

Understanding a flower colour hybrid zone - from a polygenic perspective

Nick Barton, Parvathy Surendranadh, IST Austria

David Field, Macquarie Univ., Sydney

Understanding a flower colour hybrid zone - from a polygenic perspective

Nick Barton, Parvathy Surendranadh, IST Austria

David Field, Macquarie Univ., Sydney

Tavares et al., *PNAS*, 2018

Ellis et al., *Mol. Ecol. Res.*, 2018

Ringbauer et al., *Genetics*, 2018

Surendranadh et al., *Genetics*, 2022



Institute of Science and Technology

Understanding a flower colour hybrid zone - from a polygenic perspective

Nick Barton, Parvathy Surendranadh, IST Austria

David Field, Macquarie Univ., Sydney

Tavares et al., *PNAS*, 2018

Ellis et al., *Mol. Ecol. Res.*, 2018

Ringbauer et al., *Genetics*, 2018

Surendranadh et al., *Genetics*, 2022

Frank Chan, Univ. Groningen

Daria Shipilina, Uppsala Univ.

Sean Stankowski, Arka Pal, ISTA

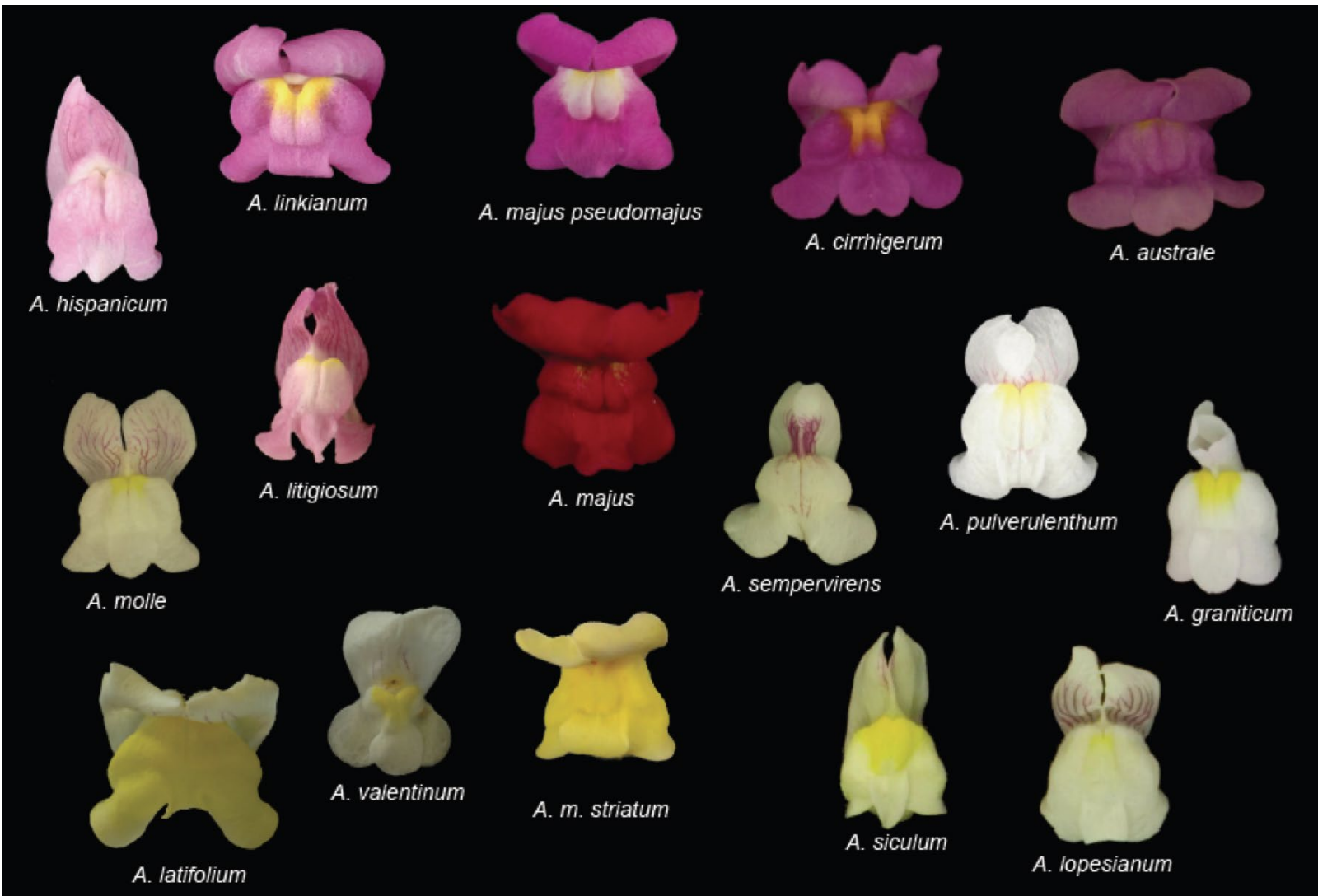


Institute of Science and Technology

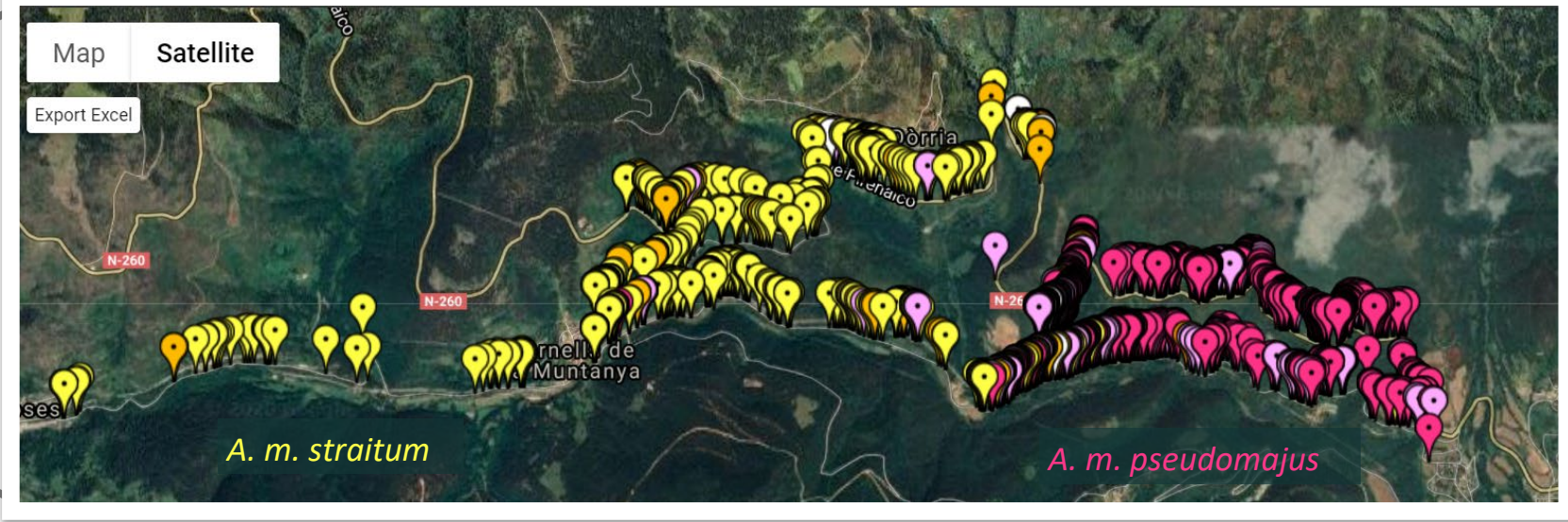
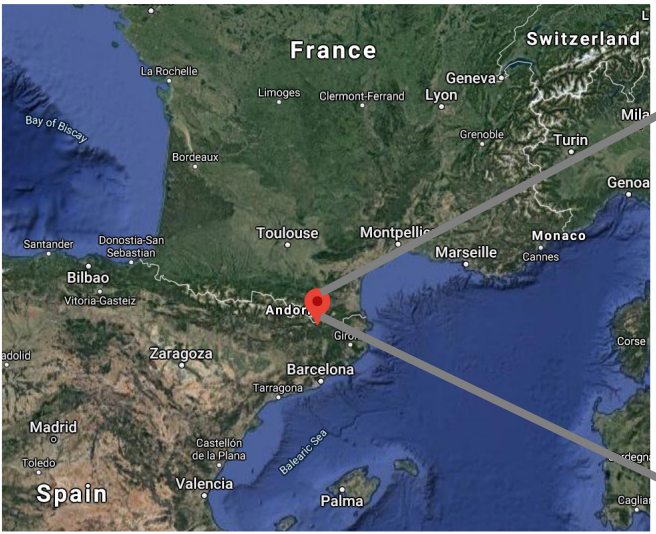
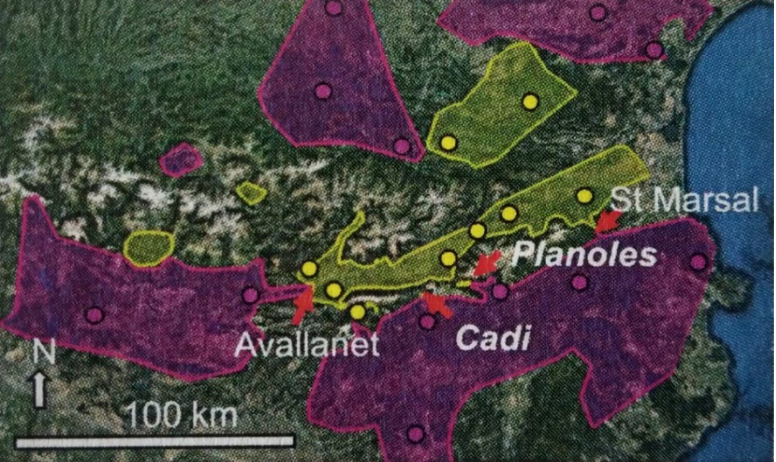
Overview

- Differences in flower pattern depend on a few genes
- Selection can maintain sharp *clines*
- Mixing generates associations between genes (*linkage disequilibrium*)
- These associations cause a barrier to gene flow
- Cline width and associations can be used to estimate selection and dispersal
- We can measure selection and dispersal *directly* from the pedigree
- Are we missing polygenic adaptation?

Flower colour in *A. majus*

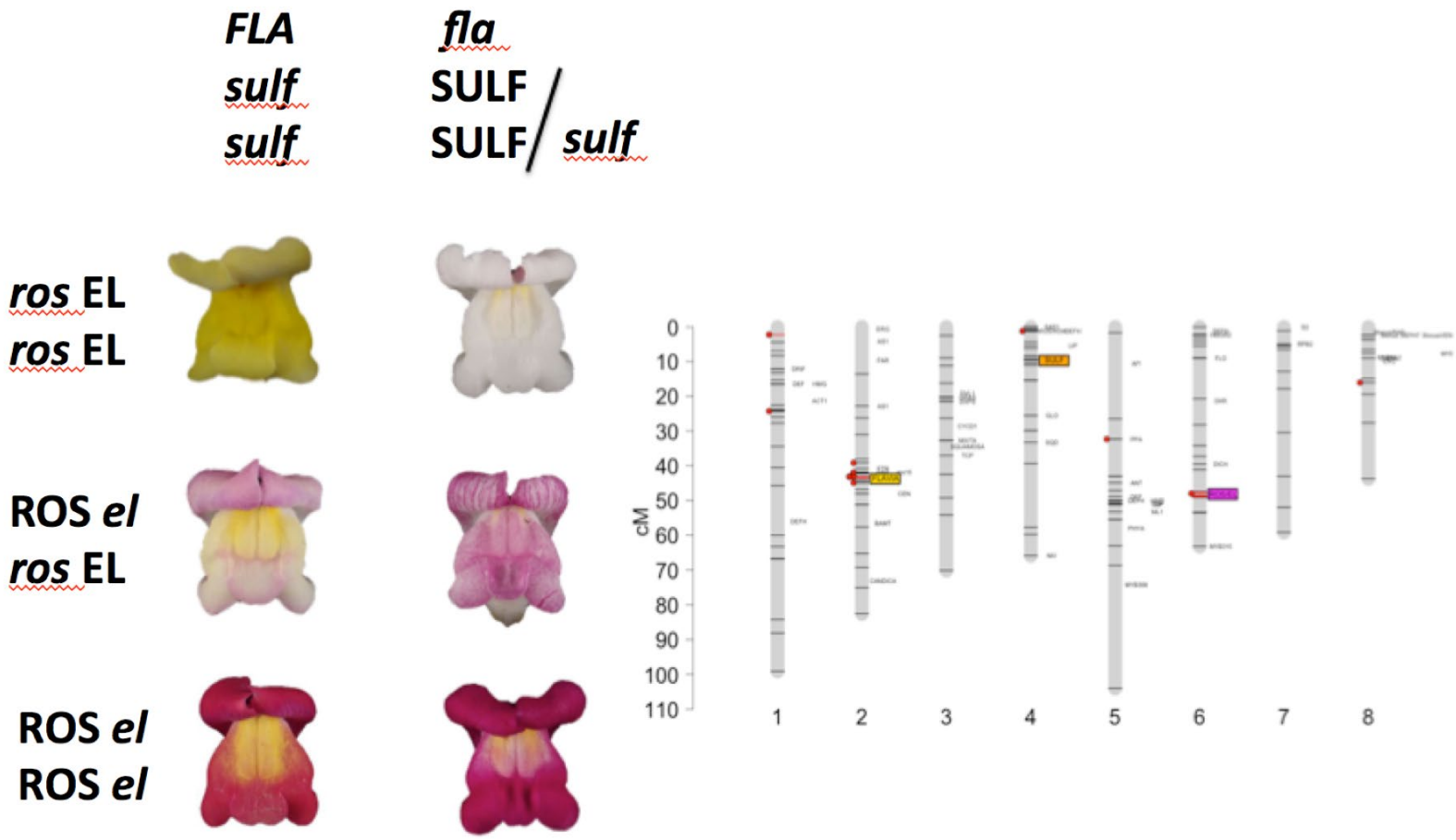


Flower colour hybrid zone in *A. majus*



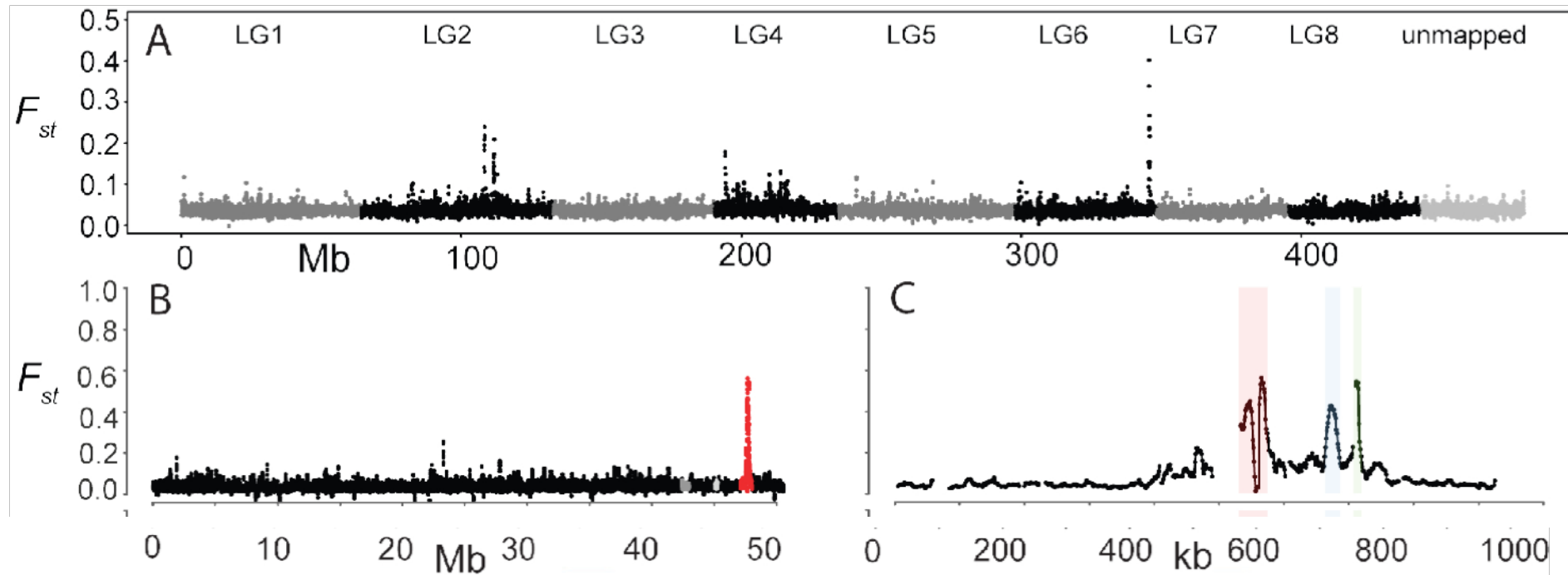
Genetics of flower colour

- *Rosea* and *Eluta* determine magenta anthocyanin pigments
- *Flavia* and *Sulfurea* determine yellow aurones



