

Genetics of sex determination in *Lysiphlebus fabarum*

Cyril Matthey-Doret

Supervised by: Casper Van Der Kooi

Directed by: Tanja Schwander



UNIL | Université de Lausanne



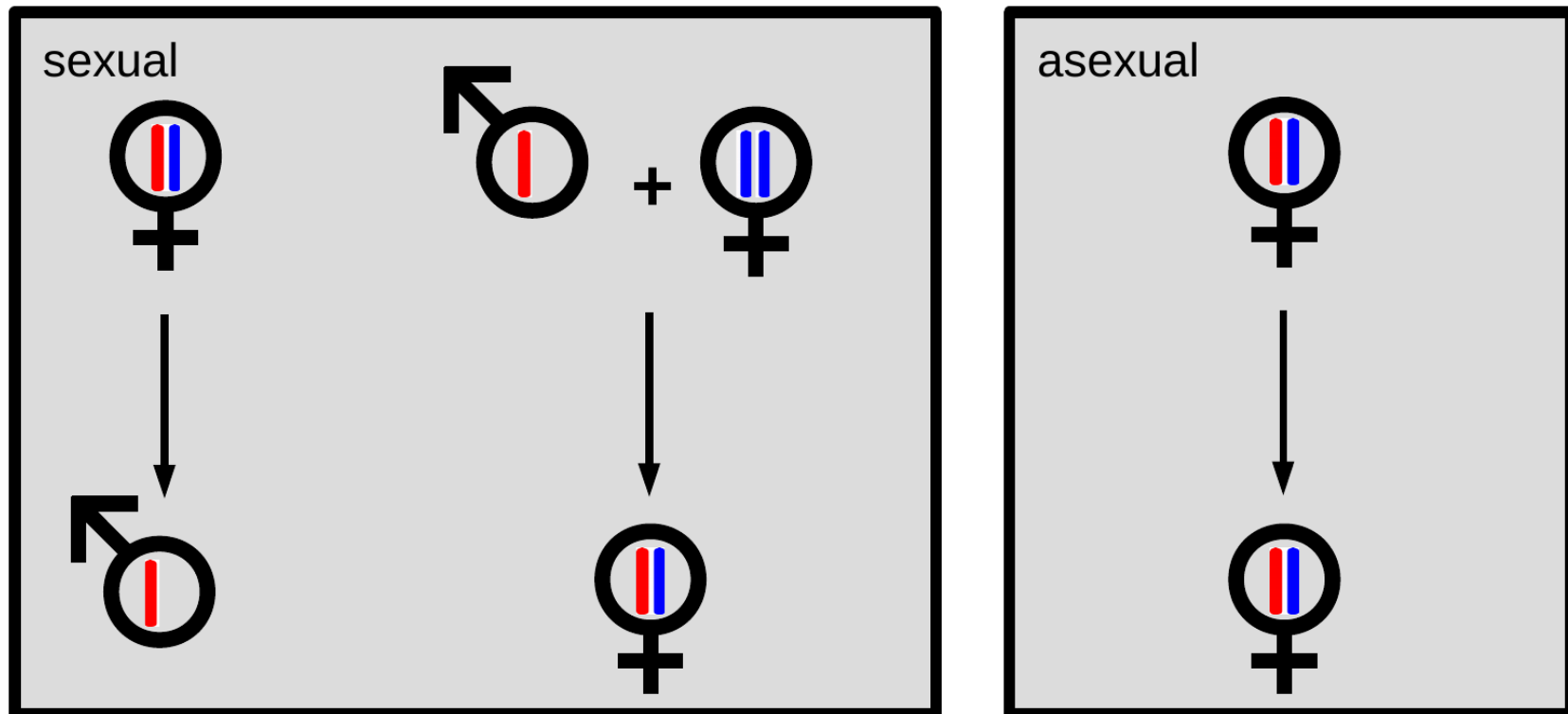
Lysiphlebus fabarum © INRA, Bernard Chaubet



Lysiphlebus fabarum © INRA, Bernard Chaubet

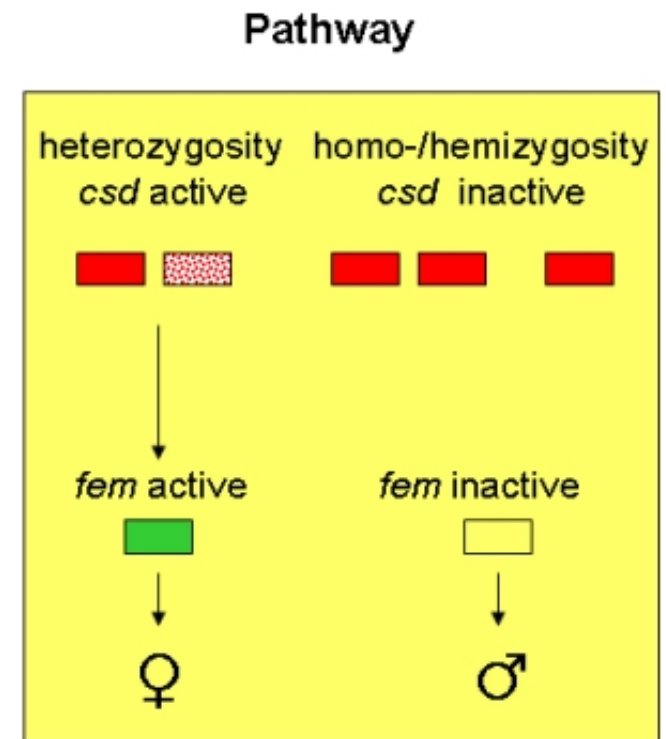
Sex determination

- Hymenoptera are haplodiploids
- In sexual lineages: egg fertilization --> female



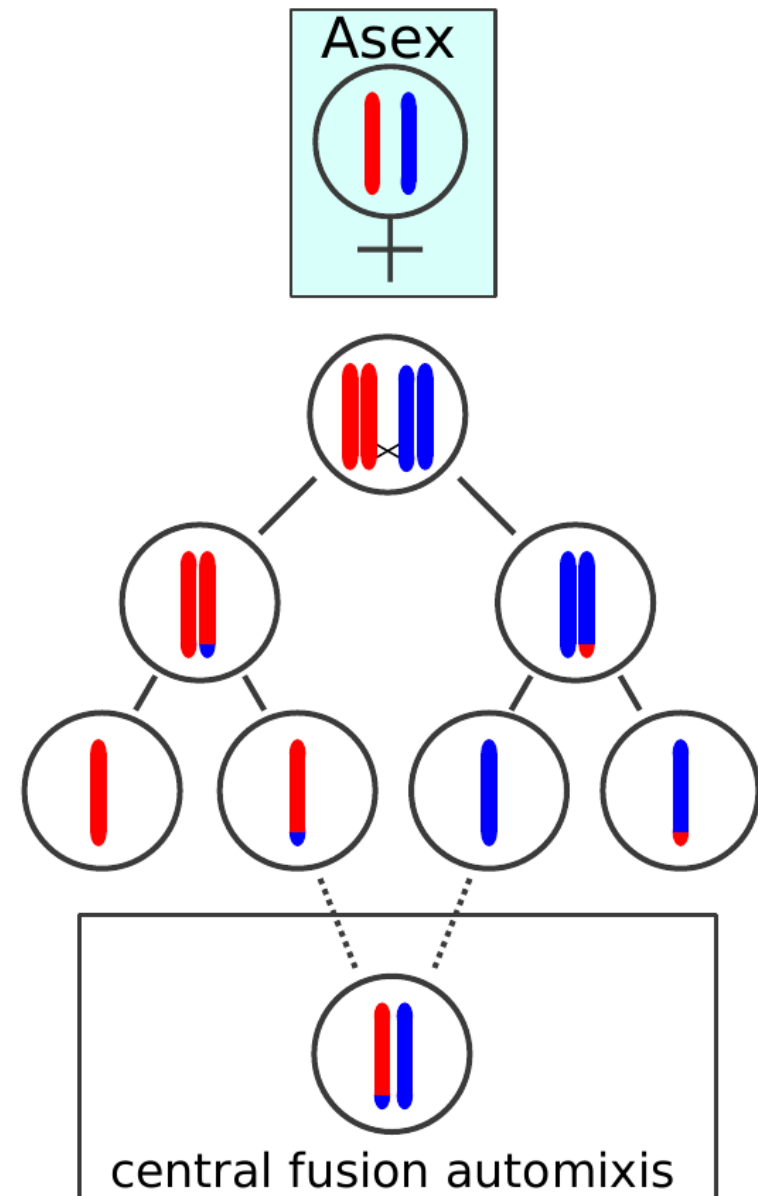
CSD mechanism

- 2 different alleles required to trigger female development
- Males are normally haploid: single copy
- Diploid individuals with homozygous CSD --> males
- Can be single or multi-locus CSD



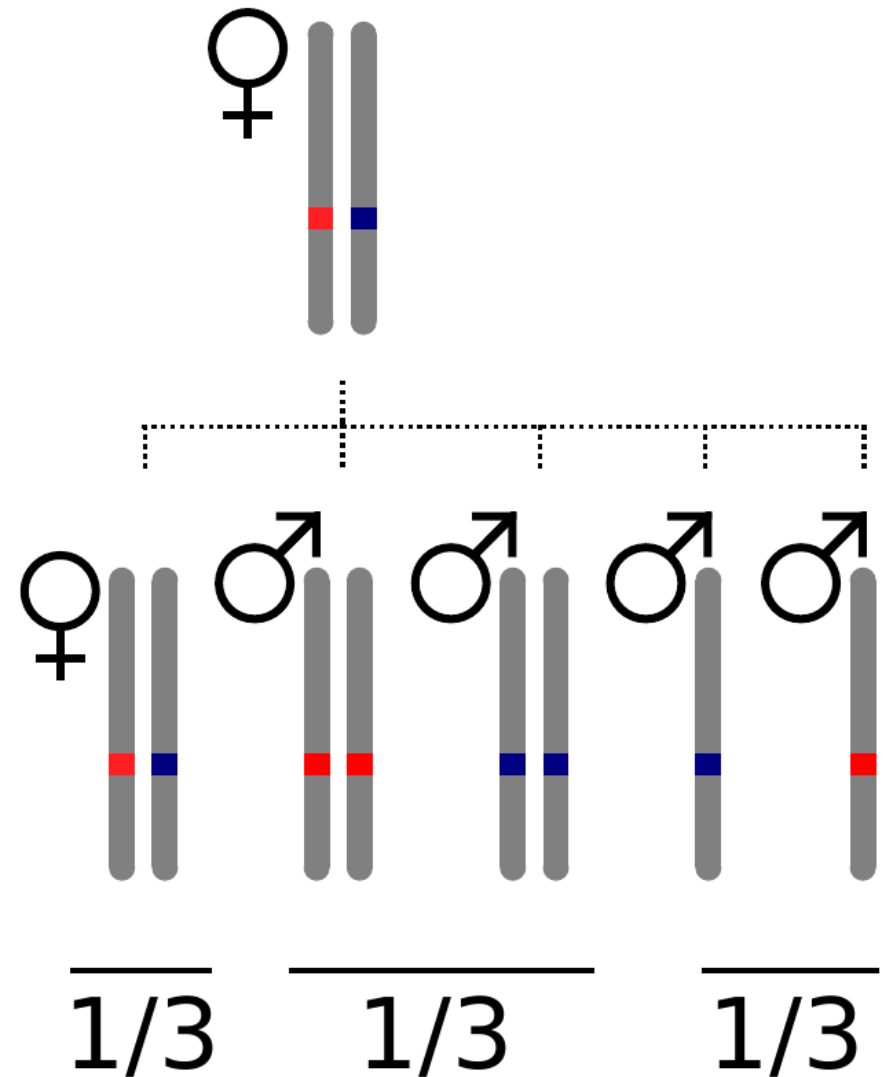
L.fabarum : Introduction

- Sex and asex populations
- Central fusion automixis: recombination
- Multi-locus or single-locus *csd* ?



