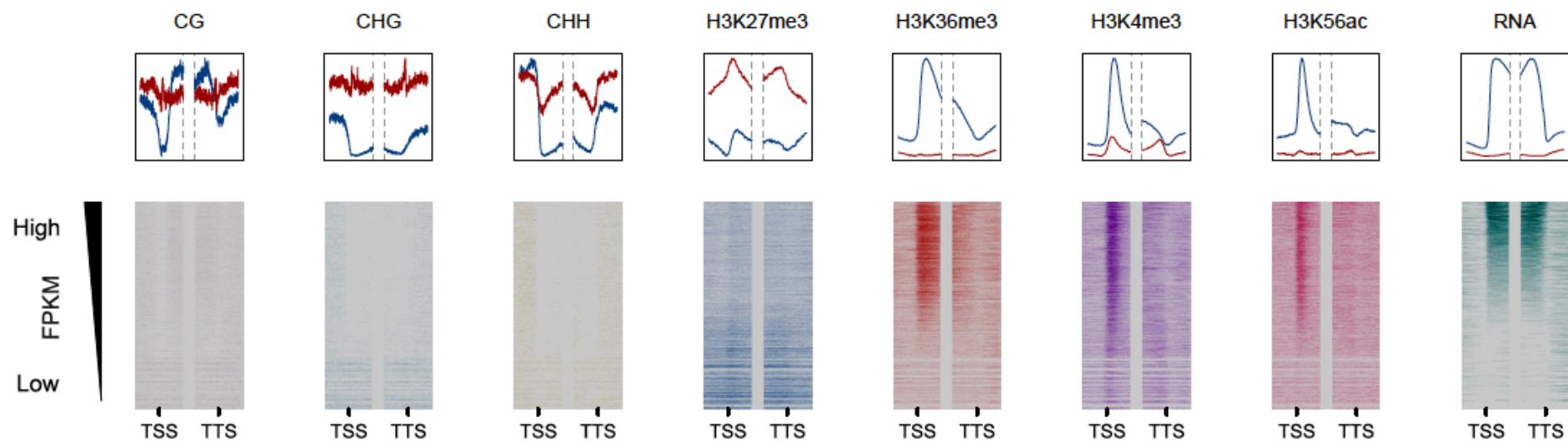




Top 20% gene (highly expressed)



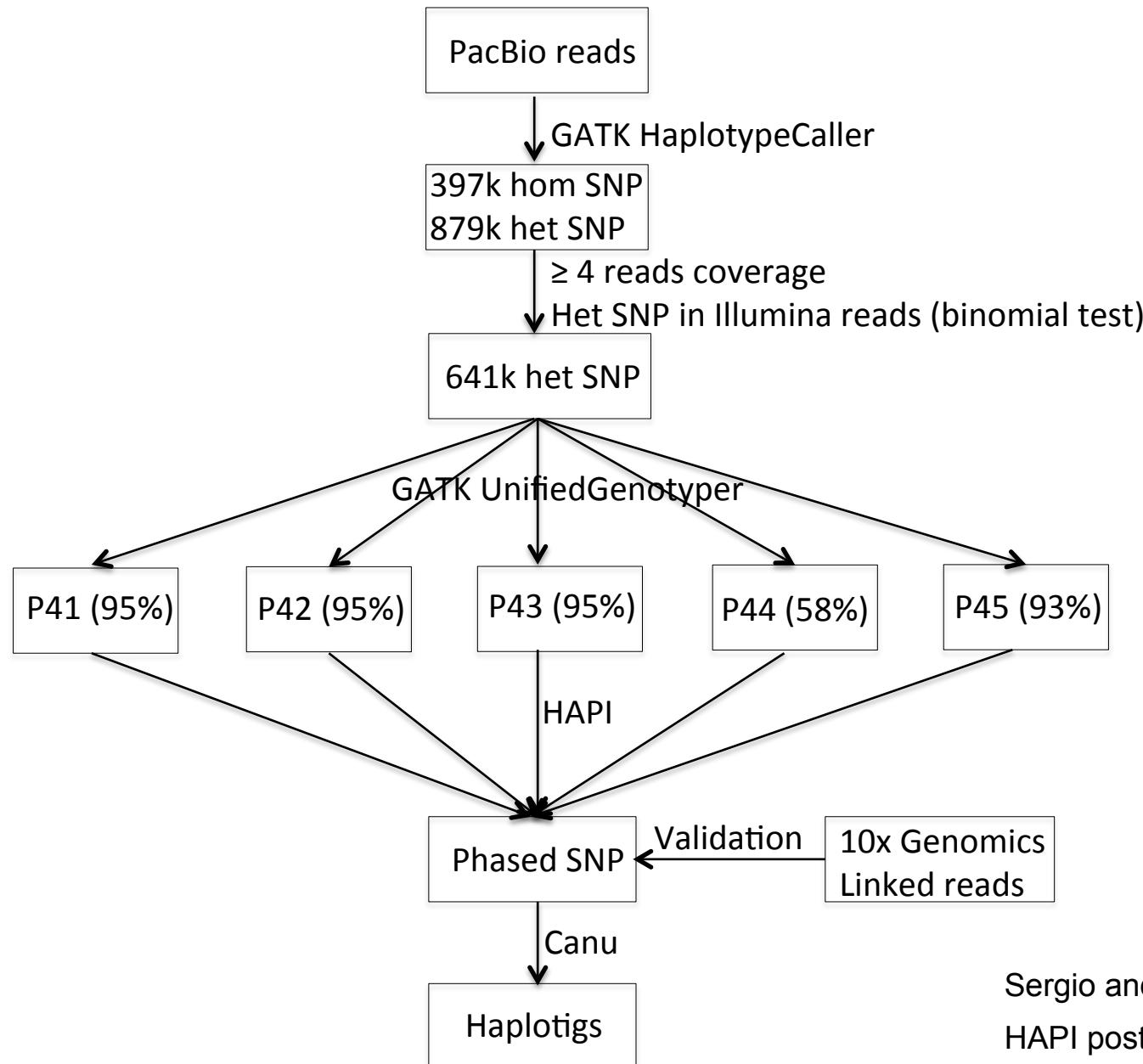
Bottom 20% gene (barely expressed)