Scanning the genome

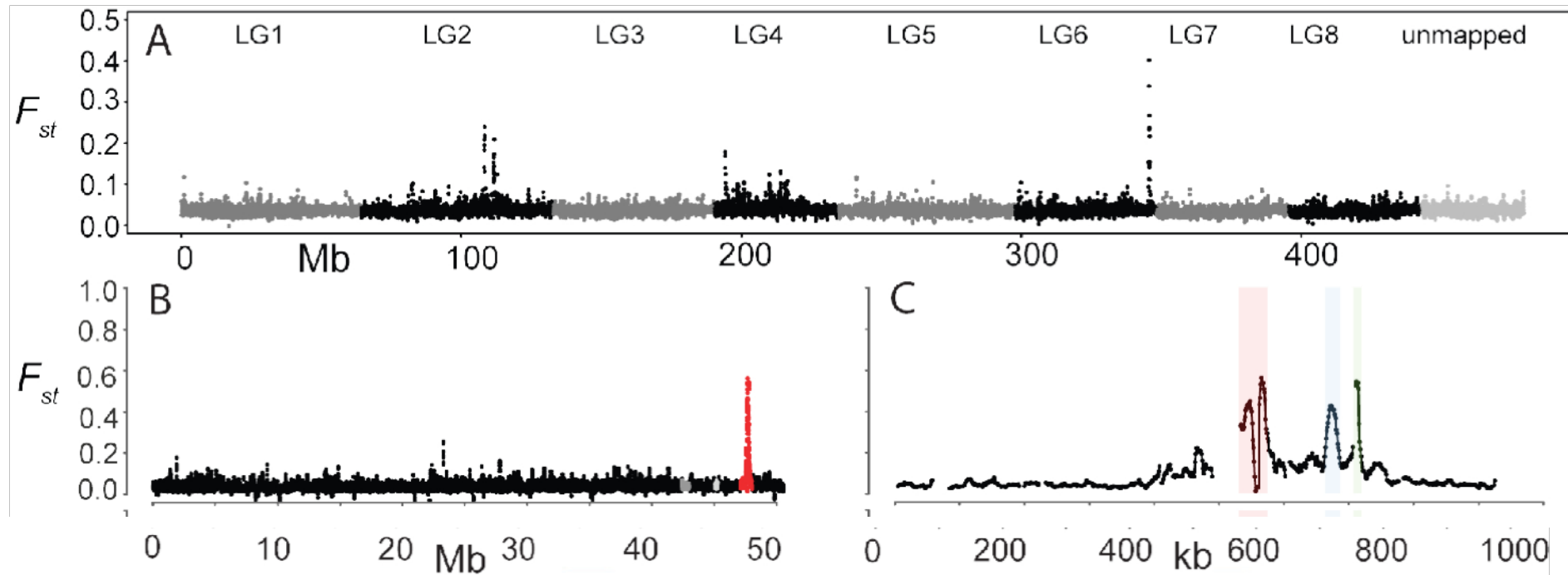
- Pooled sequences 2Km apart across the hybrid zone
- Increased relative divergence (F_{st}) indicates flower colour loci



- e.g. *Rosea/Eluta* (0.5cM apart) on LG6

Scanning the genome

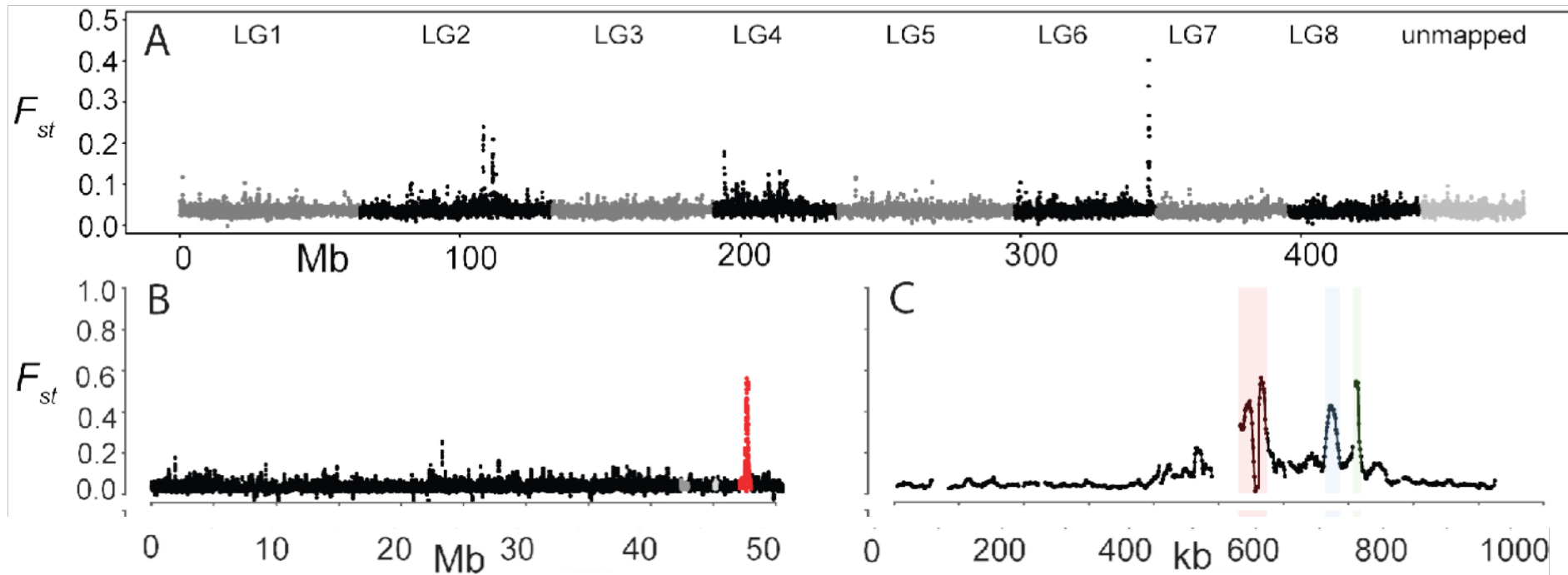
- Pooled sequences 2Km apart across the hybrid zone
- Increased relative divergence (F_{st}) indicates flower colour loci
- Narrow clines, inferred from 8 pools, also indicate these loci



- e.g. *Rosea/Eluta* (0.5cM apart) on LG6

Scanning the genome

- F_{st} measures divergence between populations relative to diversity within:
$$F_{st} = \frac{\pi_B - \pi_W}{\pi_B + \pi_W}$$
- Sharp peaks in F_{st} are due to reduced diversity (selective sweeps?) (Cruikshank & Hahn, 2013)
- A broader increase may be due to a barrier to gene flow (Tavares et al., 2018)

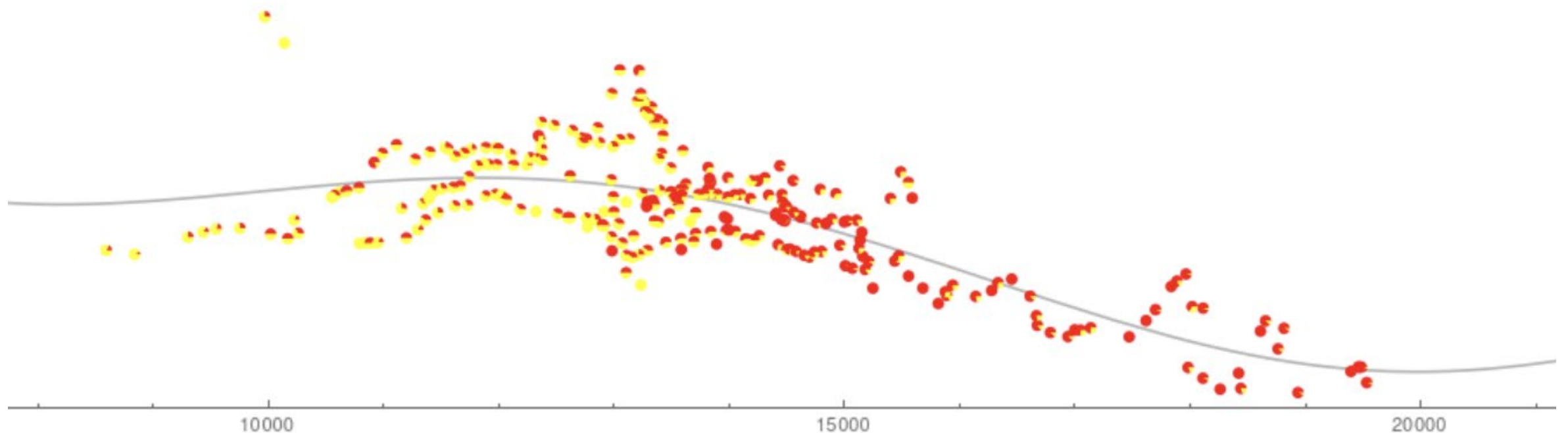


- e.g. *Rosea/Eluta* (0.5cM apart) on LG6

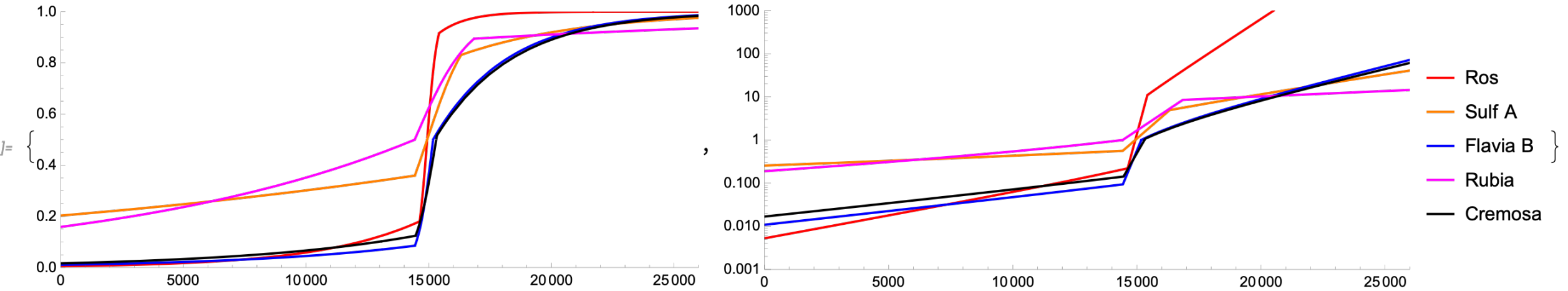
Flower colour SNPs

- Flower colour SNPs from 5 chromosomes
- 22500 individuals from 11 years
- 999 demes (25m); 343 of at least 10 plants

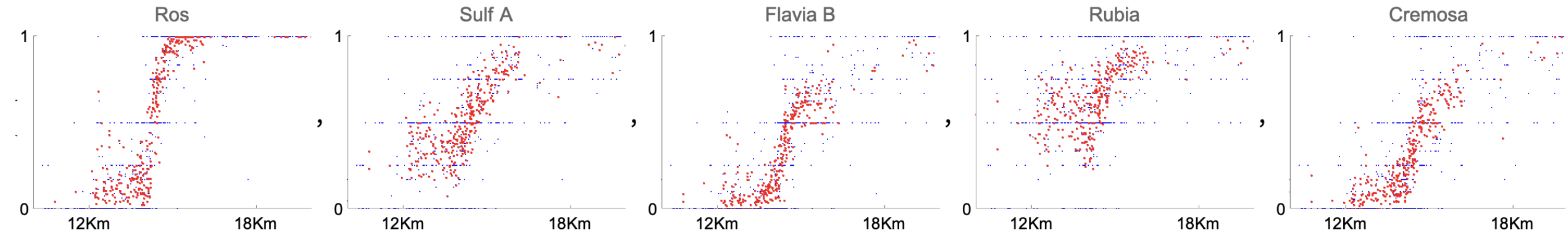
SNP	Chromosome	cM
Rosea /Eluta	6	49.29
Sulfurea	4	9.37
Flavia	2	43.18
Rubia	5	32.32
Cremosa	1	2.18