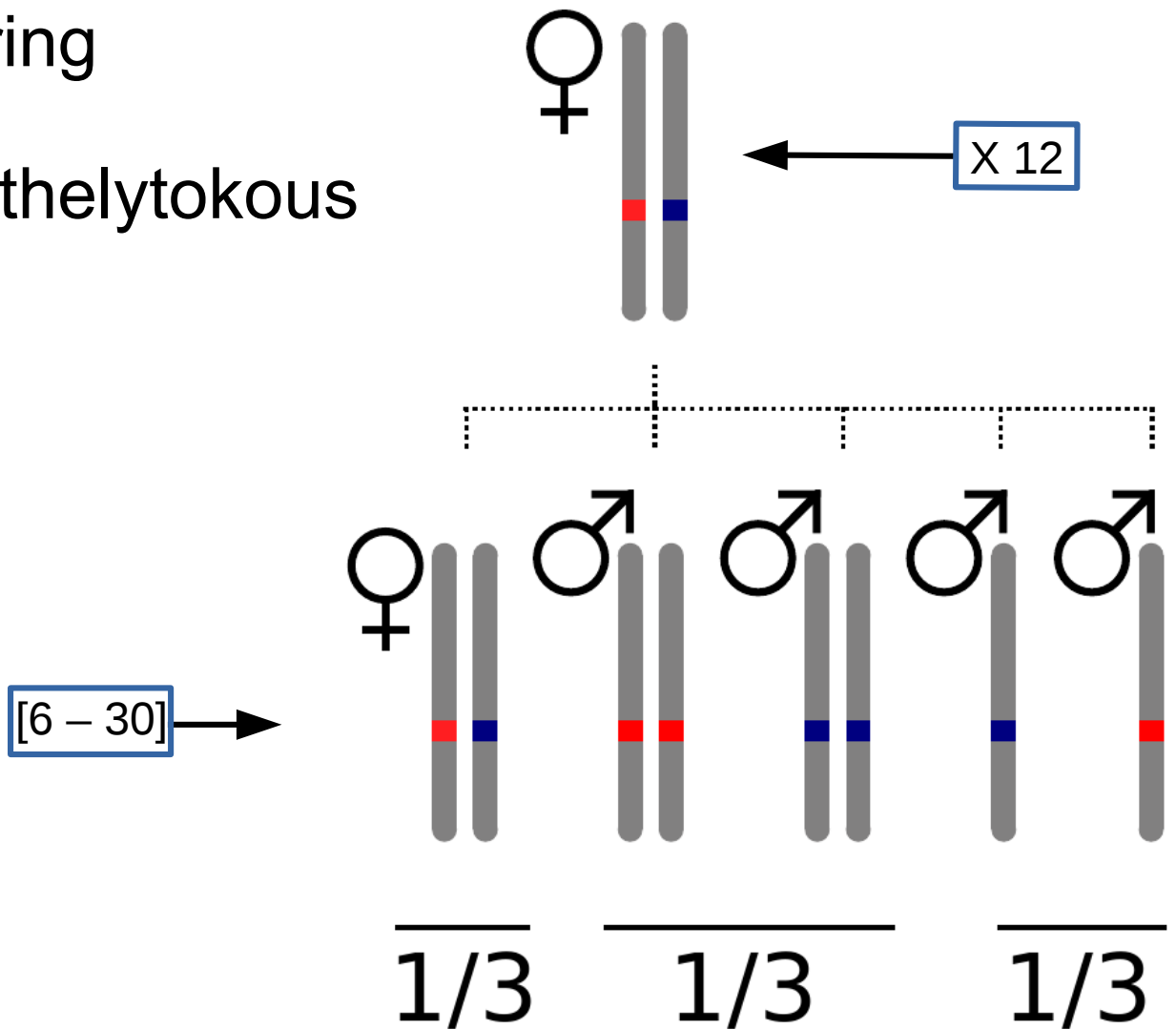
Crossed individuals

- Thelytokous females
- Strongly inbred
 - High rate of diploid males production
 - Highly homozygous background



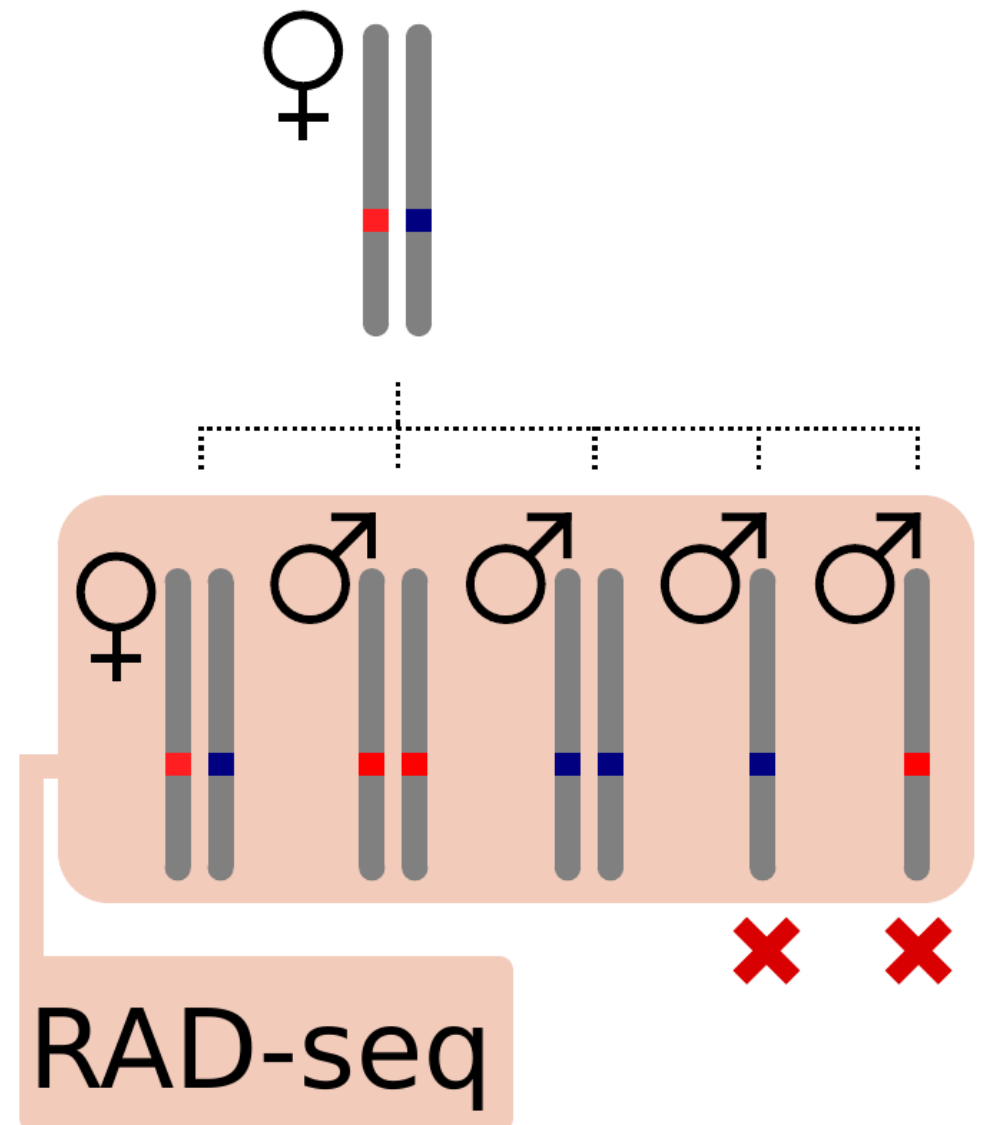
Crossed individuals

- Total of 201 offspring
- From 12 different thelytokous mothers



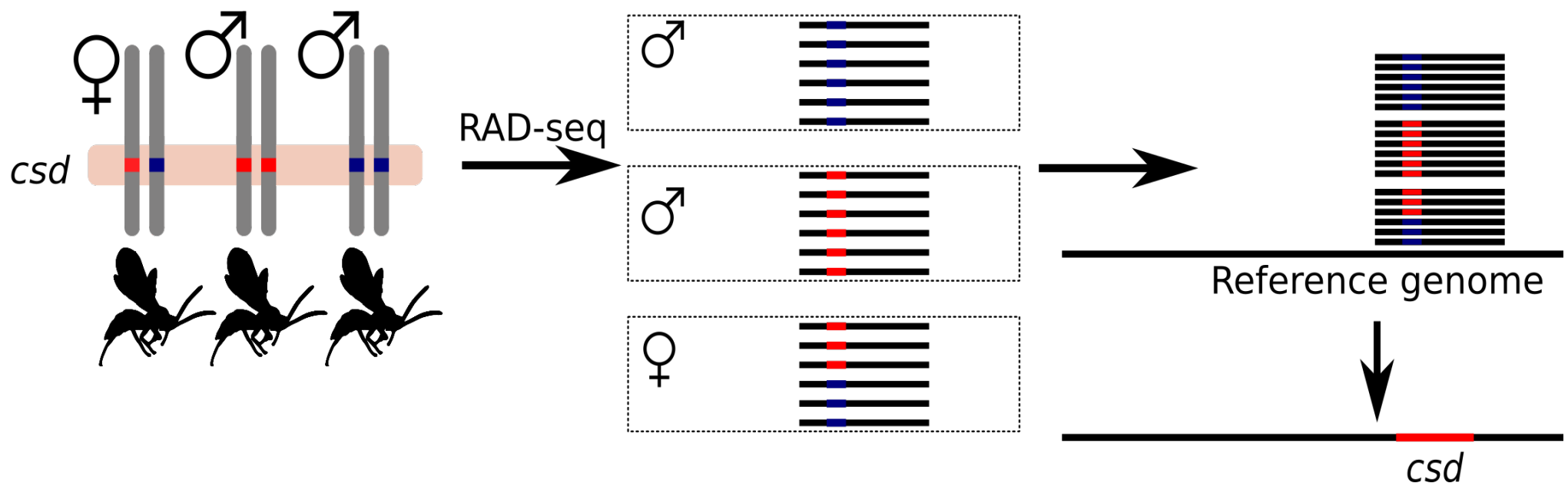
Crossed individuals

- Different allelic combinations of CSD
- Haploid males: useless
- Sequencing data:
 - 201 offspring
 - 11 mothers (1 missing)