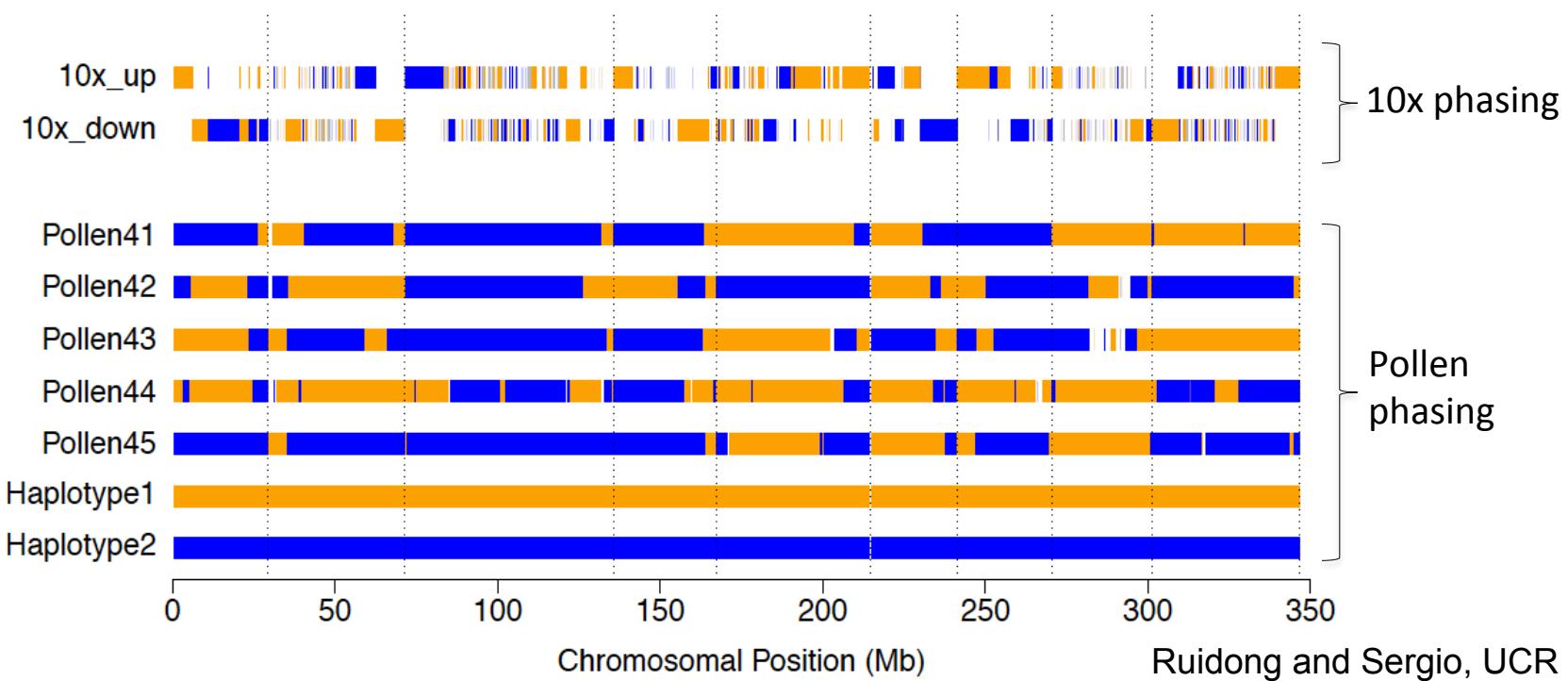
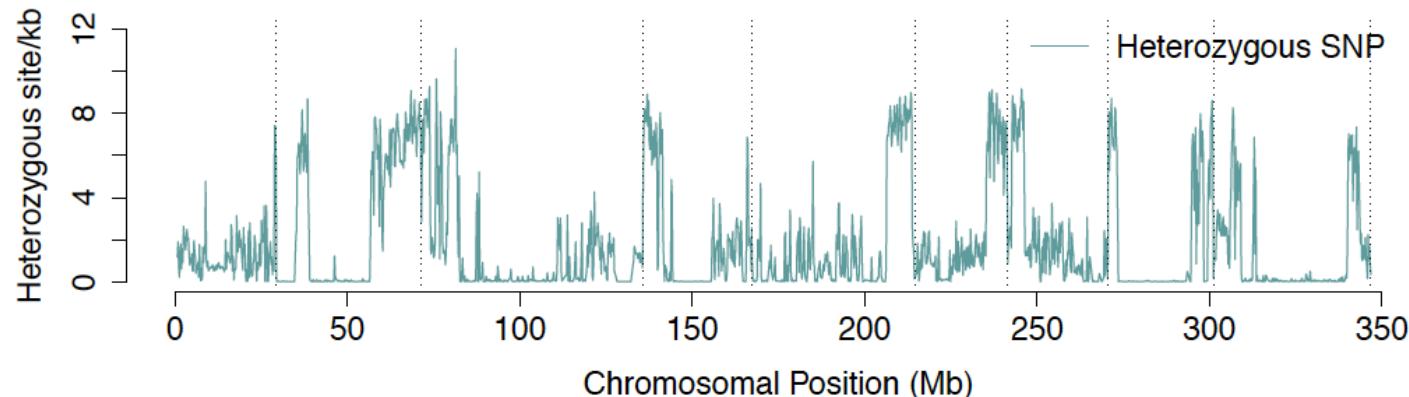
## Summary and future work on Fairchild genome assembly