Fitting clines at five loci

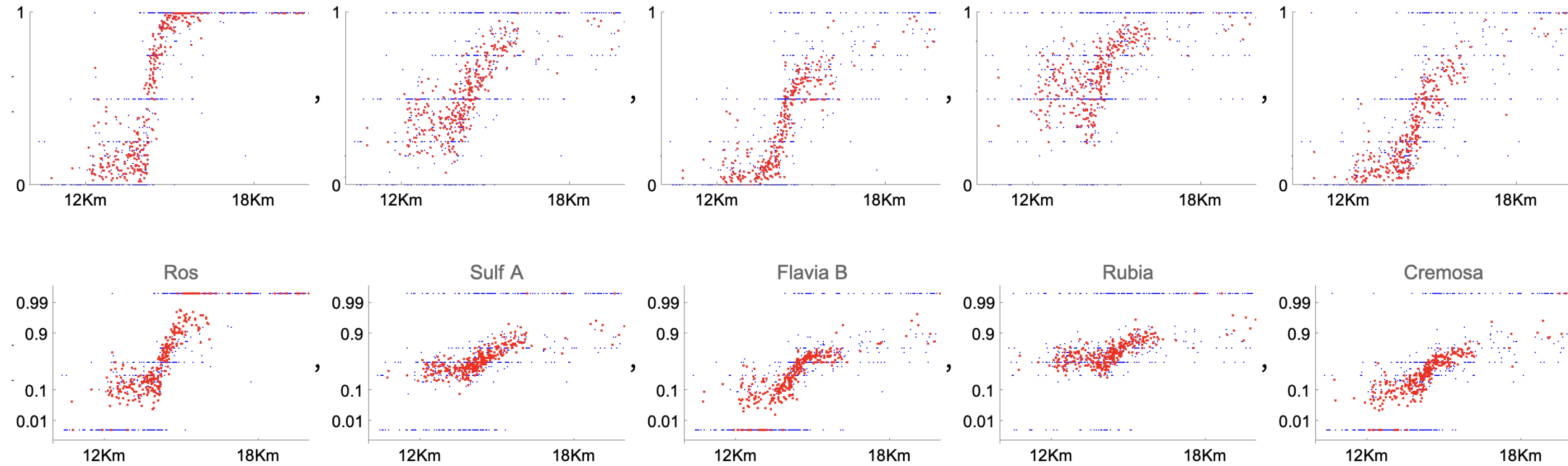


Clines at five loci



blue: all demes; red: demes with at least 10 plants

Clines at five loci



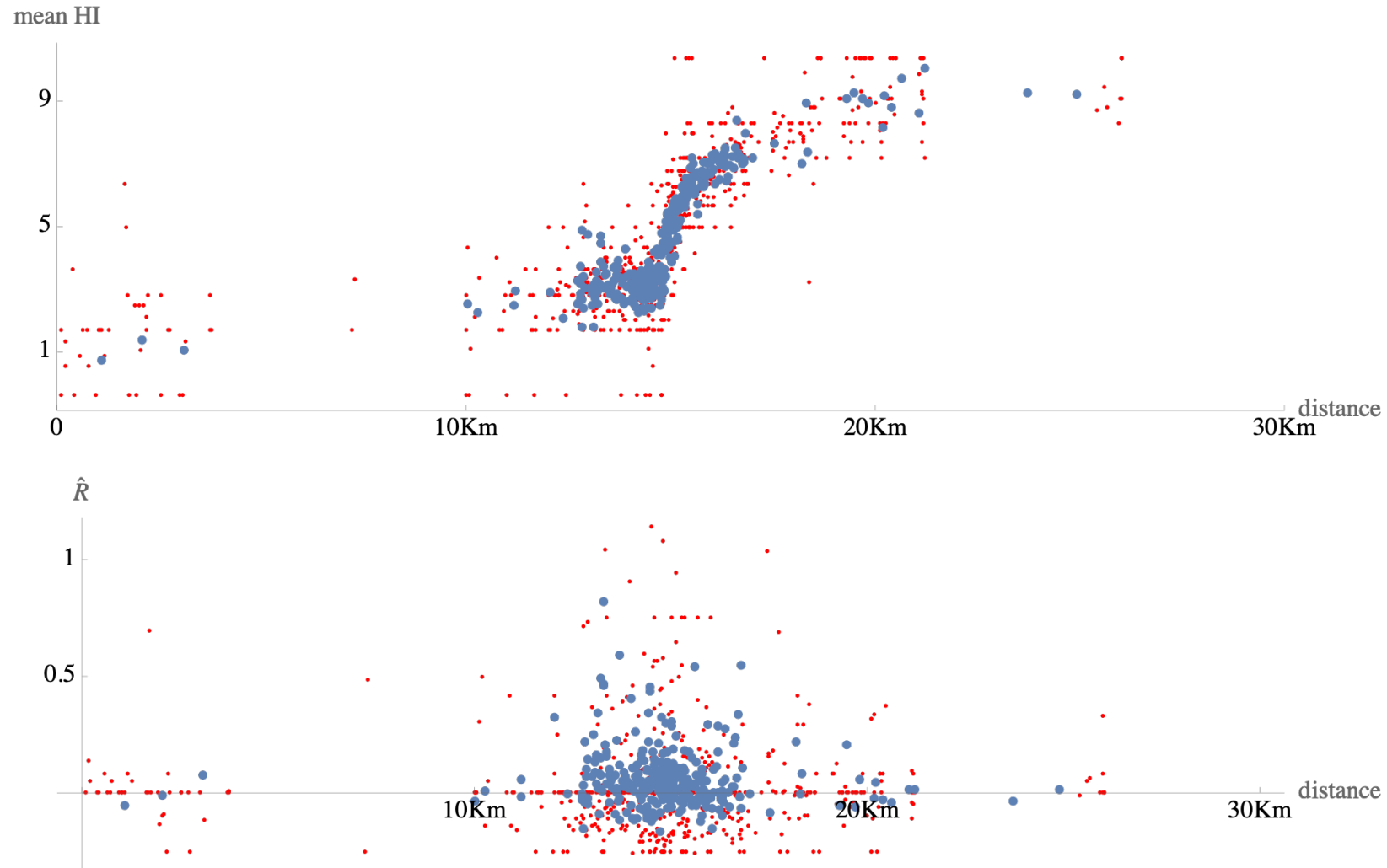
blue: all demes; red: demes with at least 10 plants

Hybrid index, based on five unlinked loci

Top: mean HI, from five unlinked loci, on a $\log(p/q)$ scale

Bottom: excess variance in HI => LD

blue: demes ≥ 10 individuals; red: demes < 10 individuals.



Inference from cline width and associations between loci

Clines coincide, and have widths $\sim 750\text{m}$ (Ros), $\sim 1000\text{m}$ (Sulf) ...

Selection coefficient can be estimated from the cline widths, $w = \sigma\sqrt{8/s}$

Inference from cline width and associations between loci

Clines coincide, and have widths $\sim 750\text{m}$ (Ros), $\sim 1000\text{m}$ (Sulf) ...

Selection coefficient can be estimated from the cline widths, $w = \sigma\sqrt{8/s}$

e.g. if $\sigma = 100\text{m}$. $w=750\text{m}$, then $s \sim 14\%$

e.g. if $\sigma = 50\text{m}$. $w=750\text{m}$, then $s \sim 3.5\%$

Inference from cline width and associations between loci

Clines coincide, and have widths $\sim 750\text{m}$ (Ros), $\sim 1000\text{m}$ (Sulf) ...

Selection coefficient can be estimated from the cline widths, $w = \sigma\sqrt{8/s}$

e.g. if $\sigma = 100\text{m}$. $w=750\text{m}$, then $s \sim 14\%$

e.g. if $\sigma = 50\text{m}$. $w=750\text{m}$, then $s \sim 3.5\%$

From the pedigree, $\sigma \sim 150\text{m}$, but is highly *leptokurtic* - and so in effect, much smaller

Inference from cline width and associations between loci

Clines coincide, and have widths $\sim 750\text{m}$ (Ros), $\sim 1000\text{m}$ (Sulf) ...

Selection coefficient can be estimated from the cline widths, $w = \sigma\sqrt{8/s}$

e.g. if $\sigma = 100\text{m}$. $w=750\text{m}$, then $s \sim 14\%$

e.g. if $\sigma = 50\text{m}$. $w=750\text{m}$, then $s \sim 3.5\%$

From correlations between loci in the centre, $R=0.033$, $\sigma = w \sqrt{R/8} \sim 50 \text{ m}$

Inference from cline width and associations between loci

Clines coincide, and have widths $\sim 750\text{m}$ (Ros), $\sim 1000\text{m}$ (Sulf) ...

Selection coefficient can be estimated from the cline widths, $w = \sigma\sqrt{8/s}$

e.g. if $\sigma = 100\text{m}$. $w=750\text{m}$, then $s \sim 14\%$

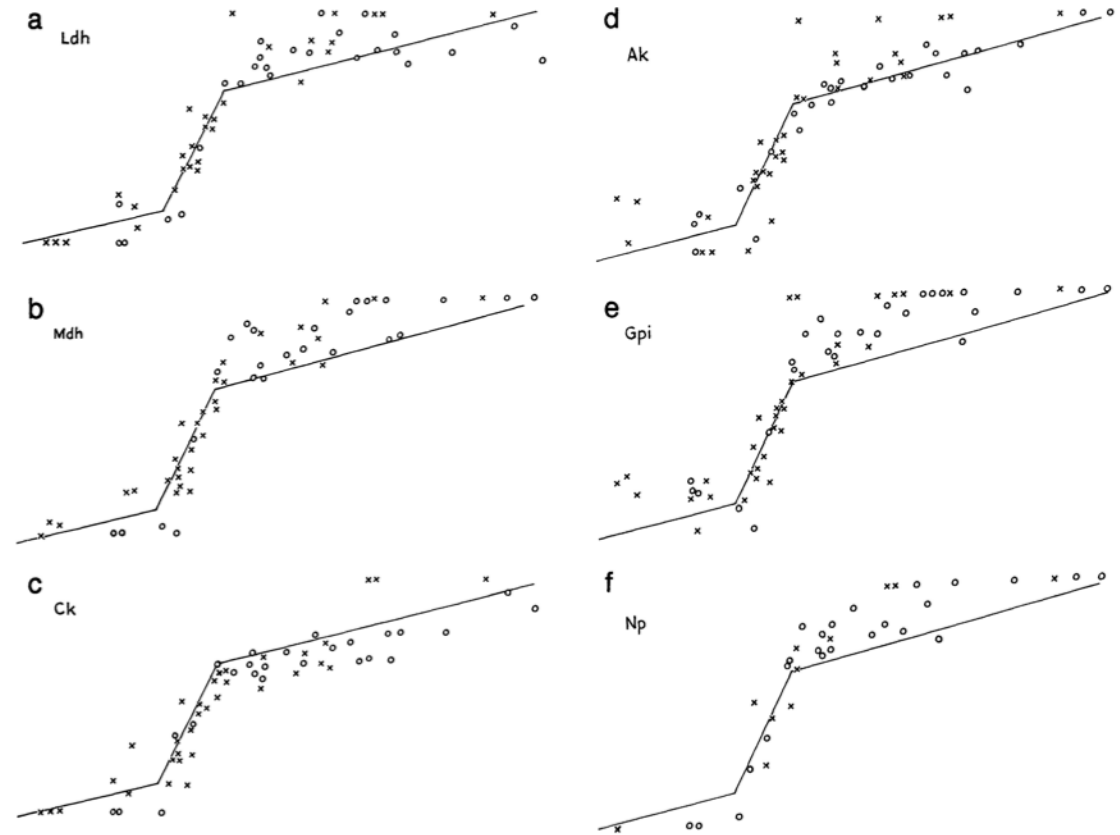
e.g. if $\sigma = 50\text{m}$. $w=750\text{m}$, then $s \sim 3.5\%$

From correlations between loci in the centre, $R=0.033$, $\sigma = w \sqrt{R/8} \sim 50 \text{ m}$

Can we explain the width, and the stepped *shape* of the clines simply from selection on 5 loci, plus dispersal?

An interlude with *Bombina bombina/variiegata*

Szymura & B, 1986, 1991



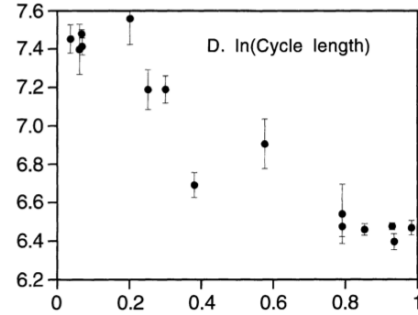
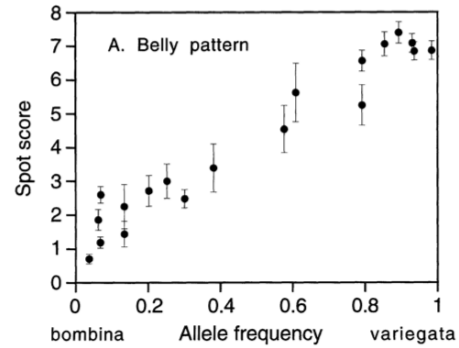
Stepped clines at 6 allozyme markers, ~ 6Km wide; strong LD, $R \sim 0.22$

From LD, infer dispersal $\sigma \sim 1\text{Km}$, and selection on individual alleles $s \sim 0.16/0.37\%$

If the step is due to linked selection, mean fitness ~ 0.6 , with $n \sim 55/300$ loci under selection

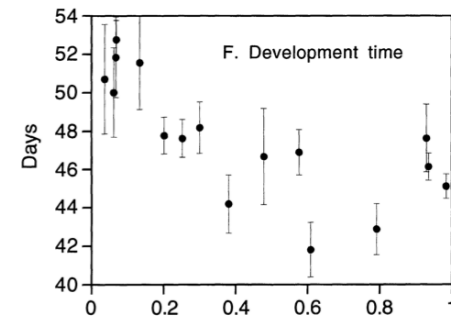
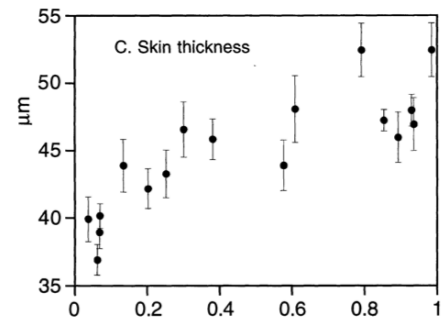
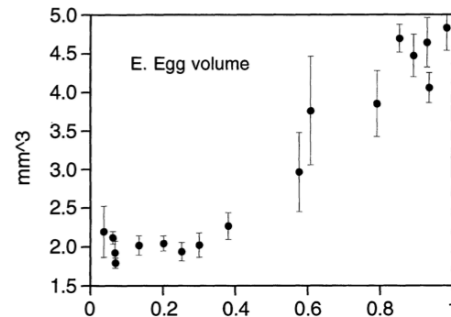
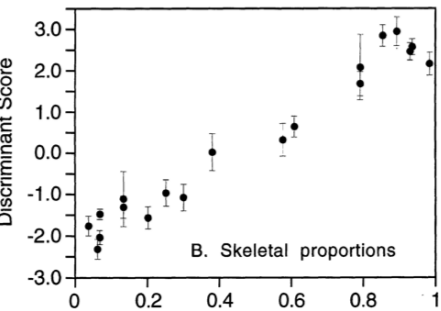
An interlude with *Bombina bombina/variiegata*

Nürnbergger et al., 1995



Δz	Enzyme 8.00	Spot score 6.31	Egg volume 3.09	ln(cycle length) -1.18	Skin thickness 11.10	Discrim. score 5.19
Enzyme	0.8677 (122) 4.2369 (109) 0.5081 (70)	0.3290 (122) 2.6226 (109)* -0.0708 (70)	-0.0398 (46) 0.8061 (46)* -0.1063 (30)	-0.0468 (27) -0.2022 (26)* 0.0079 (14)	-0.7986 (108) 0.7379 (96) -0.0559 (57)	0.1263 (96) 1.6990 (86)* -0.0034 (57)
Spot score	0.0130 0.1039 -0.0028	2.1583 (170) 5.5643 (160) 2.0029 (113)	0.0267 (52) 0.5901 (56)* -0.0630 (33)	-0.0037 (37) -0.1791 (33) -0.0810 (25)	0.0868 (130) 2.2279 (114) 1.2834 (72)	0.1403 (122) 1.7414 (115)* 0.0156 (84)
Egg volume	-0.0032 0.0653 -0.0086	0.0027 0.0605 -0.0065	0.2015 (52) 0.4670 (56) 0.5168 (33)	— — —	0.1282 (46) 1.9960 (41) 0.4693 (27)	-0.0083 (47) 0.2696 (44) -0.0991 (28)
ln(cycle length)	0.0099 0.0429 -0.0017	0.0010 0.0481 0.0218	— — —	0.0310 (37) 0.0432 (33) 0.0120 (25)	0.0999 (33) -0.0945 (28) -0.1420 (19)	-0.0103 (26) -0.0890 (25) -0.0187 (22)
Skin thickness	-0.0180 0.0166 -0.0013	0.0025 0.0635 0.0366	0.0075 0.1165 0.0274	-0.0153 0.0145 0.0217	36.8830 (130) 50.3659 (114) 41.3663 (72)	-0.0360 (105) 1.3330 (92) 1.8113 (62)
Discrim. score	0.0061 0.0819 -0.0002	0.0086 0.1063 0.0010	-0.0010 0.0337 -0.0124	0.0034 0.0295 0.0061	-0.0013 0.0463 0.0629	0.8580 (122) 1.7435 (115) 1.0538 (84)

LD causes excess covariance between traits



Six quantitative traits are concordant with allozyme markers

Inference from cline width and associations between loci

Clines coincide, and have widths $\sim 750\text{m}$ (Ros), $\sim 1000\text{m}$ (Sulf) ...