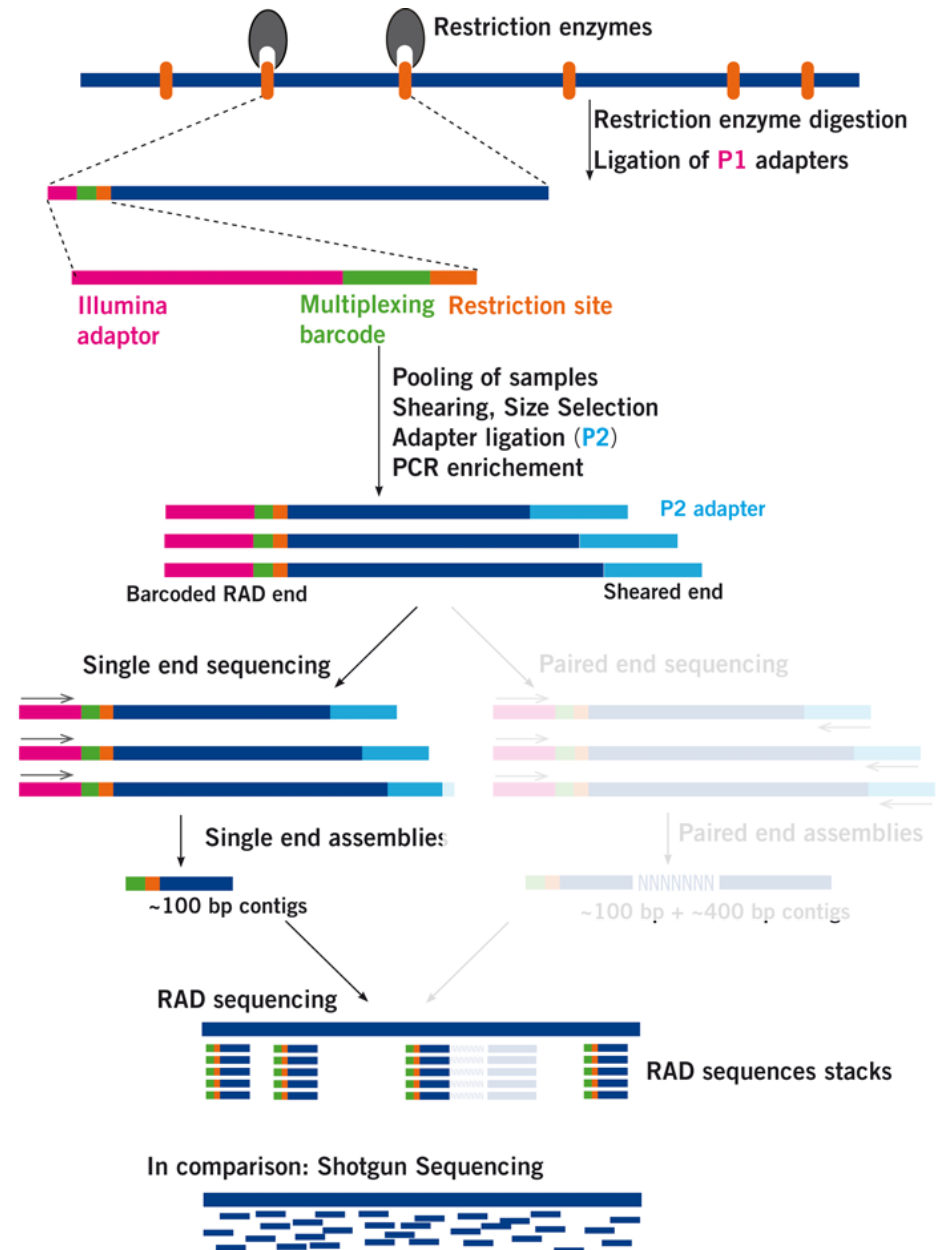
Finding CSD: General idea

- Find regions that are homozygous in males and heterozygous in females



RAD-seq

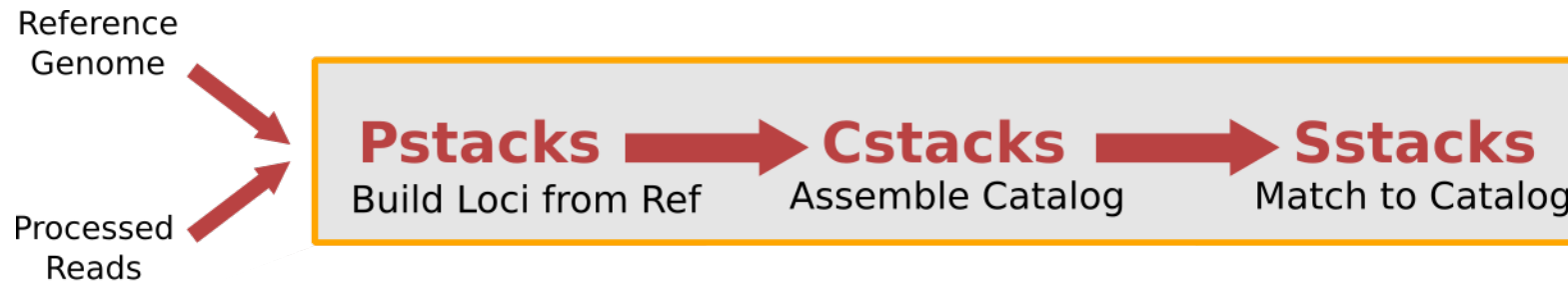
- Digest genome
- Sequence flanking regions
- Single-end sequencing



Bioinformatic analysis of RAD-seq

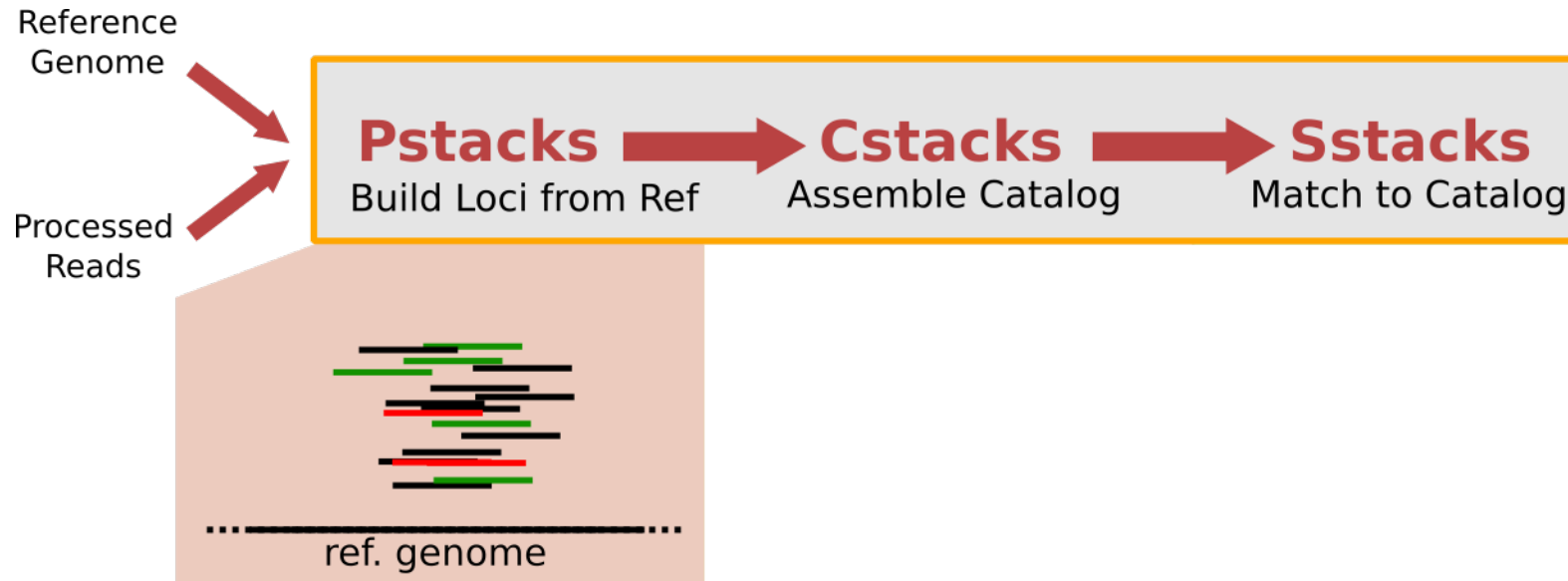
- Processing raw reads into a catalog of loci
- Different pipelines available
 - **STACKS**, PyRAD, Rtd, RAD-tools...
- Requires setting several parameters through empirical testing

STACKS Pipeline

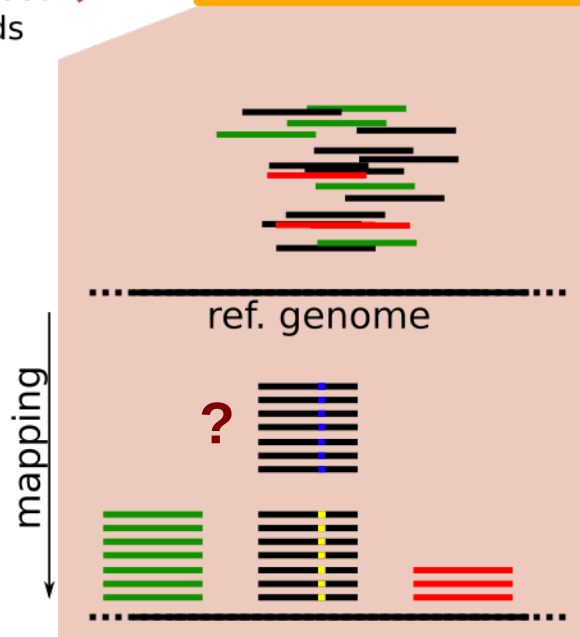
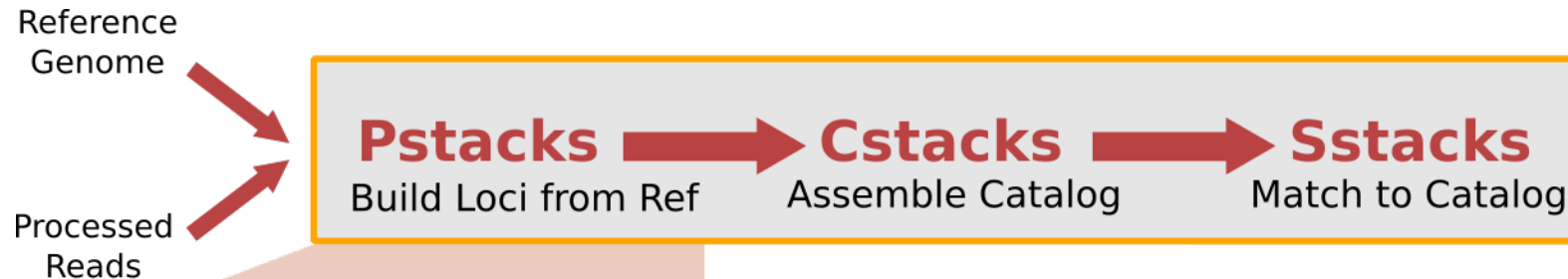


- Builds loci from reference, or *de novo*
- Each module takes arbitrary parameters