- Developed a hybrid approach to assemble reference level genome sequence from a diploid
- Discovered several Mb-size sequence rearrangements between Fairchild and Clementine
- Characterize the origin and structure of rearranged sequences

# Genotyping of heterozygous SNPs in pollen sequence



# Construct the haplotypes of Fairchild mandarin



## De novo assemblies of haplotigs

		Fairchild	
Genome type	Diploid	Haplotype1	Haplotype2
Assembled size (Mb)	366	349	162
Gap-free seq (Mb)	358	349	162
Longest scaf (kb)	64121	3716	1515
Longest ctg (kb)	22448	3716	1515
Scaf/ctg N50 (kb)	42136/9996	441/441	174/174
Assembler	Hybrid scaffold	Canu_v1.3	Canu_v1.3
Sequence	120X Pacbio + bionano + 10x	75X Pacbio	14X Pacbio
Gene completeness (CEGMA)	91.9% (228/248)	93.55% (232/248)	70.16% (174/248)
Gene completeness (BUSCO)	94.2% (1357/1440)	92.6% (886/956)	67.1% (642/956)

## Summary and future work on phasing and haplotig assembly

- Used two independent approaches to phase Mb-size haplotype blocks
- Assembled two haplotigs from phased Pacbio long reads
- Construct a haplotype-specific sequence variation map
- Study deleterious mutations within each haplotype
- Expend the deleterious mutation analysis to Mandarin population

# Acknowledgements

Wessler lab:	Jiang lab from MSU:	UCR Bioinformatics Core:
Dr. James Burnette III	Shujun Ou	Jordan Hayes
Travis R. Wrightsman	Wing lab from AGI:	Neerja Katiya
Dr. Venki Chetty	Dr. David Kudrna	UCR Genomics Core:
Roose lab:	Dr. Dario Copetti	Dr. Glenn Hicks
Dr. Sergio Pietro Ferrante	Luo lab from UC Davis:	Holly Eckelhoefer
Yoko Eck	Dr. Tingting Zhu	Clay Clark
Jia lab:	National Clonal Germplasm Repository for Citrus:	
Ruidong Li	Robert R Krueger	