Selection coefficient can be estimated from the cline widths, $w = \sigma\sqrt{8/s}$

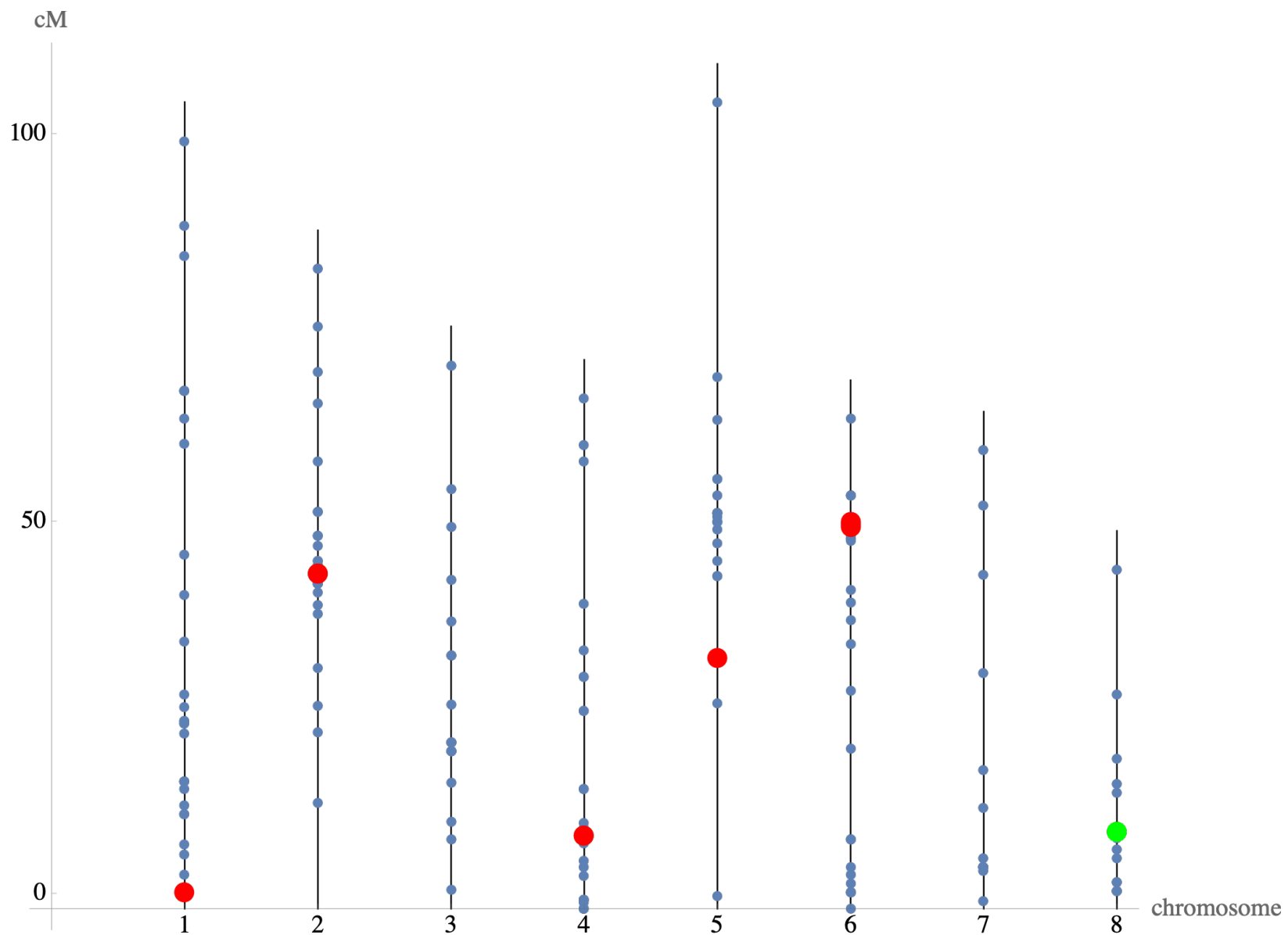
e.g. if $\sigma = 100\text{m}$. $w=750\text{m}$, then $s \sim 14\%$

e.g. if $\sigma = 50\text{m}$. $w=750\text{m}$, then $s \sim 3.5\%$

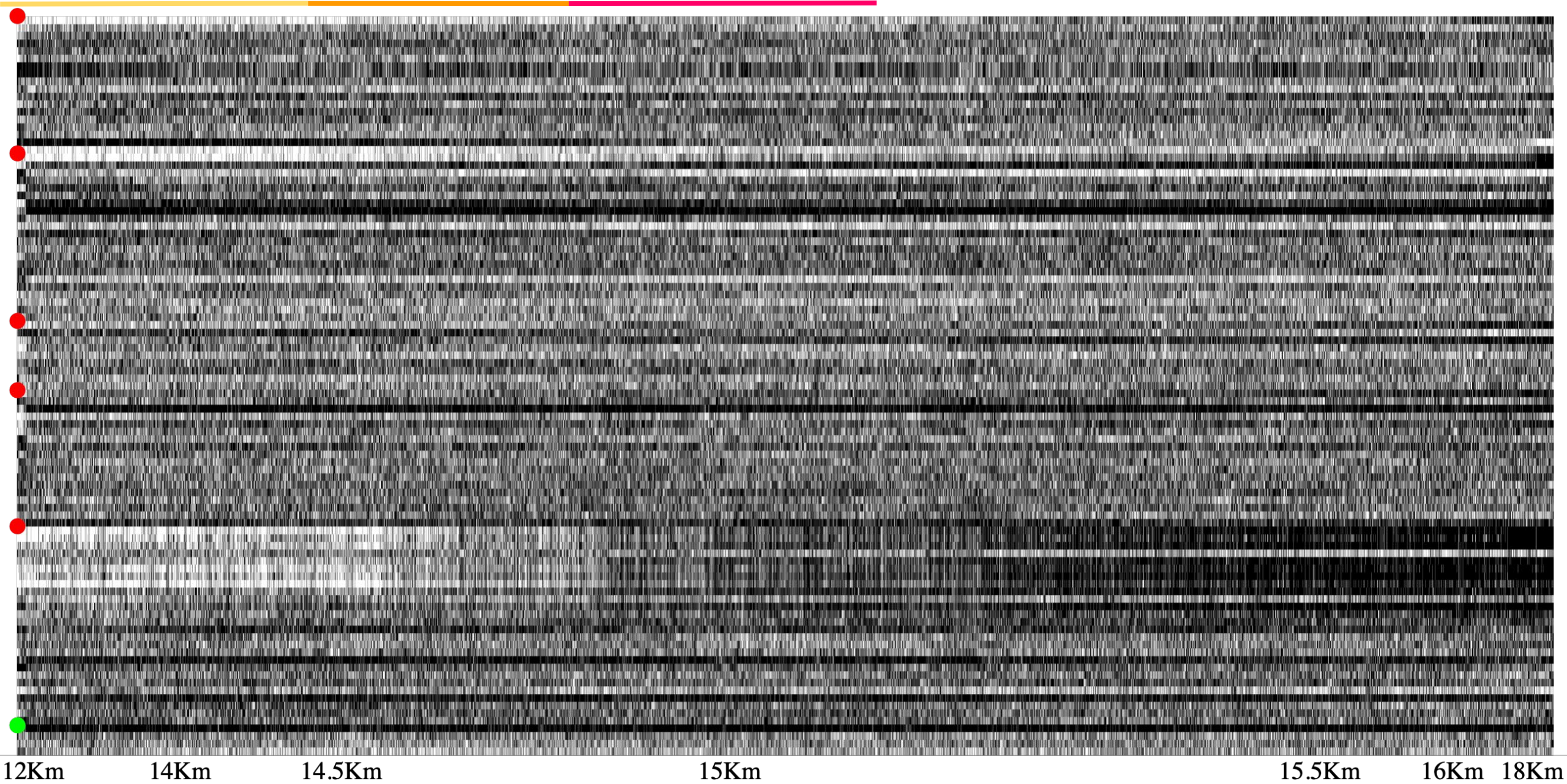
From correlations between loci in the centre, $R=0.033$, $\sigma = w \sqrt{R/8} \sim 50 \text{ m}$

Can we explain the width, and the stepped *shape* of the clines simply from selection on 5 loci, plus dispersal?

Estimate a pedigree from 98 SNP

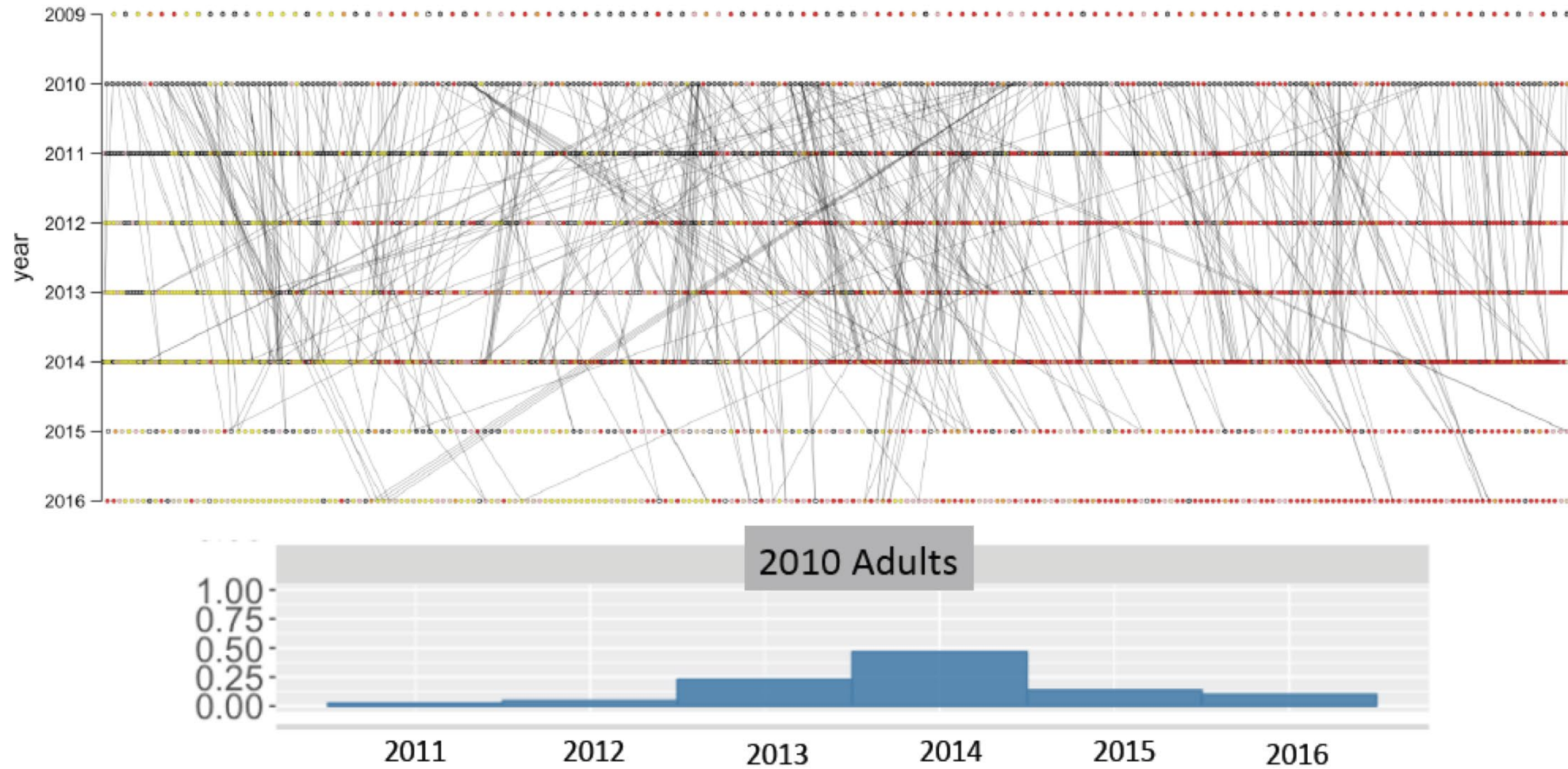


Estimate a pedigree from 98 SNP



The pedigree

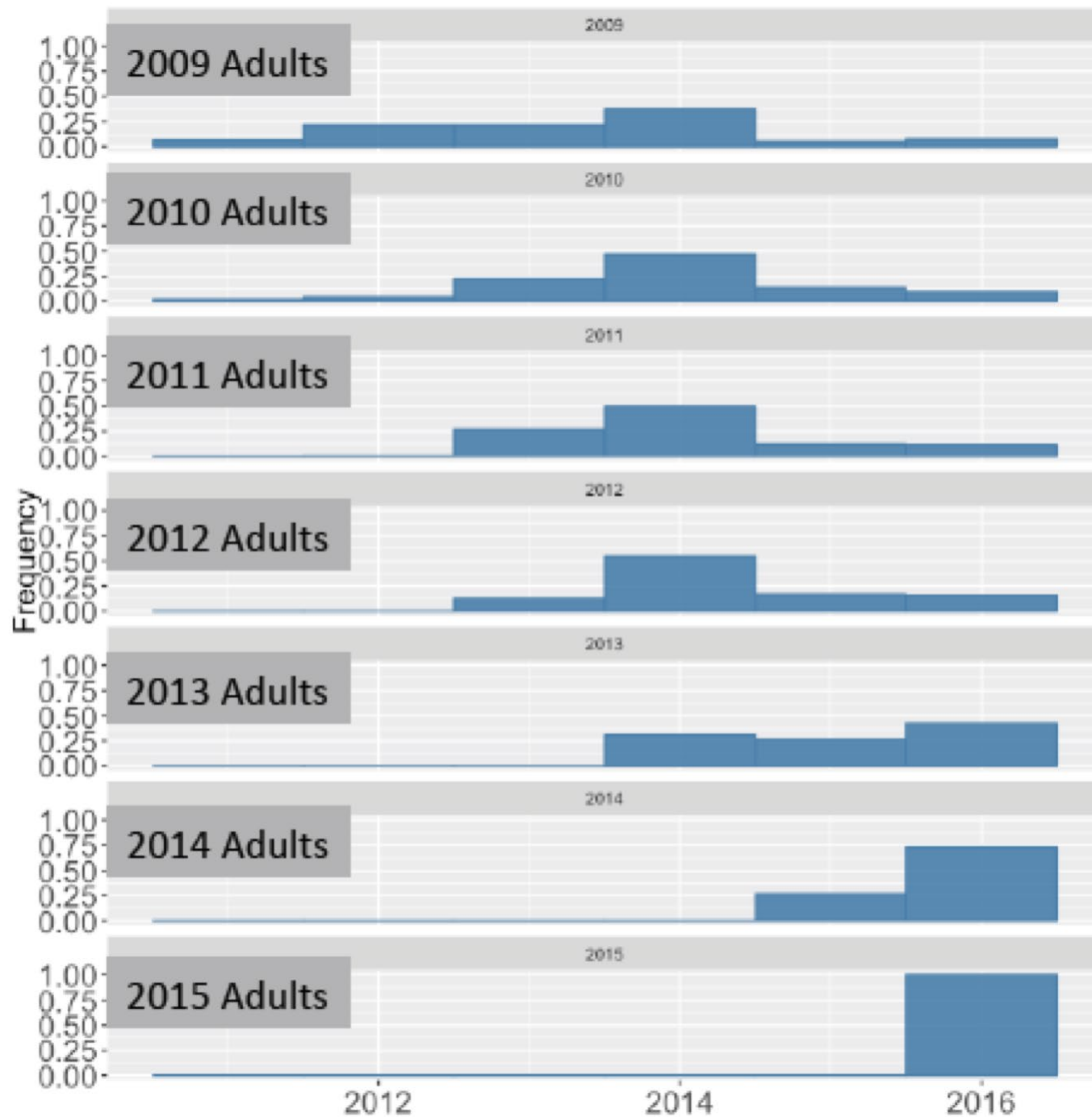
2400 trios inferred from ~ 100 SNP



The pedigree

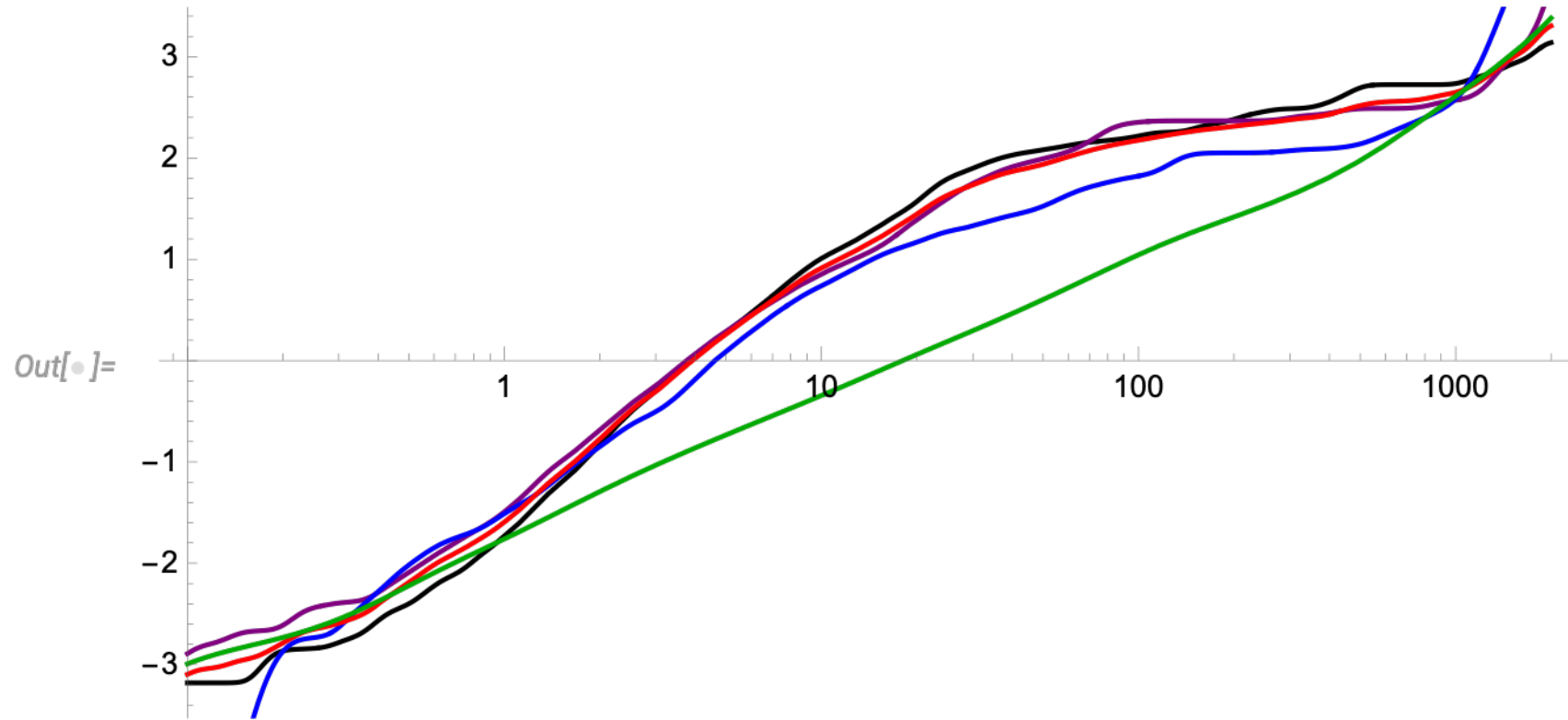
Parents inferred from ~ 100 SNP

A seed bank increases generation time



Seed and pollen dispersal

(red, green)