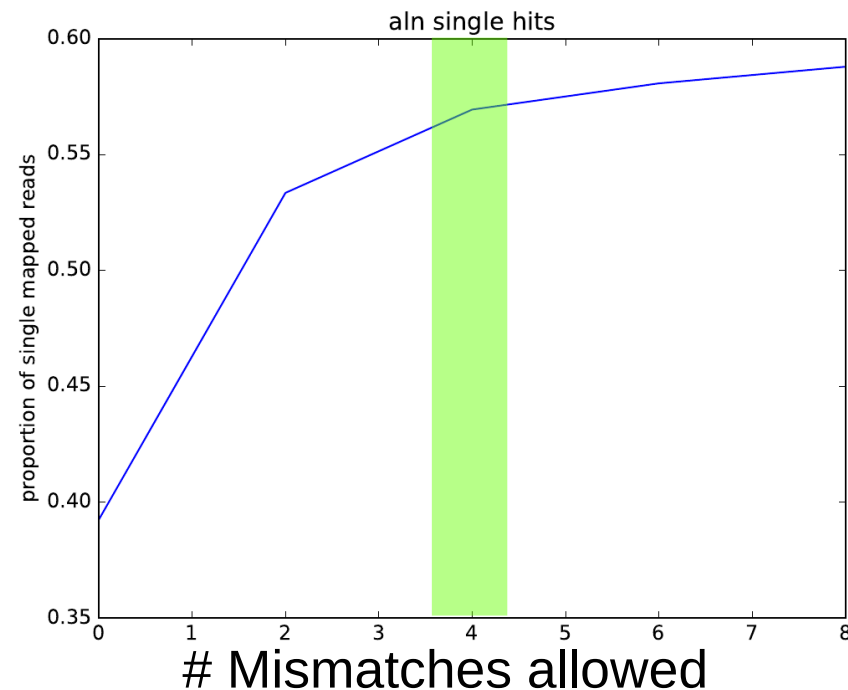
STACKS Pipeline



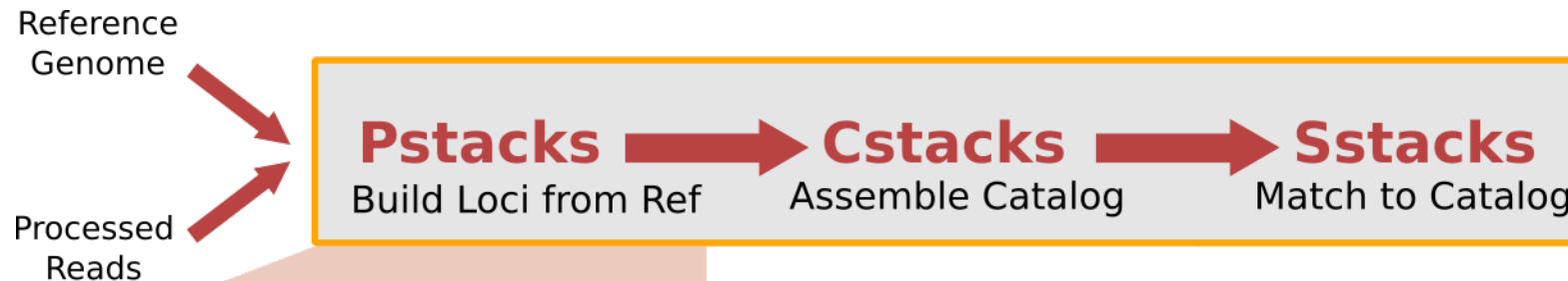
STACKS Pipeline



- How many mismatches allowed for 2 reads to map at the same position ?



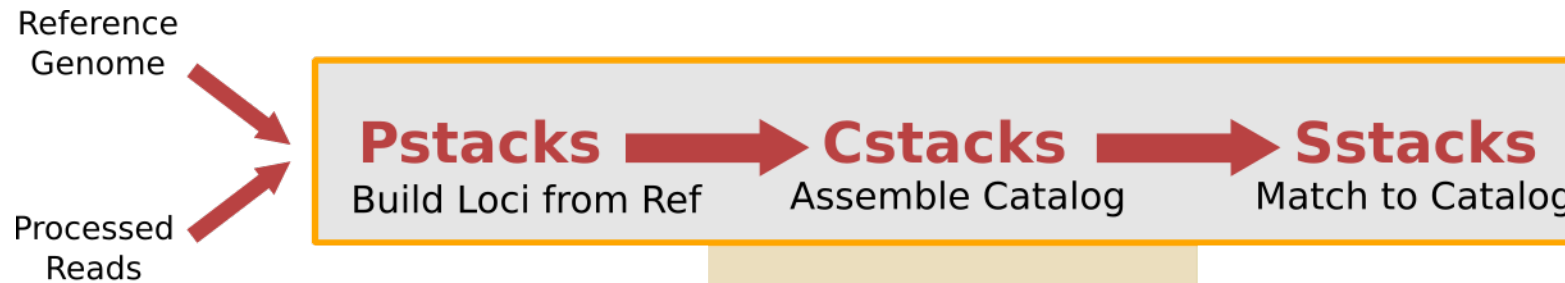
STACKS Pipeline



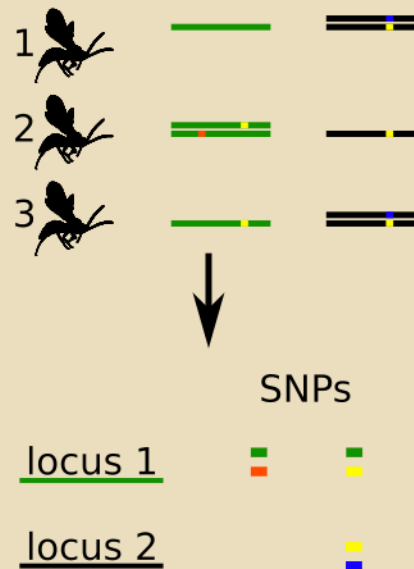
	min cov	nloci	mean cov	sd cov
1		5020.07	31.8333	58.6313
2		3659.14	42.5051	64.9899
3		2747.17	49.8687	68.1818
4		1505.35	55.5657	70.0606
5		781.869	60.101	71.3283
6		597.859	63.9141	72.2677

Minimum stack depth

STACKS Pipeline



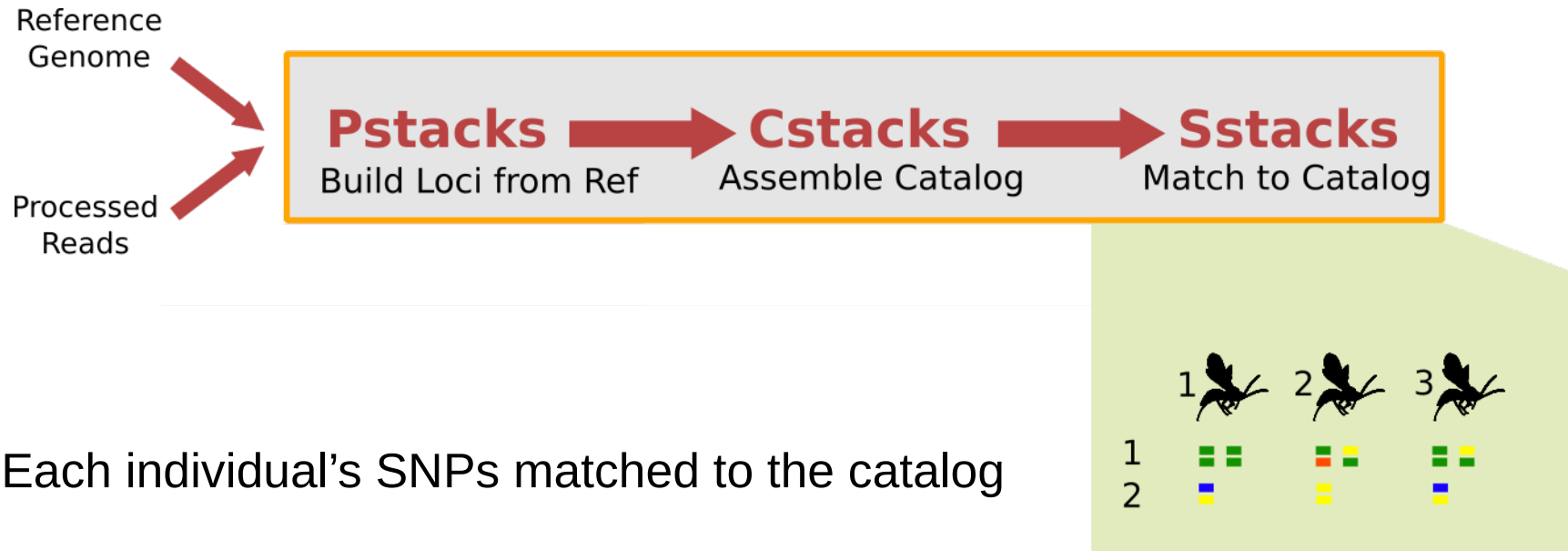
- Building catalog from all individuals
- Position and coverage of every SNP variant



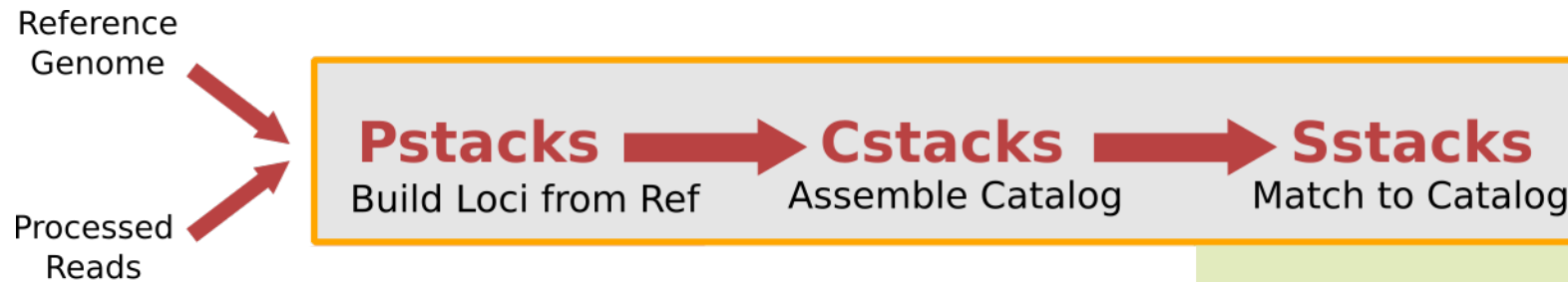
Mismatches: homologous loci

mismatch	mean loci	mean alleles
1	2696	7799
2	3129	8464
3	3198	8704
4	3263	8914

STACKS Pipeline



STACKS Pipeline



- Each individual's SNPs matched to the catalog

Methods in Ecology and Evolution

Methods in Ecology and Evolution 2017

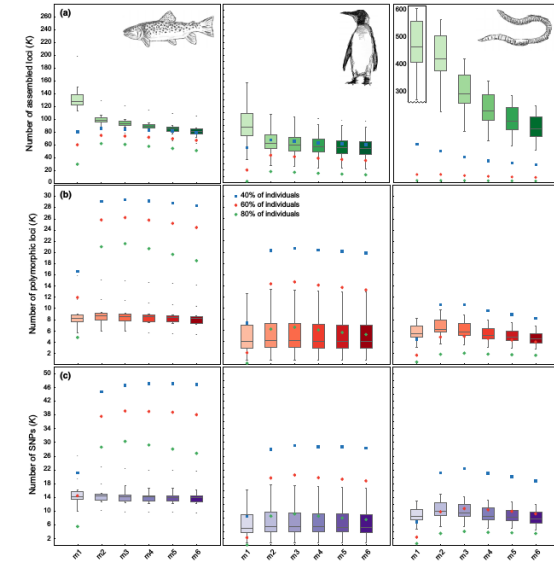
doi: 10.1111/2041-210X.12775



Lost in parameter space: a road map for STACKS

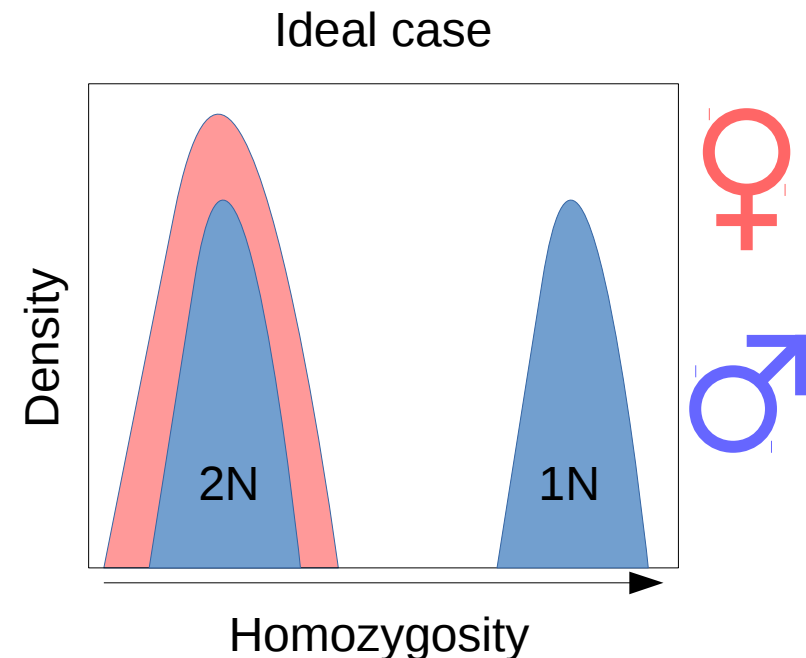
Josephine R. Paris¹ , Jamie R. Stevens¹ and Julian M. Catchen^{*,2}

¹Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK; and ²Department of Animal Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA



Excluding haploids

- Loci obtained in haploids should all be homozygous.
- But this is not the case! Why ?
 - Sequencing errors
 - Parologue merging
 - Repetitive sequences ?

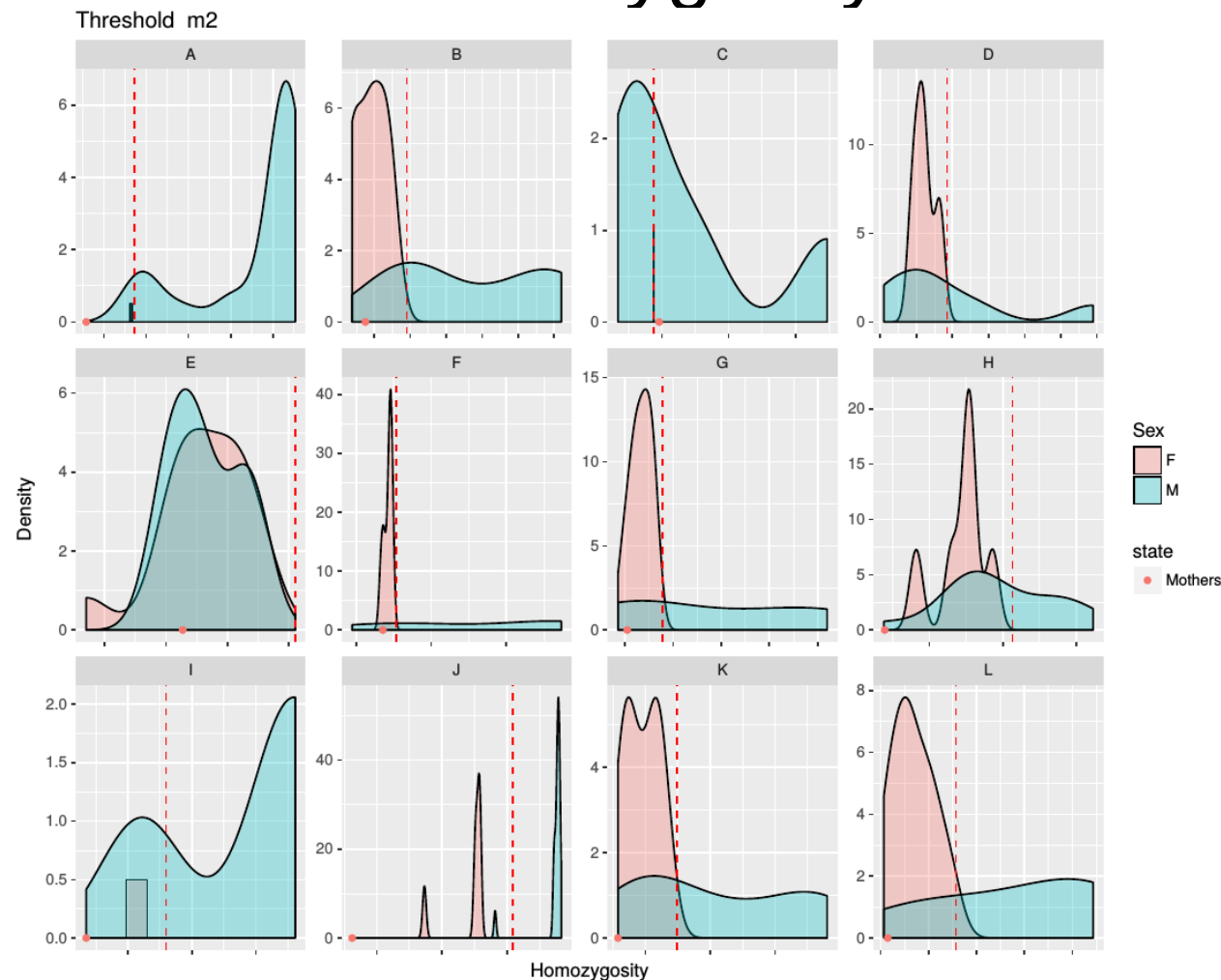


Excluding haploids

- How to split haploids and diploids ?
- Current solution: Choose a homozygosity threshold:

$$Hom_M > \mu_F + 2\sigma_F$$

Mu: mean of daughters hom.
Sigma: stdev of daughters hom.

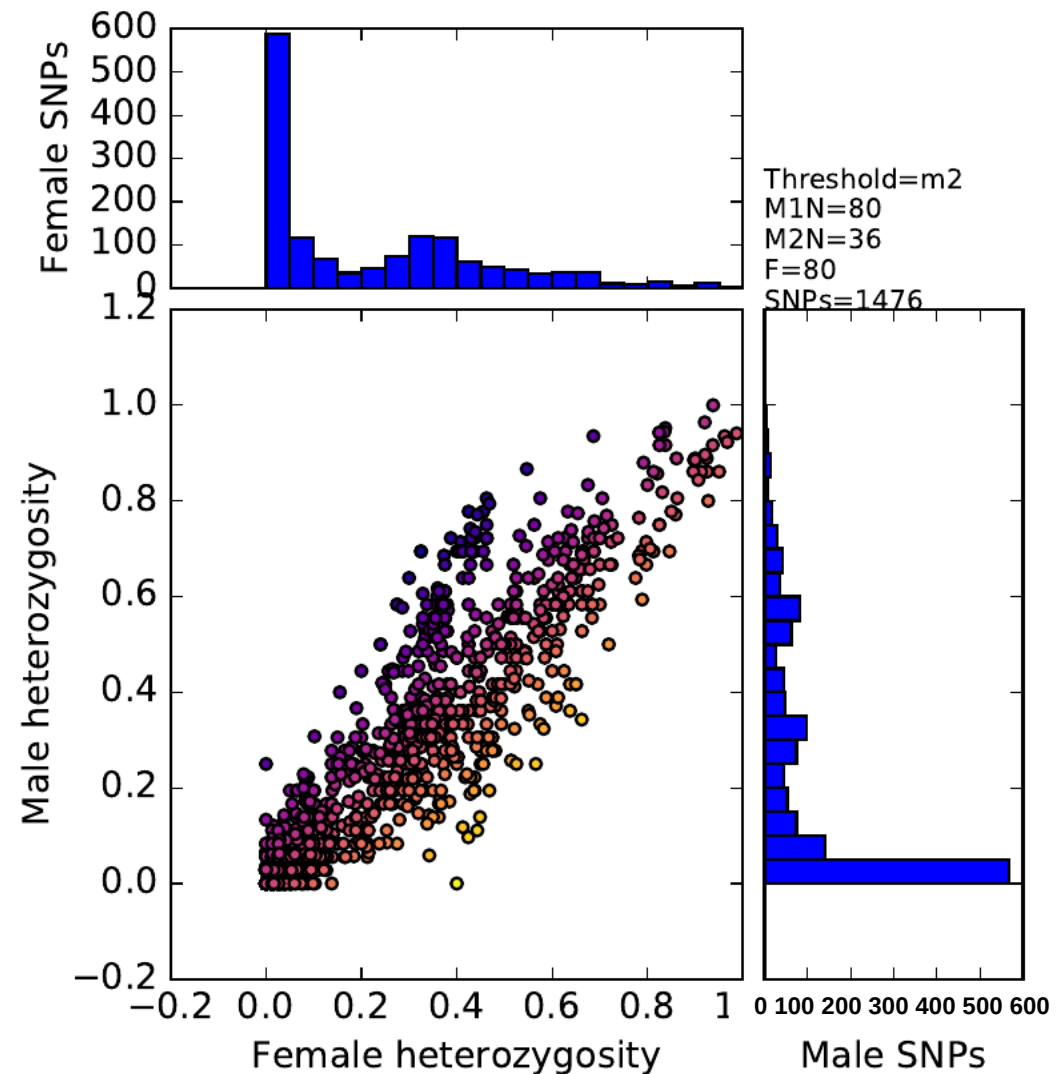


Improving ploidy separation

- Remove alleles with low number of reads relative to other allele
- Remove loci with excessive coverage (paralogues)