In addition, unlikely individuals indicate long-range dispersal:

yellow flank (< -2Km): $27/729=3.7\%$

magenta flank (> +1Km): $17/1596=1.1\%$

Simulations:

Follow haplotype frequencies, ignoring random drift

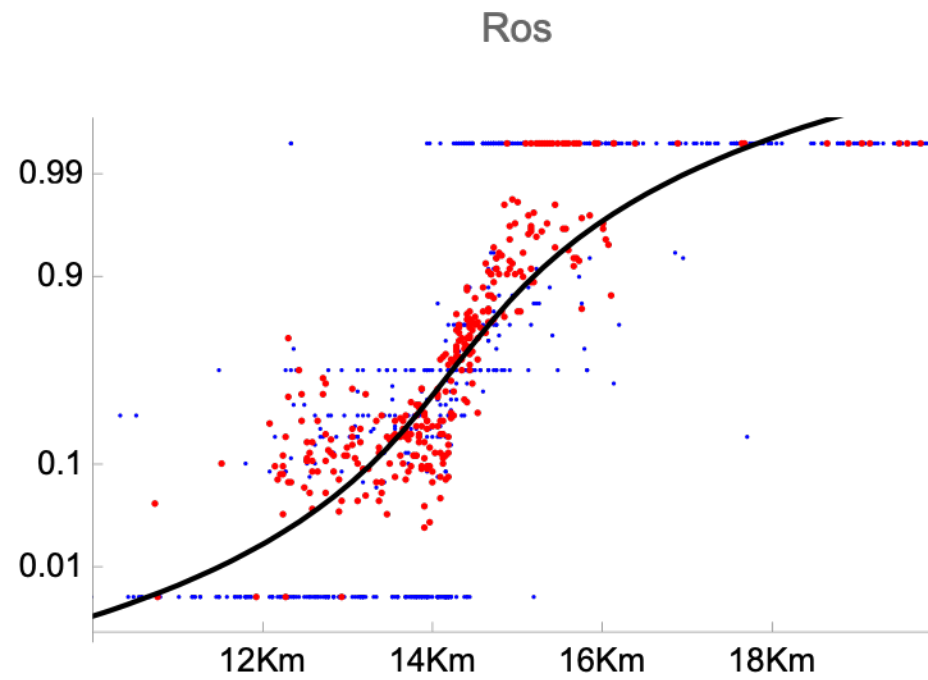
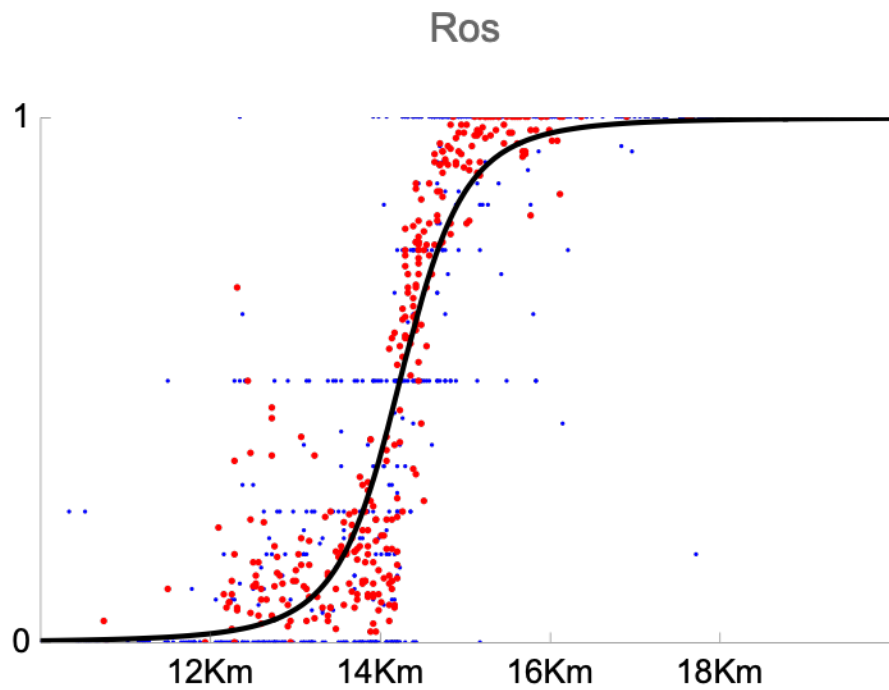
Fit allele frequencies to data, allowing for random fluctuations (F_{st})

Selection against heterozygotes $1+s_0:1:1+s_1$

Cline movement may be balanced by a density gradient

Simulations: Ros

Selection against heterozygotes $1+s:1:1+s$
 $s=0.02$ no tails (black)

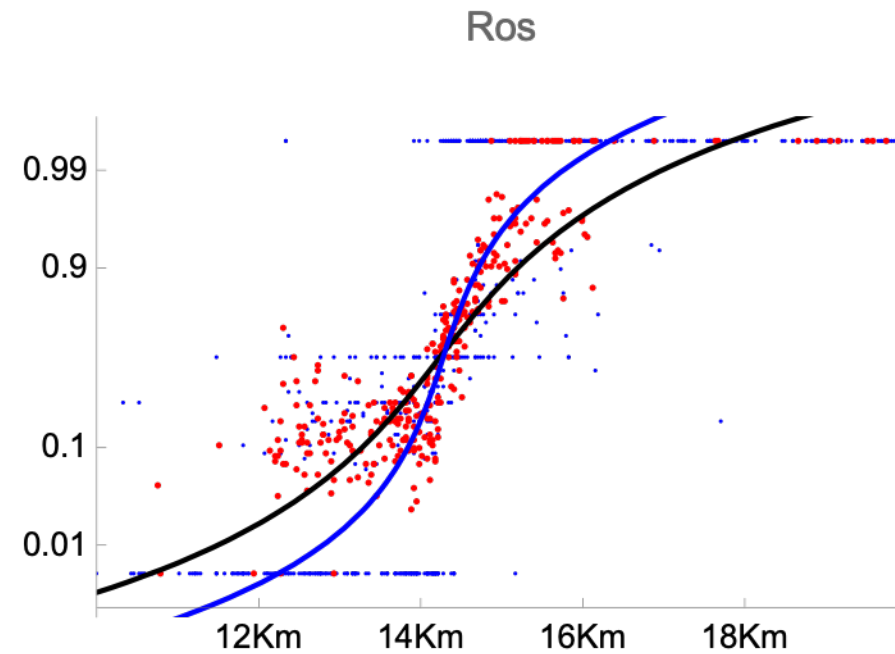
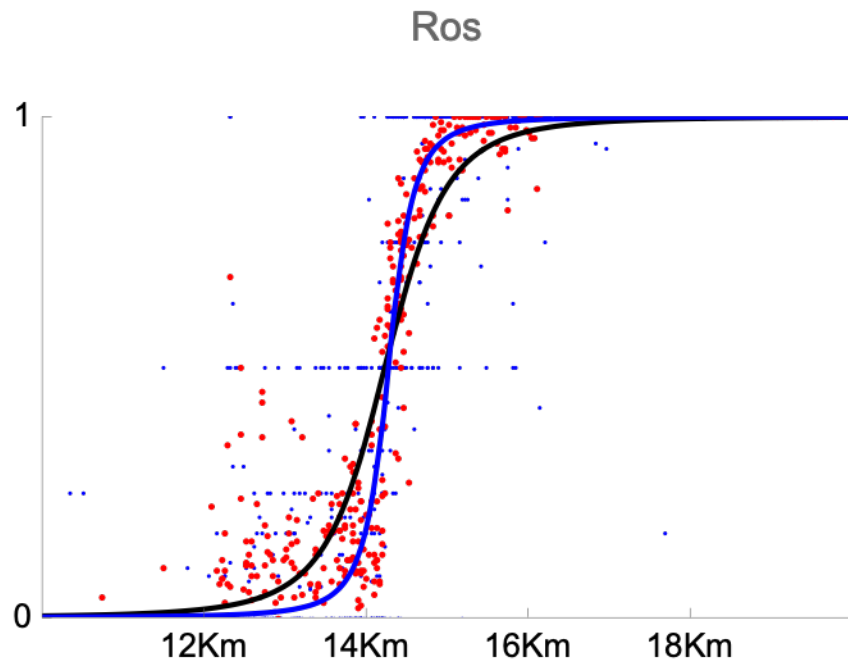


Simulations: Ros

Selection against heterozygotes $1+s:1:1+s$

$s=0.02$ no tails (black)

$s=0.05$ no tails (blue)



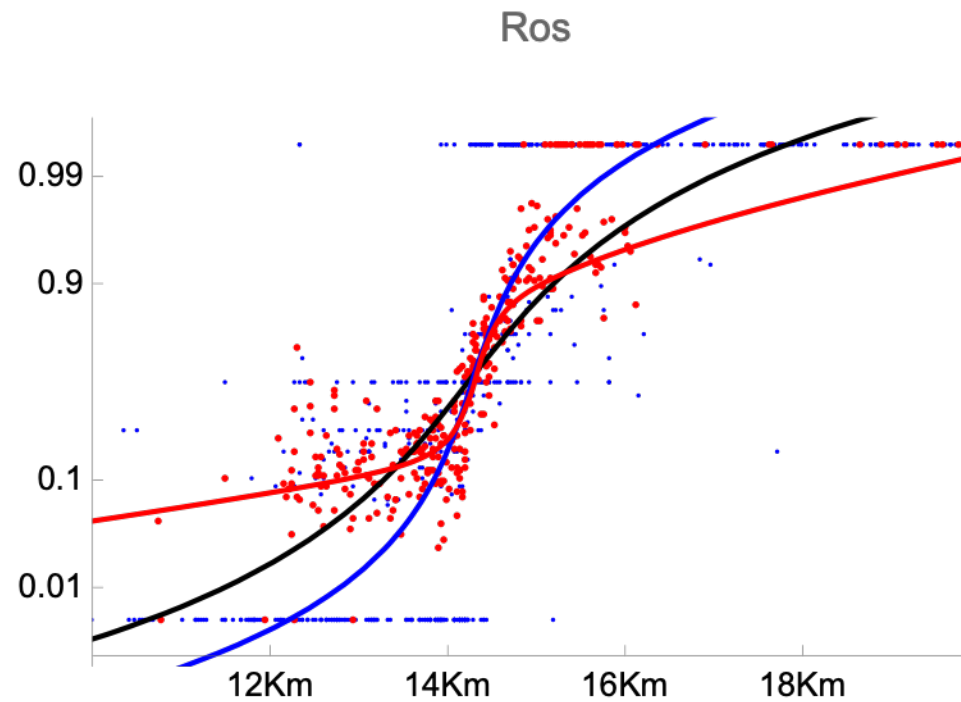
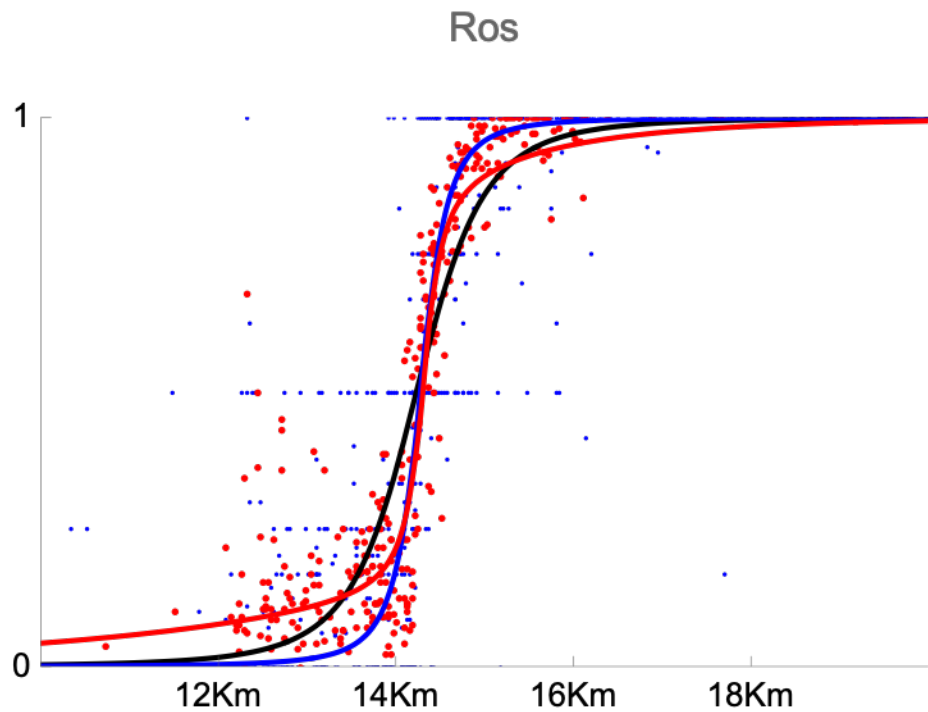
Simulations: Ros

Selection against heterozygotes $1+s:1:1+s$

$s=0.02$ no tails (black)

$s=0.05$ no tails (blue)

$s=0.19$ with tails (red)



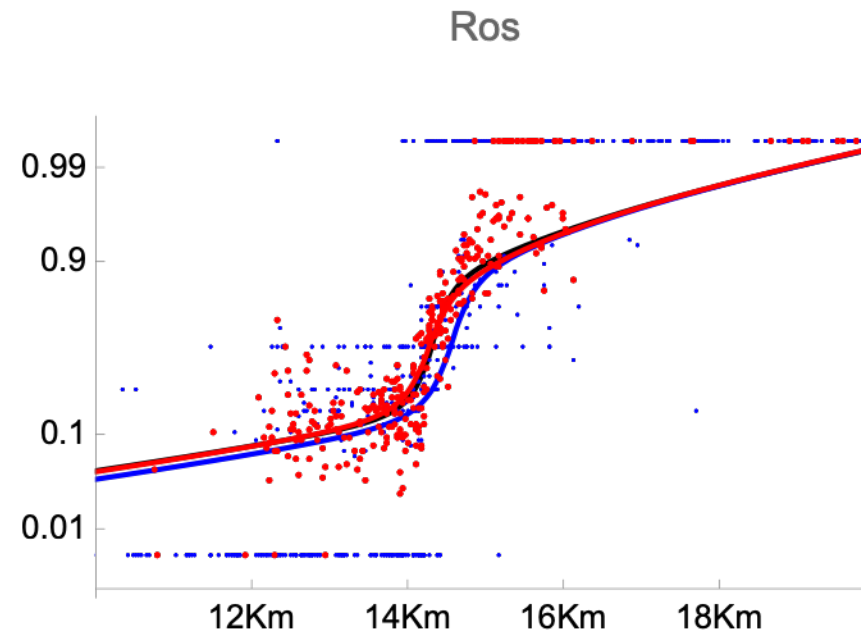
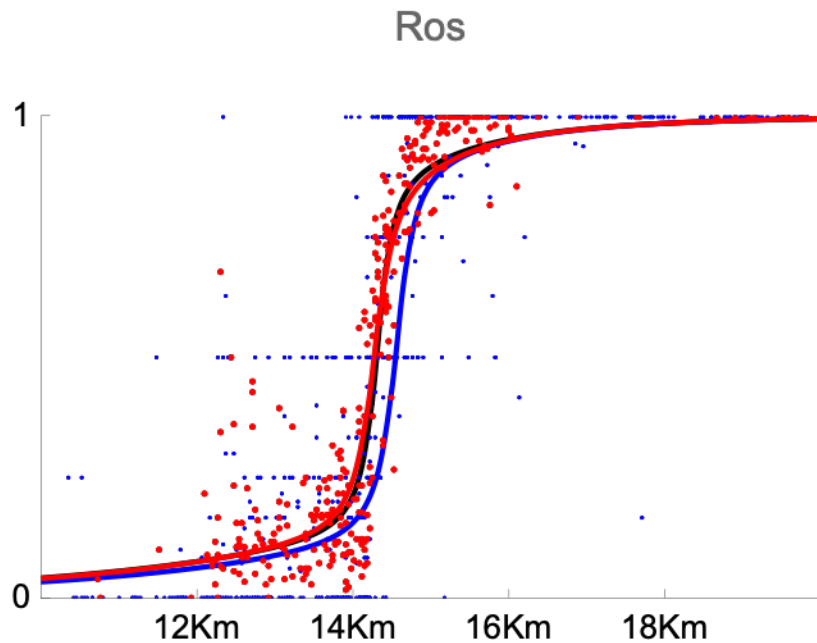
Simulations: five loci