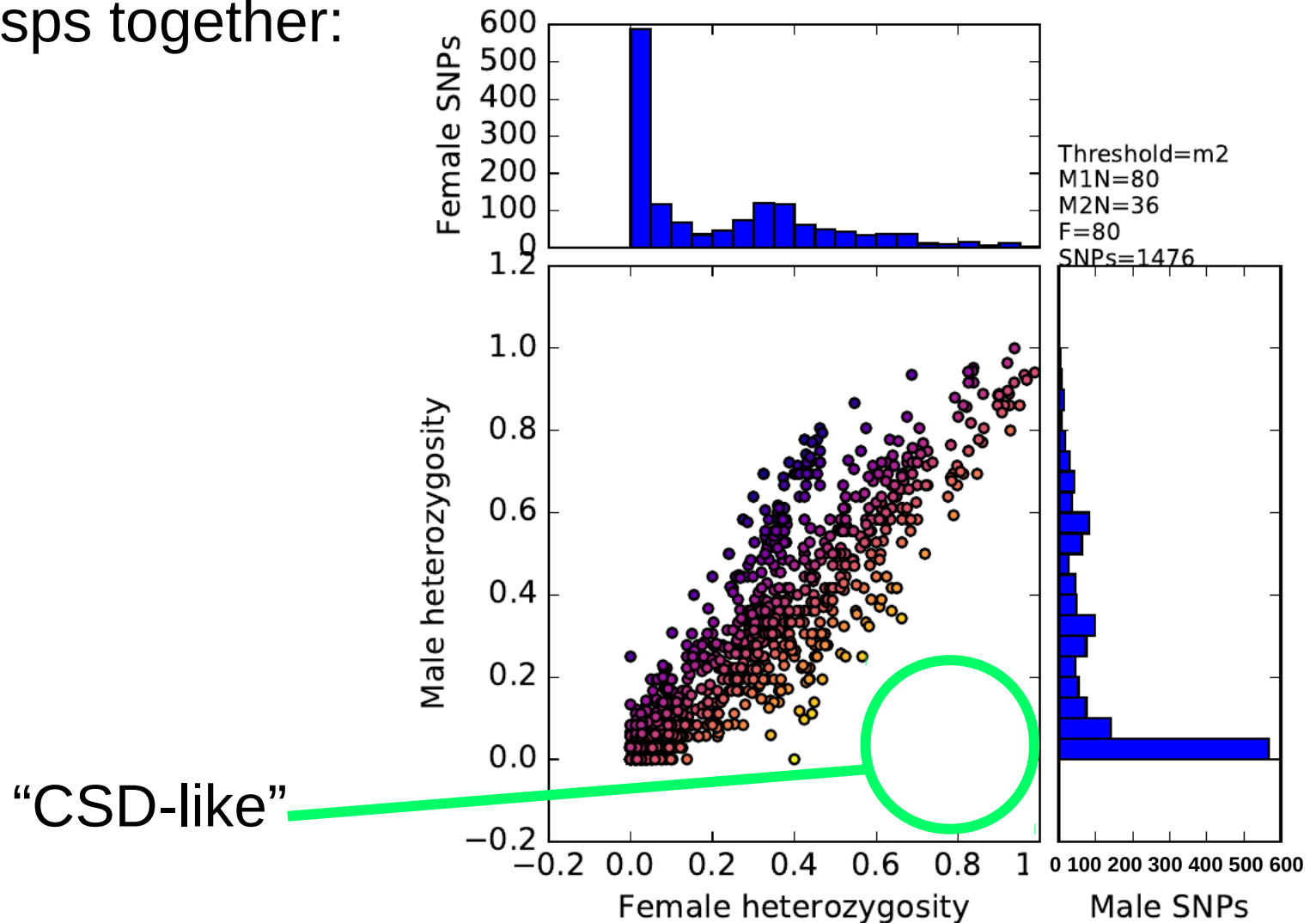
Back to CSD

- How many SNPs are “CSD-like” in diploids ?



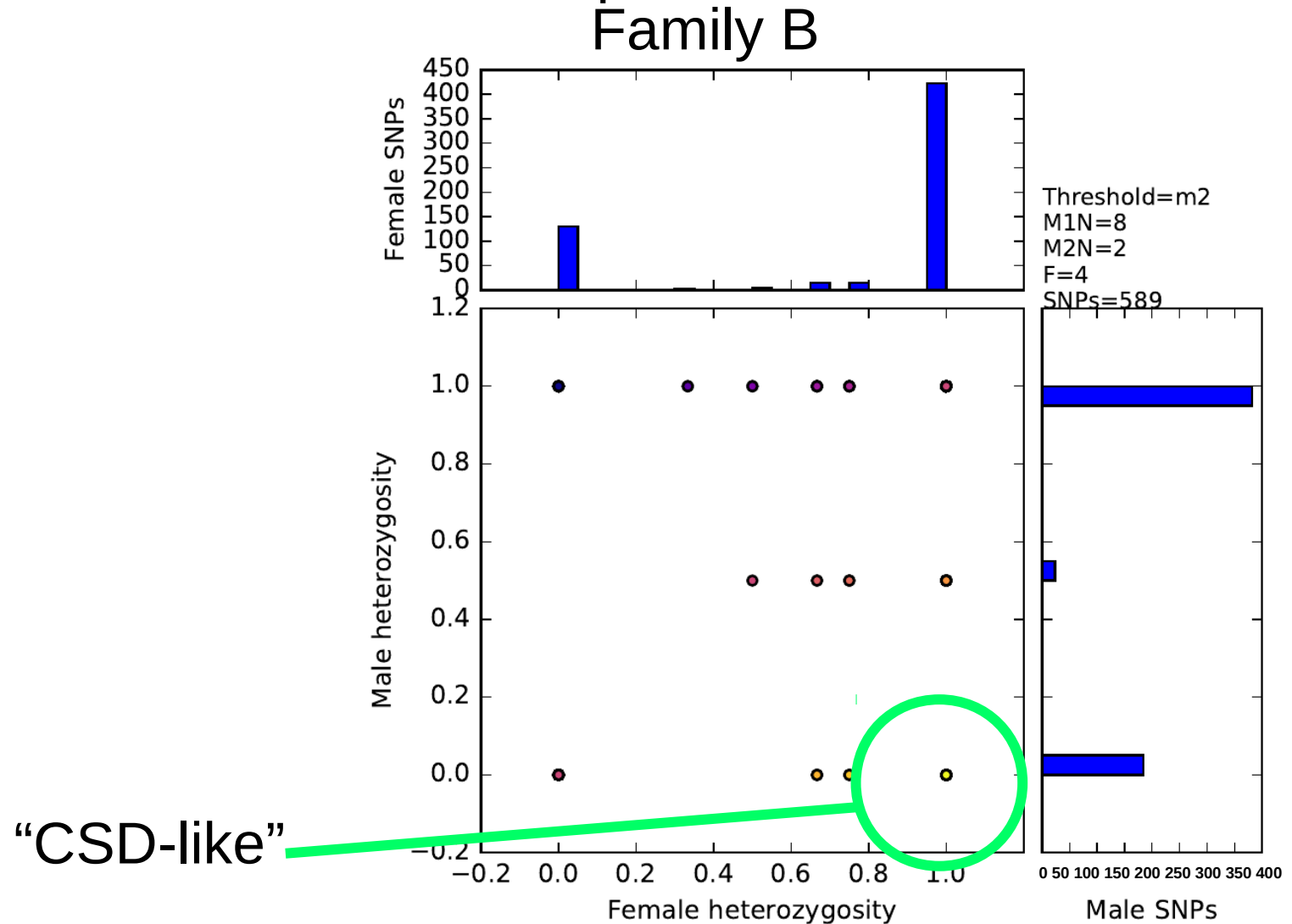
Back to CSD

- How many SNPs are “CSD-like” in diploids ?
- Pooling all wasps together:
 - None



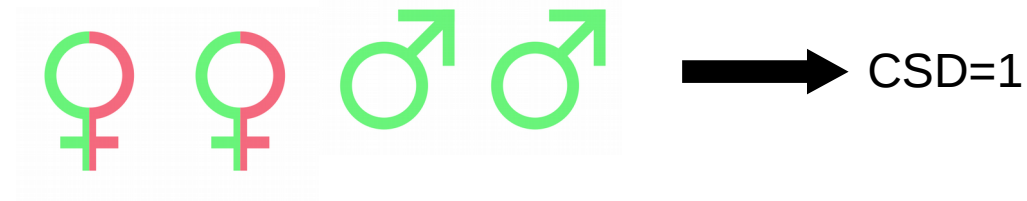
Back to CSD

- How many SNPs are “CSD-like” in diploids ?
- Per family:
 - Too many

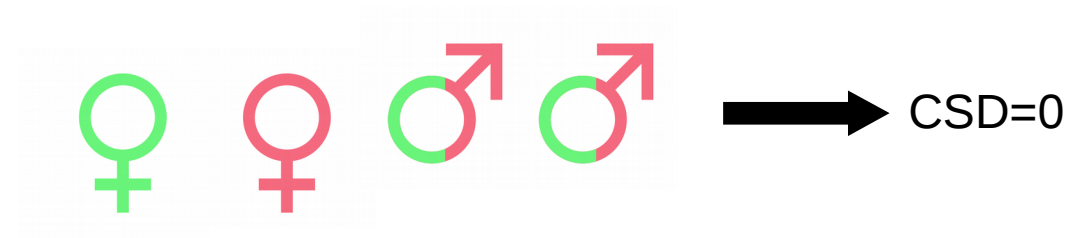


Finding CSD

- What regions are inherited like CSD ?
- For each SNP: proportion of individuals where SNP fit CSD pattern



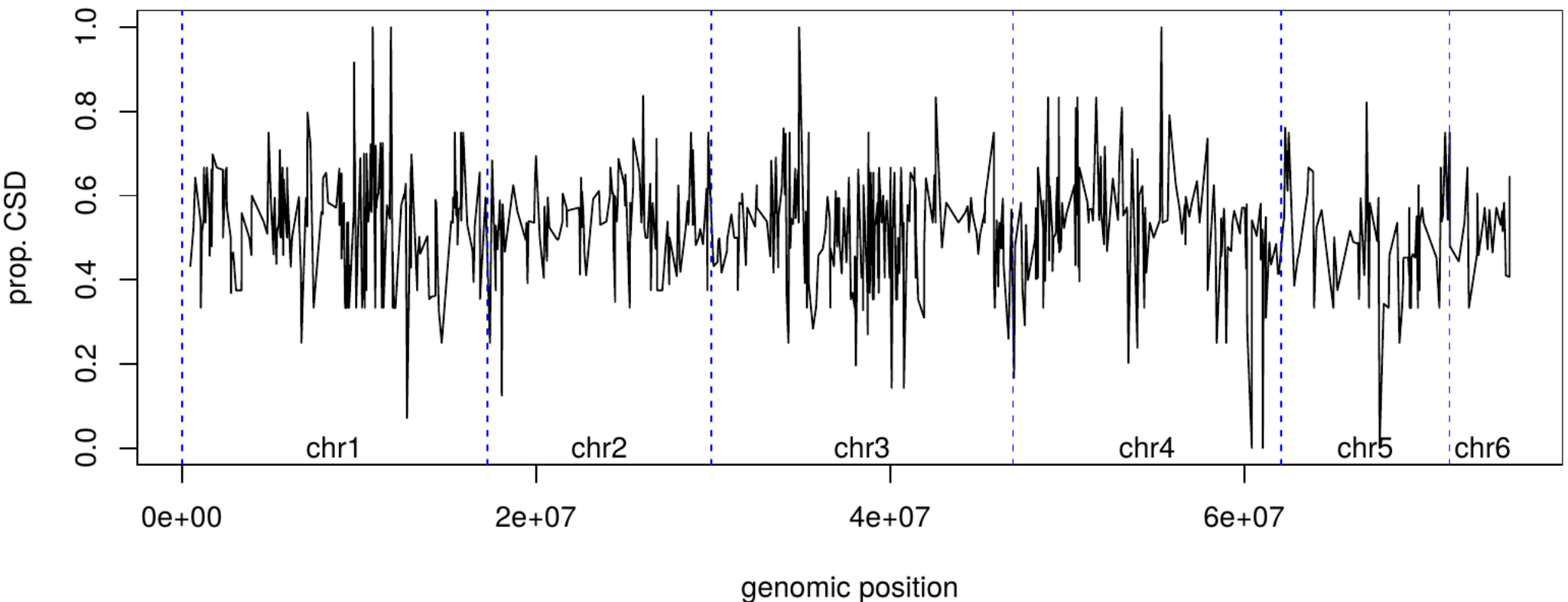
$$\frac{Hom(M) + Het(F)}{2}$$



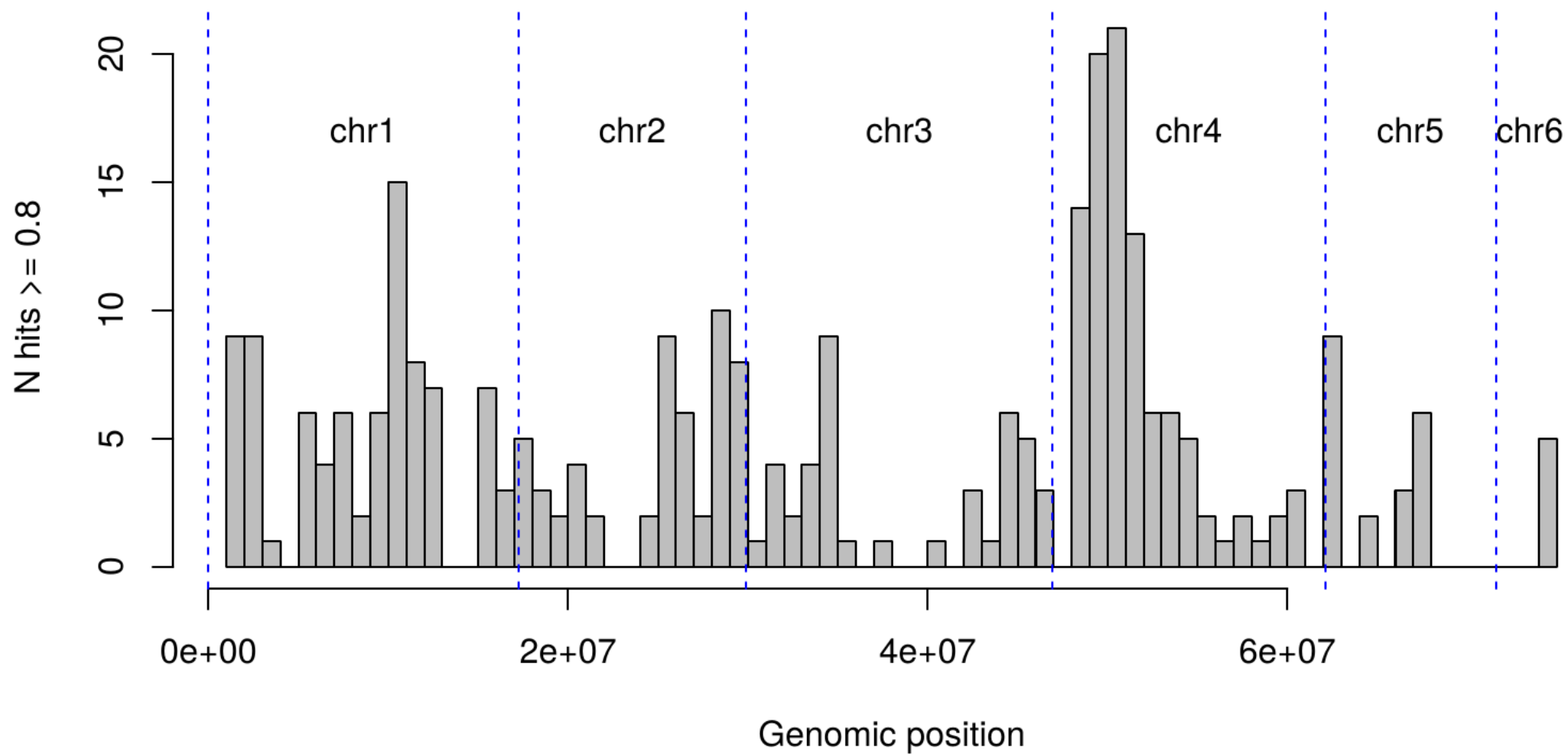
Finding CSD

- Computing CSD prop. in each family
- Average each position's value across families

Proportion of CSD individuals, averaged across families



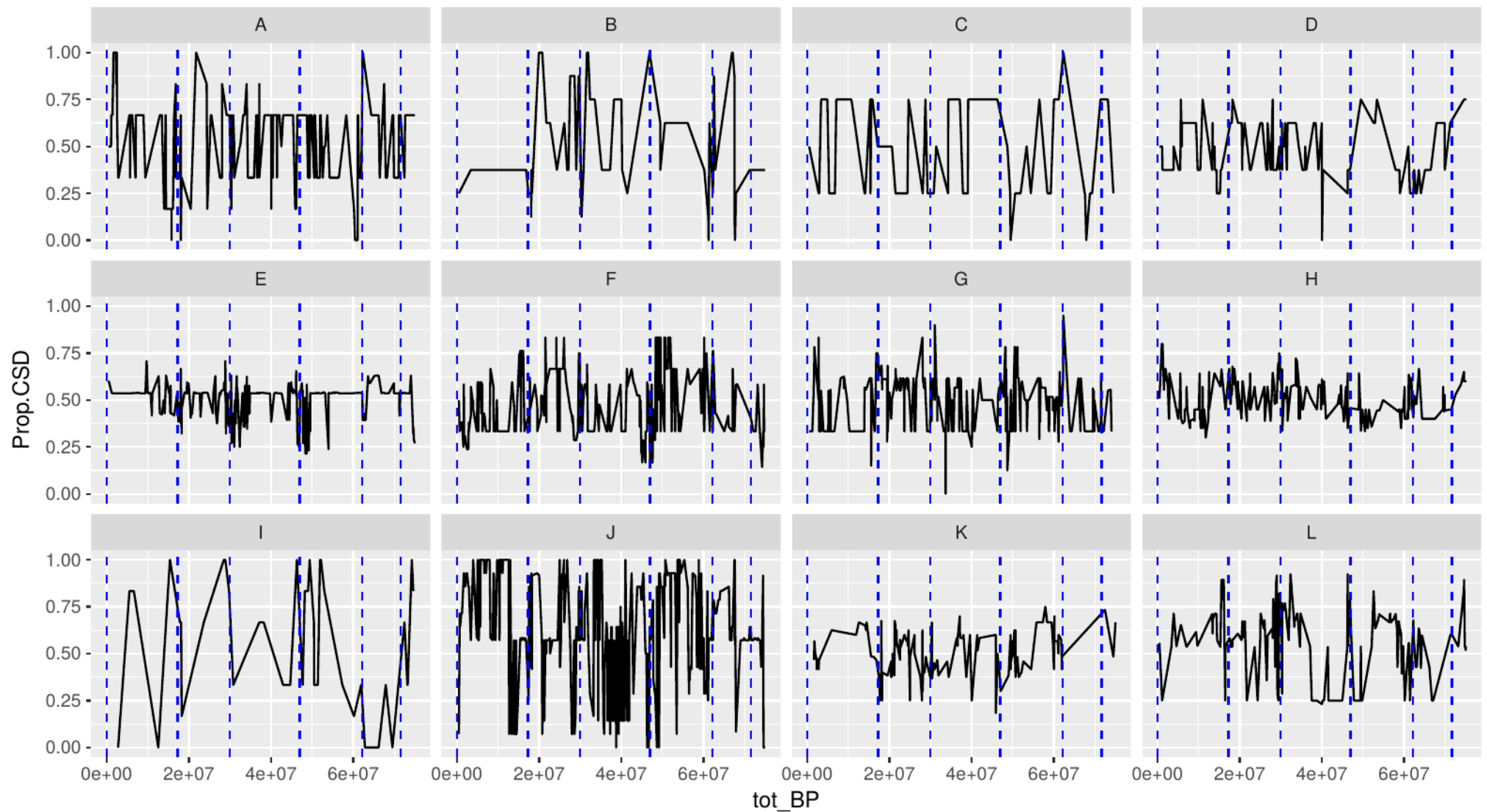
Top CSD candidates



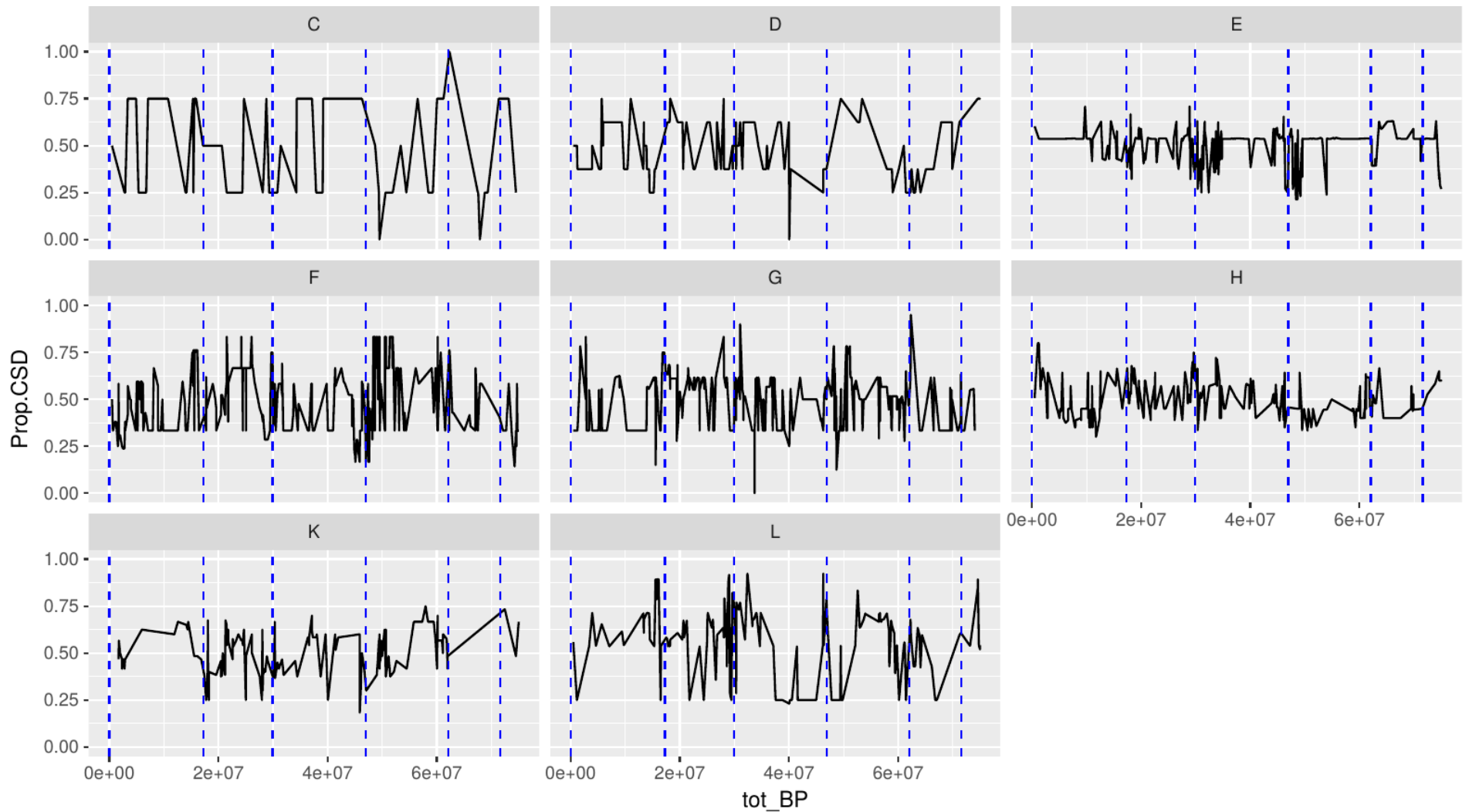
Challenges

- Probably no restriction site inside CSD locus
 - Imperfect male-female segregation
- May be multi-locus CSD
 - Different causative loci in different clutches