Selection against heterozygotes $1+s_0: 1: 1+s_1$

individual loci (LE; black)

multiple loci, symmetric (LD; blue)

multiple loci, asymmetric (LD; red)



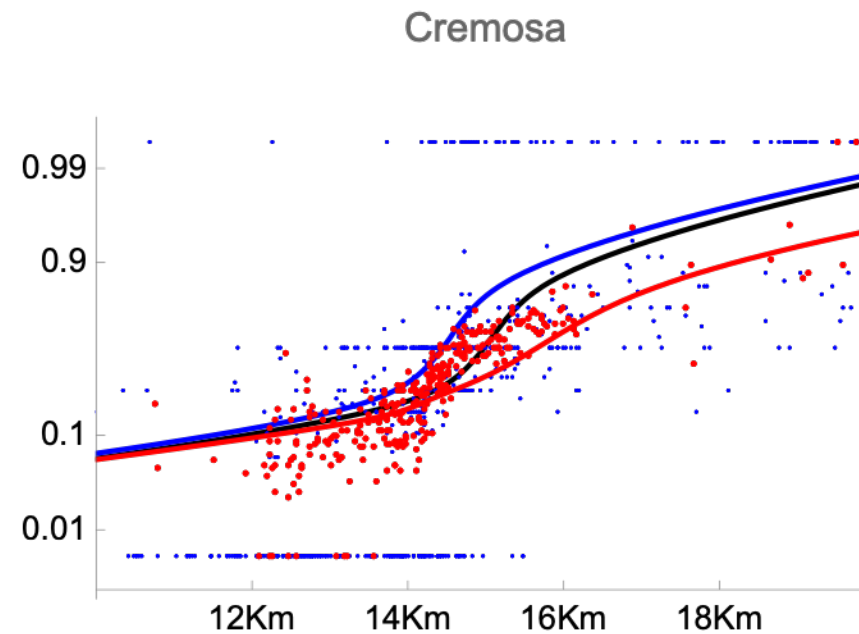
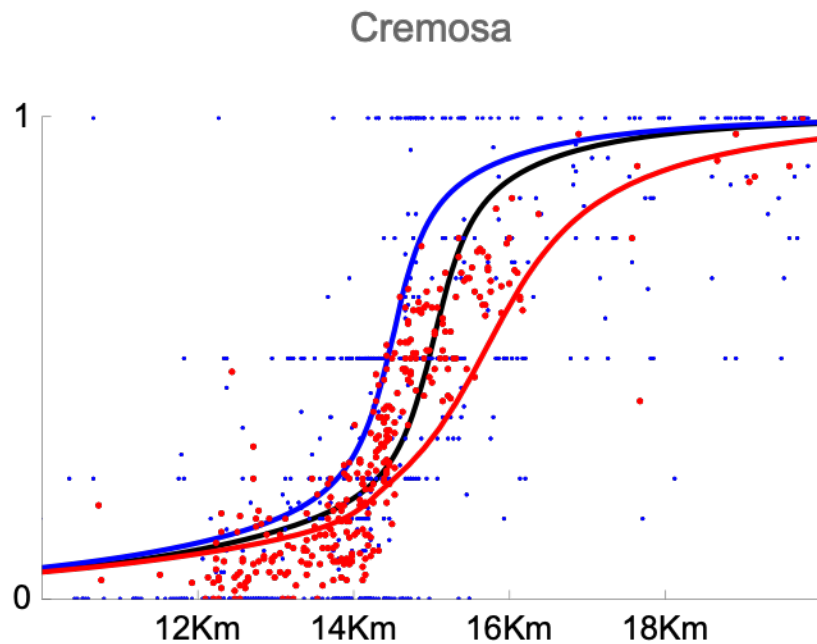
Simulations: five loci

Selection against heterozygotes $1+s_0: 1: 1+s_1$

individual loci (LE; black)

multiple loci, symmetric (LD; blue)

multiple loci, asymmetric (LD; red)



Simulations: five loci

Best fit includes LD, and allows asymmetric selection at two loci:

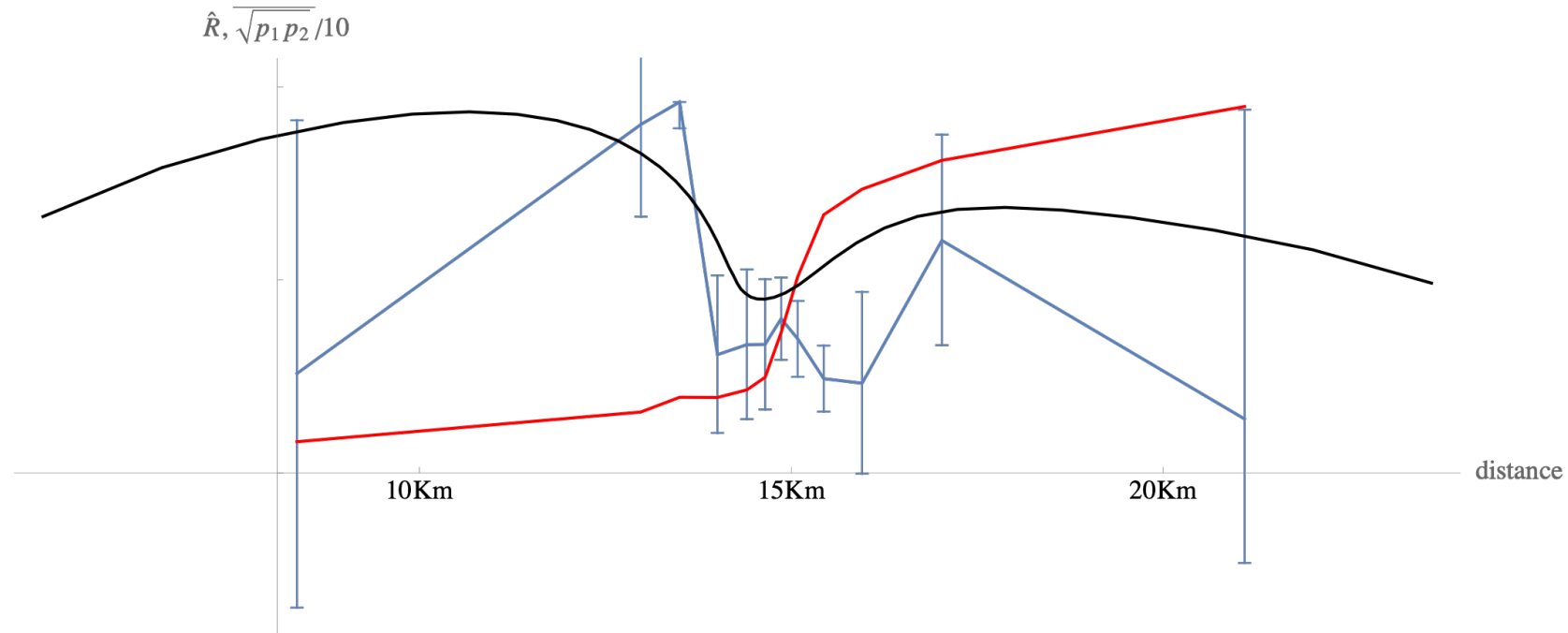
		Ros	Sulf	Flavia	Rubia	Cremosa	log (L)	F _{st}	df
	LE	0.195	0.057	0.115	0.03	0.115	-9491.72	0.070	15
symmetric	LD	0.16	0.091	0.111	0.019	0.073	-10327.8	0.115	7
asymmetric	LD	0.16	0.064	{0.09, 0.07}	0.019	{0.06, 0.04}	-9510.7	0.076	9

Simulations: five loci

Best fit includes LD, and allows asymmetric selection at two loci:

		Ros	Sulf	Flavia	Rubia	Cremona	log (L)	F _{st}	df
	LE	0.195	0.057	0.115	0.03	0.115	-9491.72	0.070	15
symmetric	LD	0.16	0.091	0.111	0.019	0.073	-10327.8	0.115	7
asymmetric	LD	0.16	0.064	{0.09, 0.07}	0.019	{0.06, 0.04}	-9510.7	0.076	9

Observed LD ($R_{ij} = D_{ijj} / \sqrt{p_i q_i p_j q_j}$; blue), simulated LD (black), mean allele frequency (red)



Direct estimates of fitness

Compare the genotype of plants assigned as parents with their neighbours

Fit a model of frequency-dependent selection $p^* = p + pq(s_0 q + s_1 p)$

Out[]//TableForm=

Ros

s	σ	s/σ	P value
-0.284894	0.0670969	-4.24601	1.
0.113408	0.050365	2.25173	0.0104

Sulf A

s	σ	s/σ	P value
-0.017014	0.115532	-0.147267	0.561
0.0773246	0.103306	0.748502	0.2278

Flavia A

s	σ	s/σ	P value
-0.120361	0.0665252	-1.80925	0.9598
0.144175	0.0932438	1.54622	0.064

Rubia

s	σ	s/σ	P value
-0.311008	0.215916	-1.44042	0.923
0.192839	0.124381	1.55039	0.0608

Cremona

s	σ	s/σ	P value
-0.125669	0.0666614	-1.88518	0.974
0.178139	0.100384	1.77457	0.0388

Summary and open questions

- Genome scans: excess F_{st} and sharp clines indicate flower colour loci
- Associations (LD) between unlinked loci indicate a genetic barrier
- Stepped clines are mainly due to long-tailed dispersal, rather than linked selection
- Dominance, epistasis, frequency dependent selection are hard to distinguish
- SNP markers are most likely not causal: introgression also contributes to stepped clines

How does selection act ?

How much variation in flower pattern is captured by these population genetic methods?

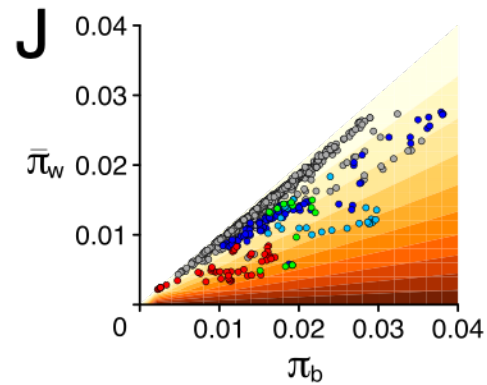
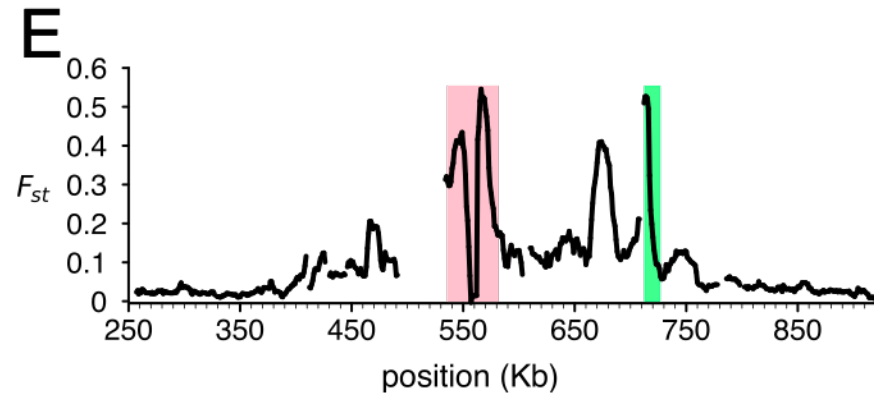
Is there polygenic adaptation for other traits?



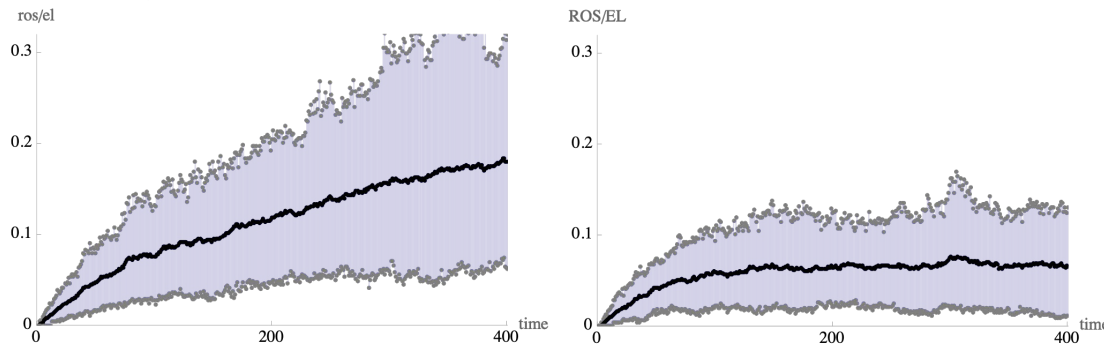
Thank you for your attention!