Zooming into families

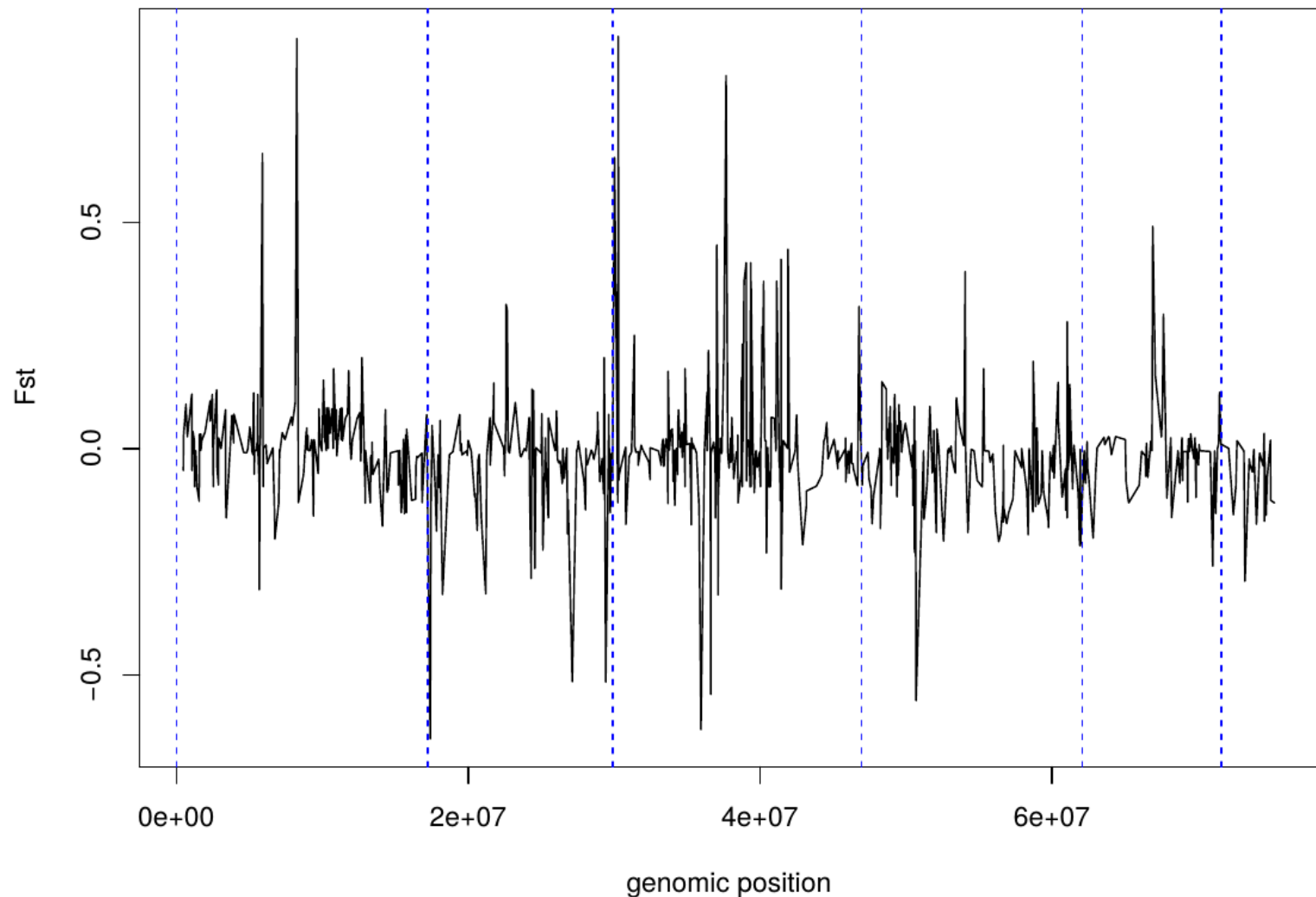


Zooming into families

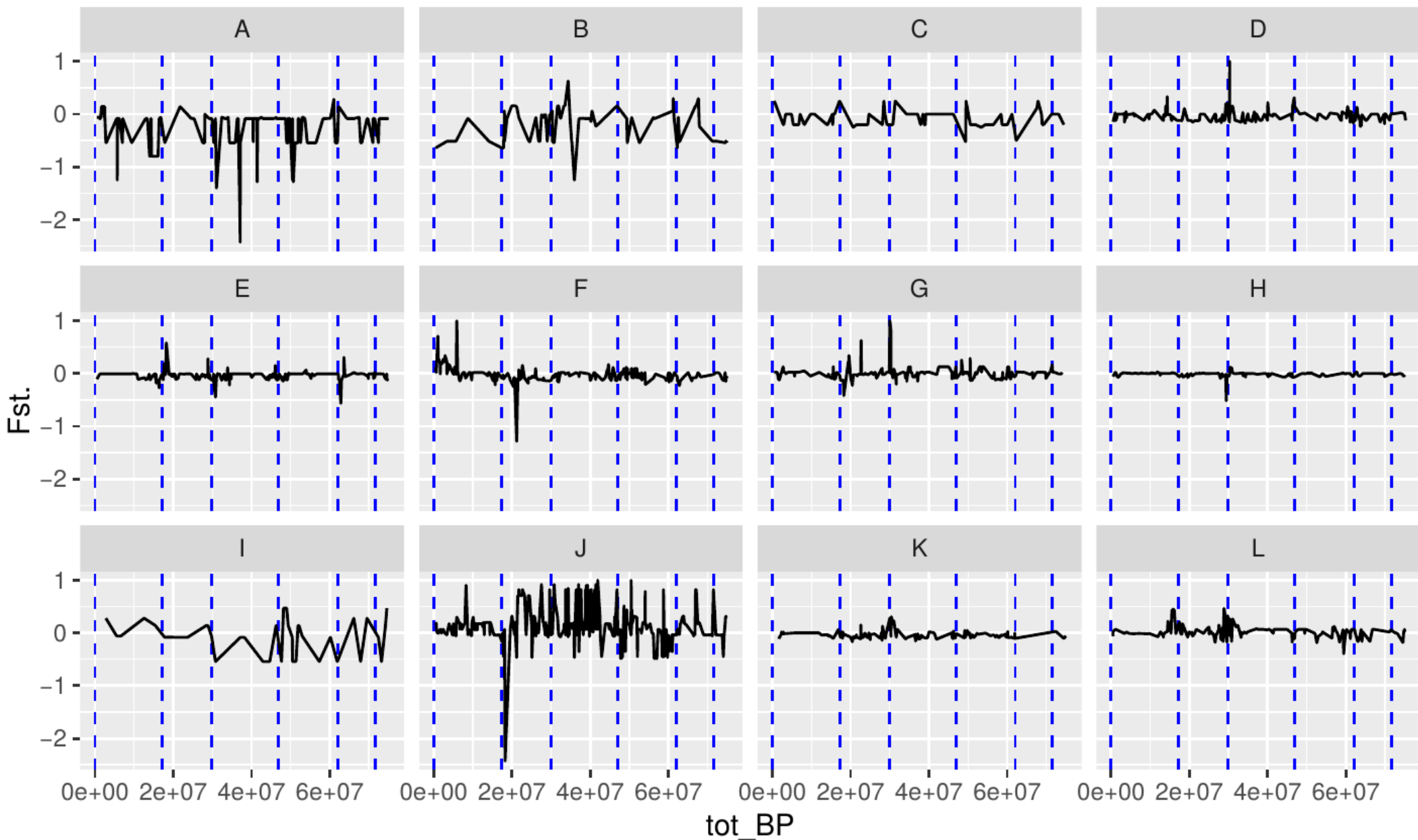


M-F Fst along genome

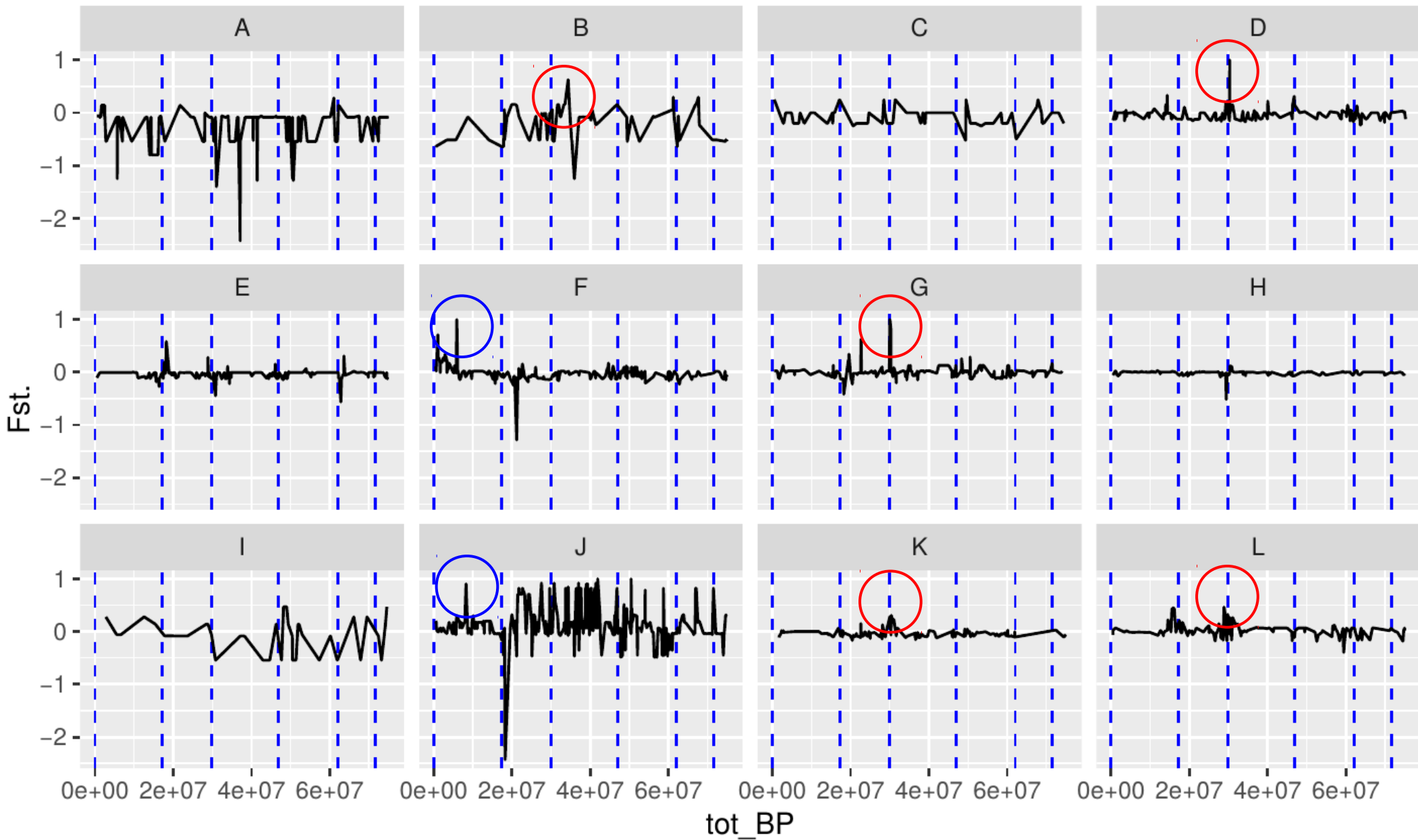
- Pooling all families together



Splitting by family



First results



Family	Mother	Sons	Daughters
A	1	26	4
B	1	10	4
C	1	6	1
D	1	6	4
E	1	3	15
F	1	10	7
G	1	6	9
H	1	15	6
I	1	4	2
J	1	20	7
K	1	14	4
L	1	4	14

213 individuals sequenced

Family	Mother	Sons	Daughters
A	1	1	2
B	1	1	3
C	1	2	1
D	0	4	4
E	1	3	13
F	1	3	6
G	1	3	9
H	1	10	6
I	1	1	2
J	1	1	6
K	1	6	4
L	1	2	13

117 individuals left for downstream analyses.

Back to CSD

- Exclusion of SNPs that are homozygous in mother