Understanding the *Rosea/Eluta* region

F_{st} shows a broad increase + sharp peaks



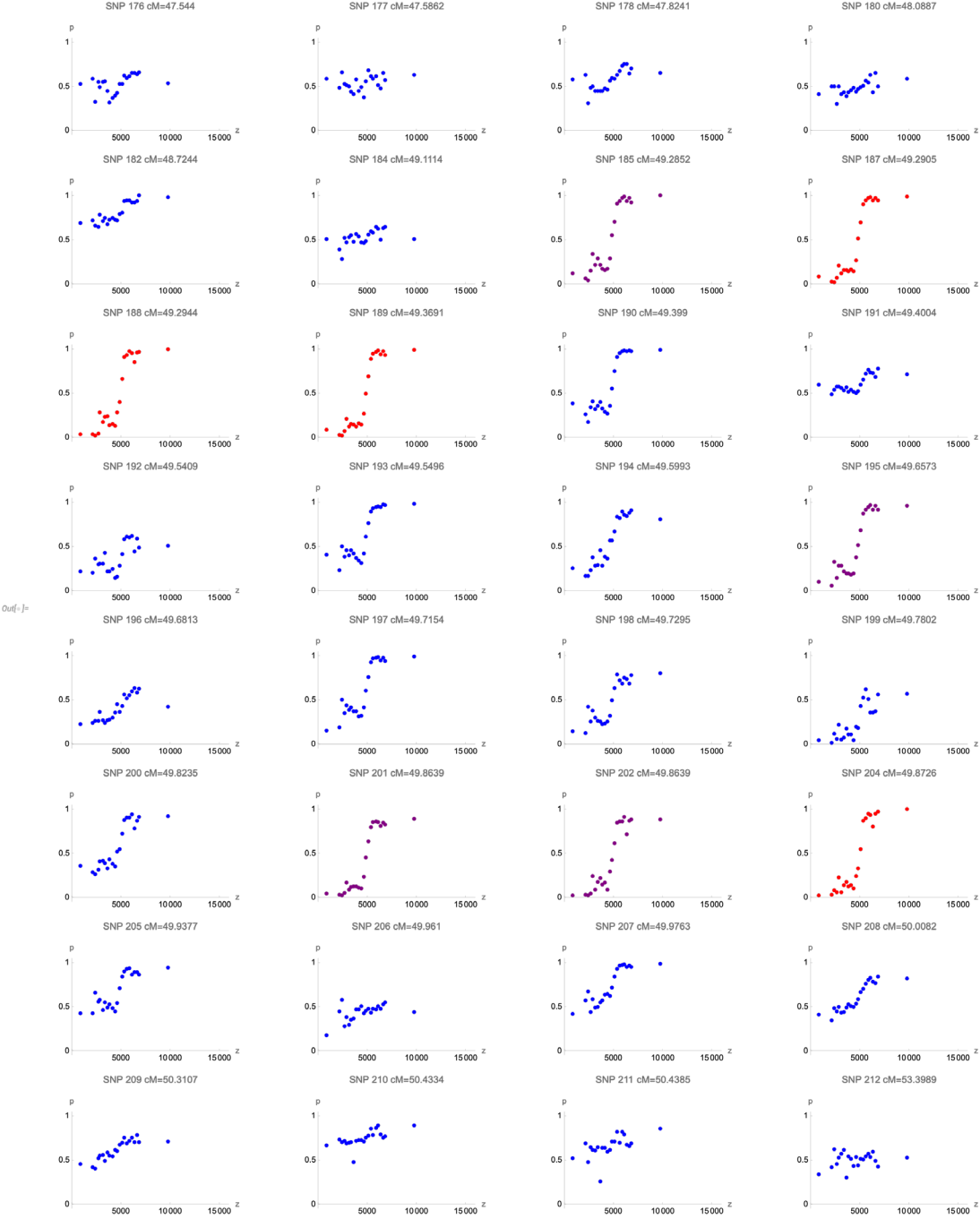
Recombinants (ros/el, ROS/EL) reach $\sim 10\%$ in the centre, indicating > 80 generations



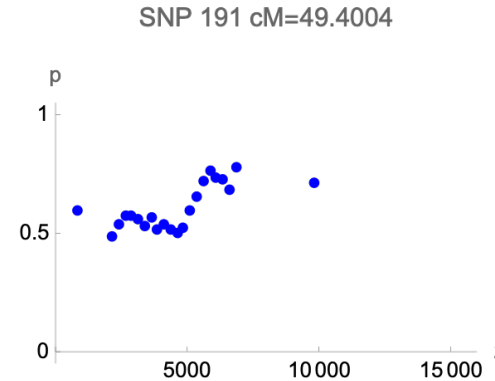
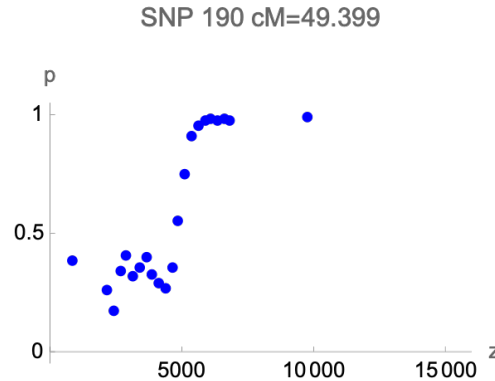
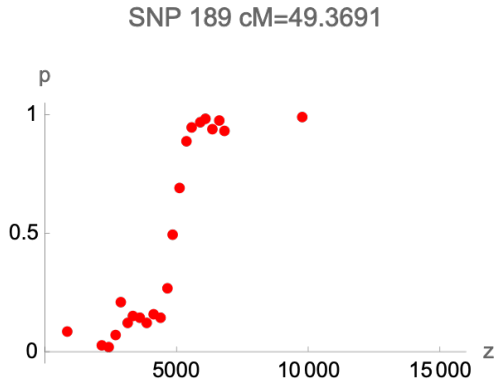
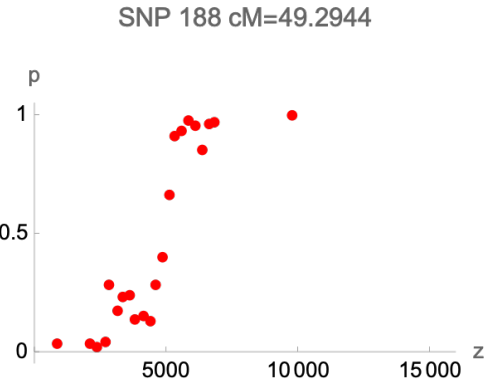
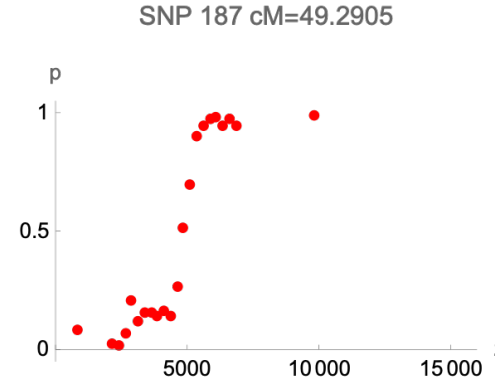
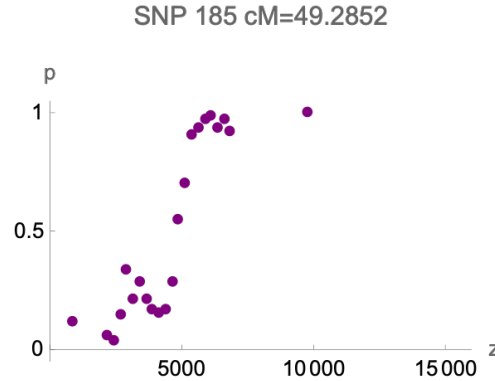
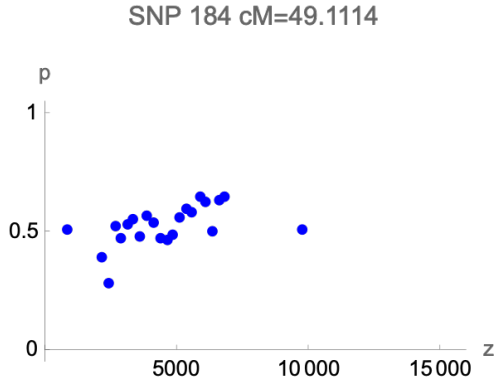
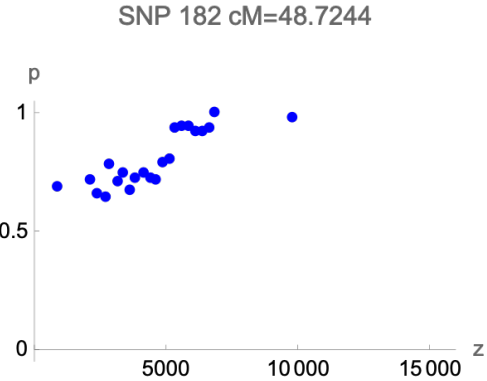
Simulations with 10% +ve fds

Clines along the genome

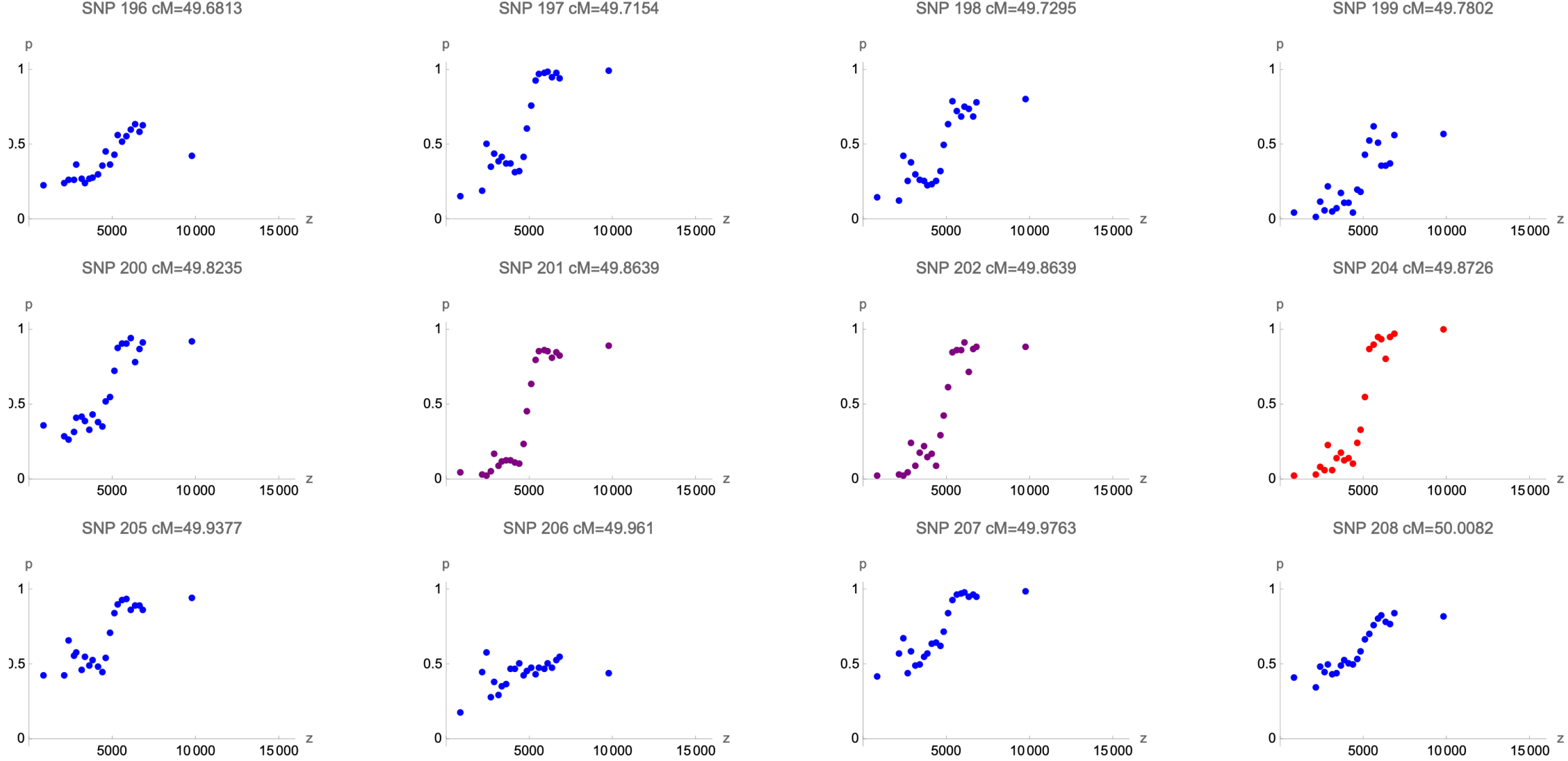
Clines along the genome: around *Rosea/Eluta*



Clines along the genome: around *Rosea*

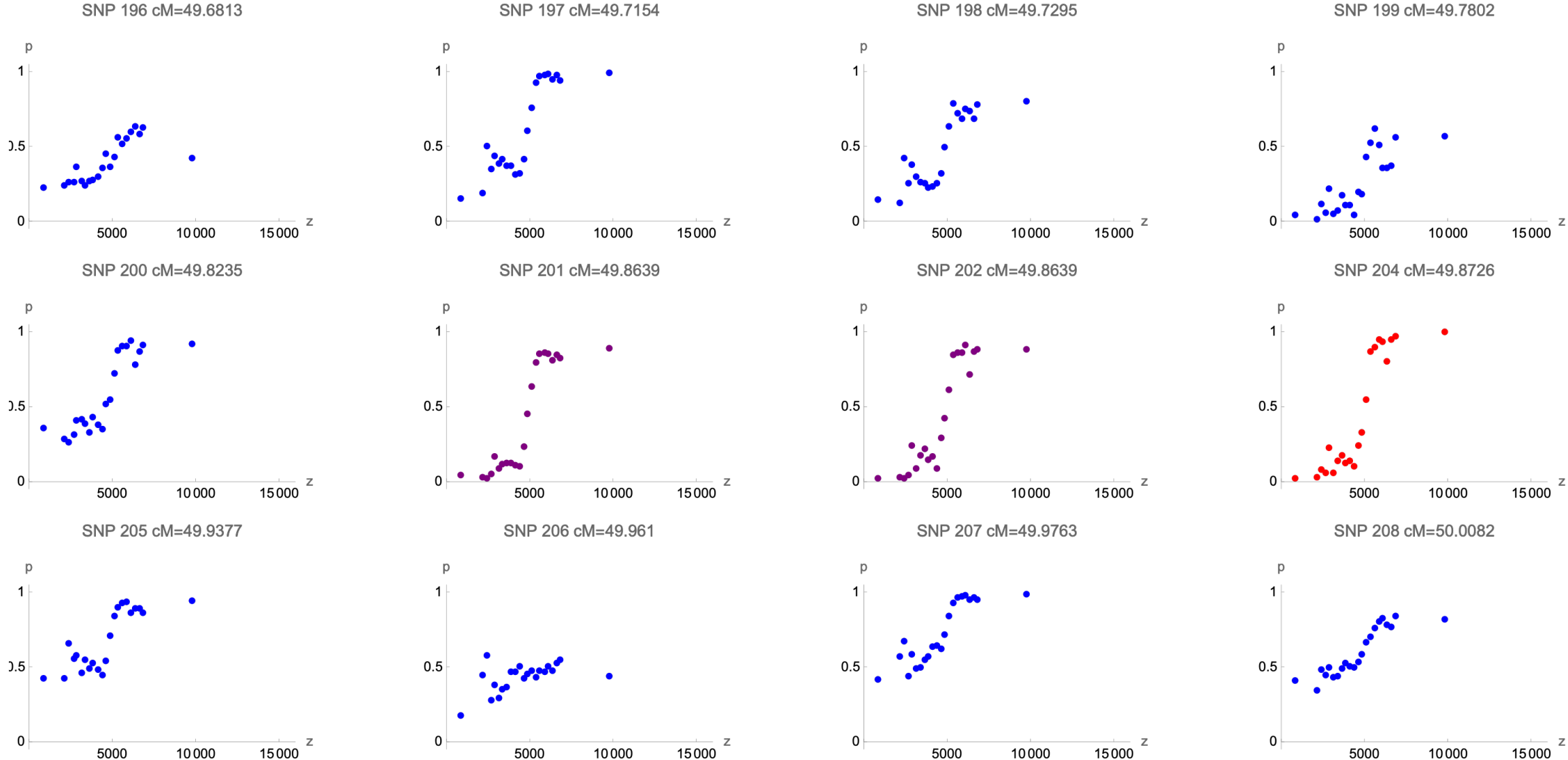


Clines along the genome: around *Eluta*



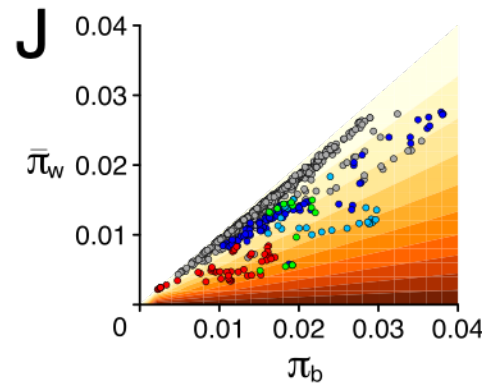
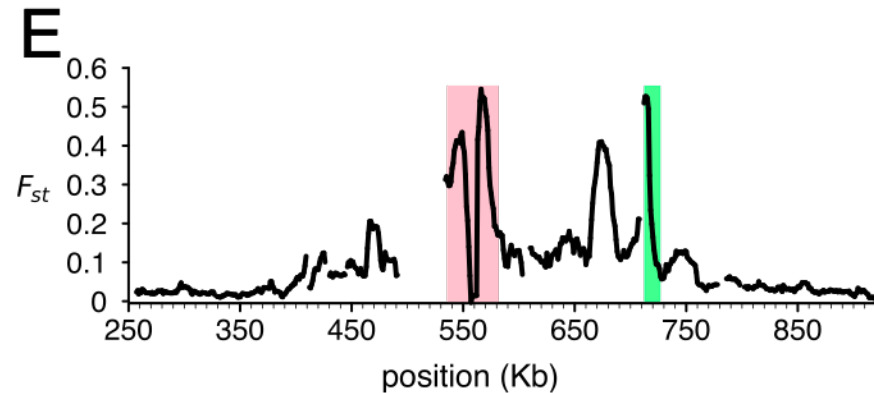
Clines along the genome: around *Eluta*

How are clines reflected in the underlying haplotypes ?

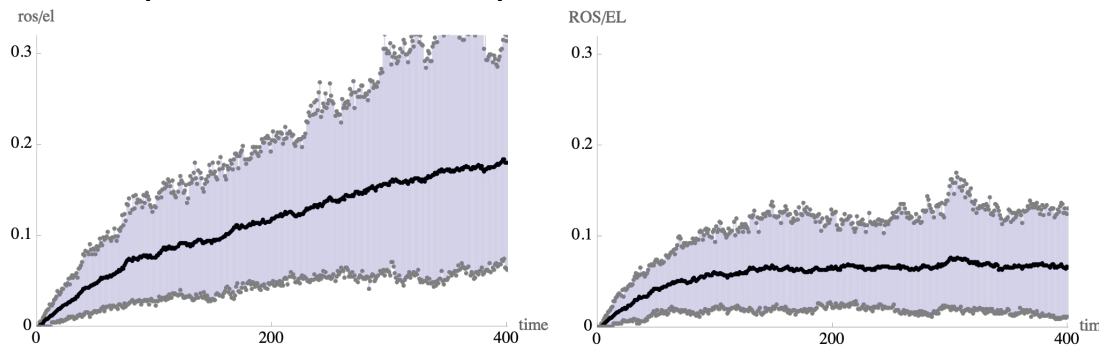


Understanding the *Rosea/Eluta* region

F_{st} shows a broad increase + sharp peaks



Recombinants (ros/el, ROS/EL) reach ~10% in the centre, indicating > 80 generations

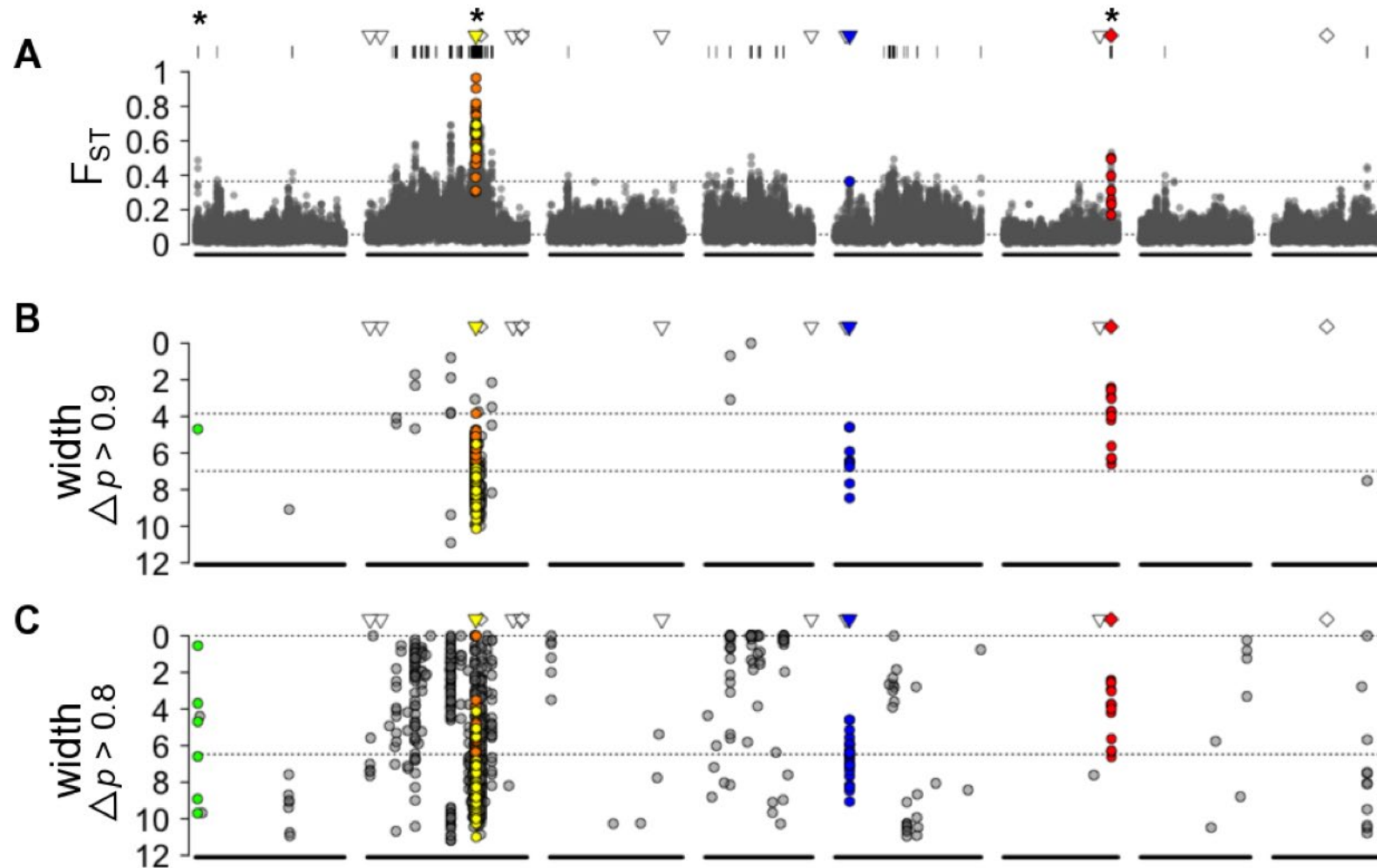


Simulations with 10% +ve fds

Clines along the genome

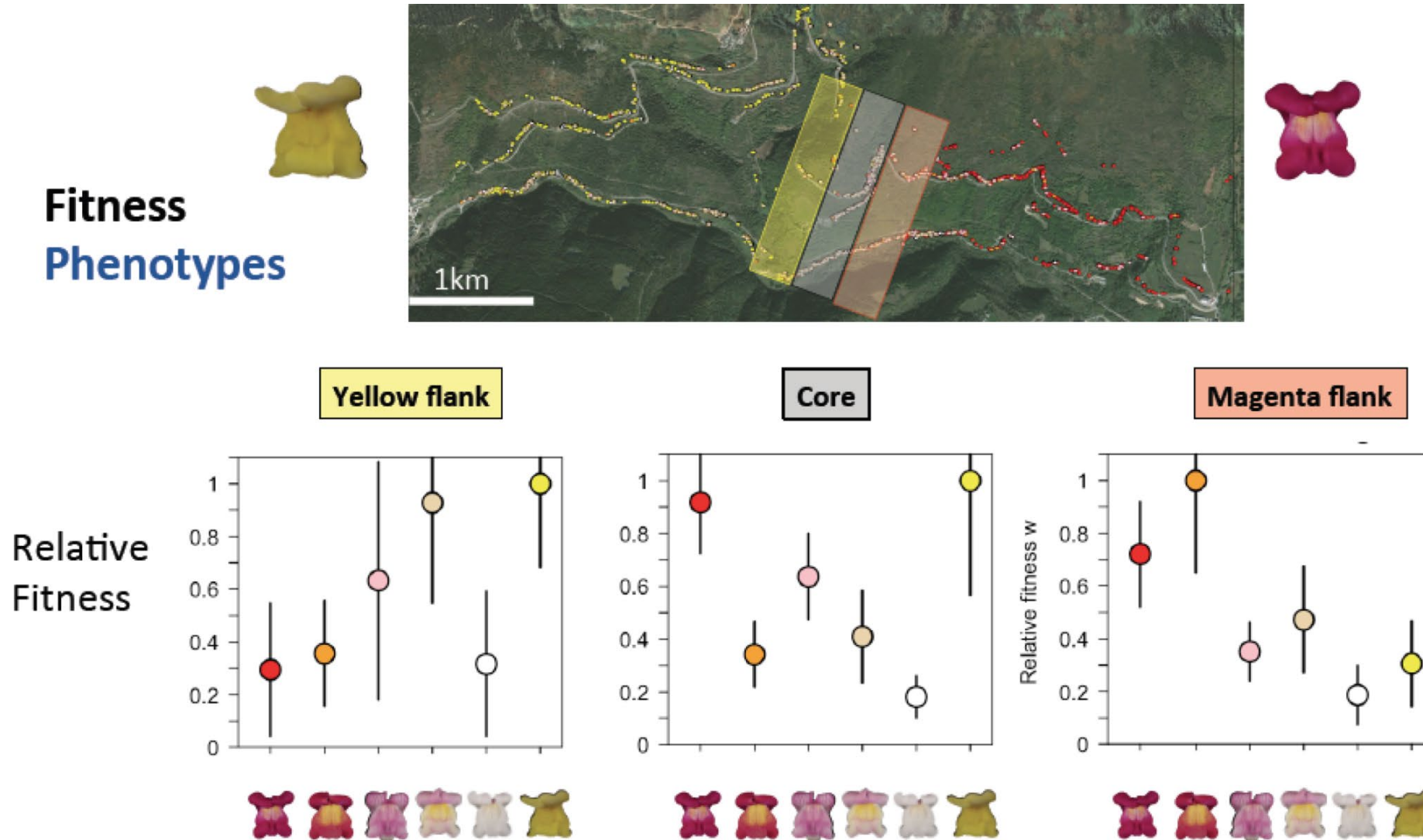
What haplotype structure do we expect, after multiple sweeps through diverse populations?
How much more can we infer from haplotypes ???

- Narrow clines inferred from 8 pooled samples correspond to known colour genes



The pedigree

Strong selection against hybrids, and for common phenotypes

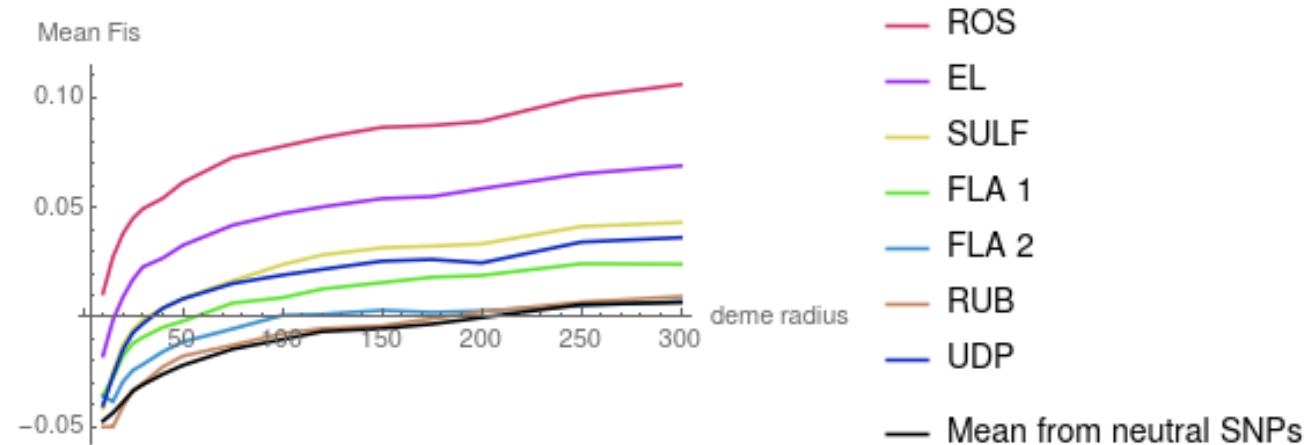


Analyzing F_{is}

$F_{is} = 1 - \text{observed} / \text{expected \# of heterozygotes}$

- F_{is} increases with “deme” size due to Wahlund effect
- Heterozygote deficit at ROS is higher.
- F_{is} of the flower color SNPs is higher than the neutral ones – higher fluctuation of allele frequency or deviation from random mating?

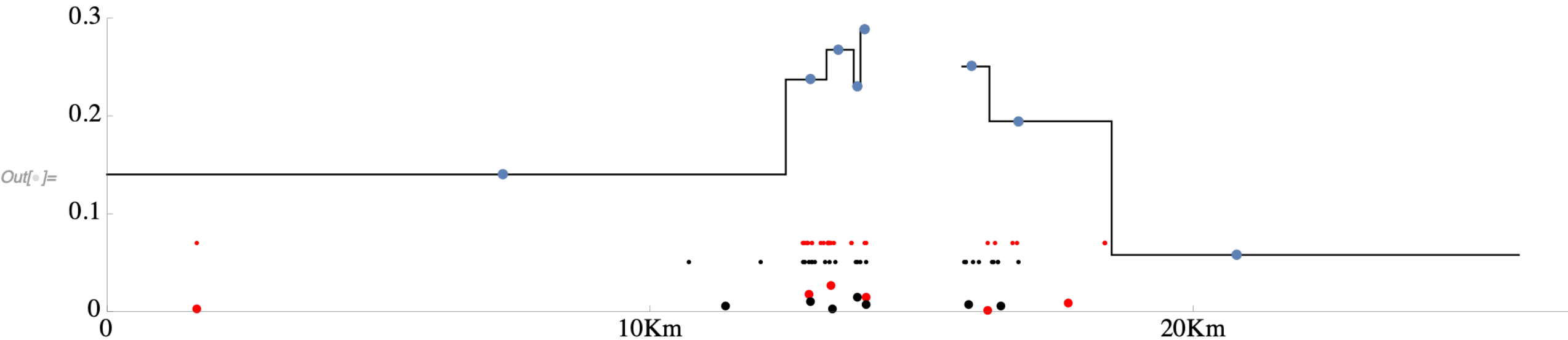
Heterozygosity of offspring generated by the leptokurtic dispersal kernel is consistent with the data.



Data	Mean H	Std
Field data	0.35	0.19
Simulated offspring	0.34	0.20

Associations between loci: Locations of unlikely individuals

Small dots: individuals with $P < 10^{-4}$ Red: those with ≥ 3 foreign homozygotes (seed dispersal?); black: < 3 foreign homozygotes (F1 or backcross?). Large dots show means for each bin. Blue dots (above): average position and introgressing allele freqs.



Associations between loci: Finding unlikely individuals

There are 51 individuals with $P < 10^{-4}$ on the yellow side, 20 on the magenta slide
 \log_{10} probabilities of various classes

genotype	Z	P_0	$P_{B \times 1}$	P_{F1}	P_{foreign}
{2, 1, 2, 0, 2}	12 830.8	-7.05444	-5.1952	-4.78533	-2.99915
{2, 1, 2, 0, 2}	12 904.6	-7.05444	-5.1952	-4.78533	-2.99915
{2, 2, 2, 2, 2}	13 279.3	-6.93074	-4.81918	-3.87577	-0.8208
{2, 2, 2, 2, 2}	13 286.5	-6.93074	-4.81918	-3.87577	-0.8208
{2, 2, 2, 2, 2}	13 287.	-6.93074	-4.81918	-3.87577	-0.8208
{2, 2, 2, 2, 2}	13 289.9	-6.93074	-4.81918	-3.87577	-0.8208
{2, 2, 2, 2, 2}	13 291.2	-6.93074	-4.81918	-3.87577	-0.8208
{2, 2, 2, 2, 2}	13 295.6	-6.93074	-4.81918	-3.87577	-0.8208
{2, 2, 2, 1, 2}	13 284.8	-6.82148	-4.82428	-3.9826	-1.37701
{2, 2, 2, 1, 2}	13 291.	-6.82148	-4.82428	-3.9826	-1.37701
{2, 2, 2, 1, 2}	13 299.	-6.82148	-4.82428	-3.9826	-1.37701
{2, *, 2, 2, 2}	13 294.5	-6.51503	-4.62532	-3.77126	-0.969783
{2, 1, 2, 2, 2}	13 326.2	-6.40978	-4.46008	-3.61305	-1.28467
{2, 2, 1, 2, 0}	1646.62	-6.26334	-3.96534	-3.417	-1.83962

Associations between loci: Locations of unlikely individuals

$\log(L)$ for % parental and F1 individuals on the yellow and magenta flanks (left, right).

MLE for parental on *striatum* flank: 0.0327 {0.0236,0.0427}

MLE for parental on *pseudomajus* flank: 0.0124 {0.0080,0.0176}

