

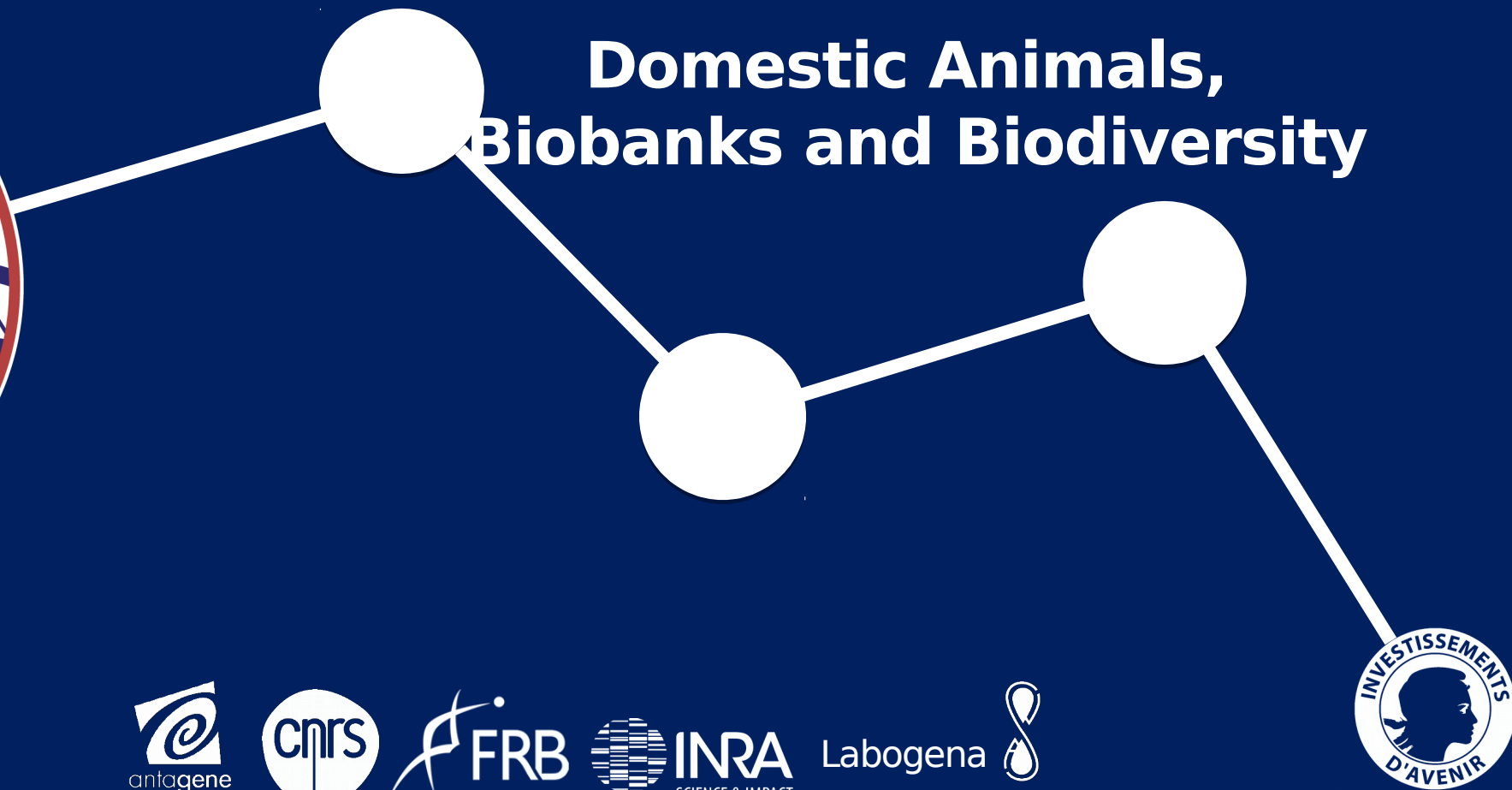
CRB Anim

Centres de Ressources Biologiques

3rd International Seminar of CRB-Anim Infrastructure



Domestic Animals, Biobanks and Biodiversity



Agence Nationale de la Recherche
ANR



MINISTÈRE
DE L'AGRICULTURE
DE L'ALIMENTAIRE
ET DE LA FORÊT



INRA
SCIENCE & IMPACT



VetAgro Sup

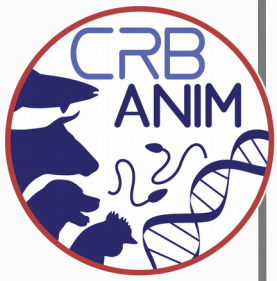


Time trends in genomic variation for commercial pig lines

Simon Boitard^{1,*}, Cyriel Paris¹, Patrice Dehais¹,
Alban Bouquet², Laurence Liaubet¹, Juliette
Riquet¹, Marie-José Mercat²

¹ : GenPhySE, Université de Toulouse, INRA, ENVT, Castanet-Tolosan, ² : Alliance R&D/IFIP, Le Rheu

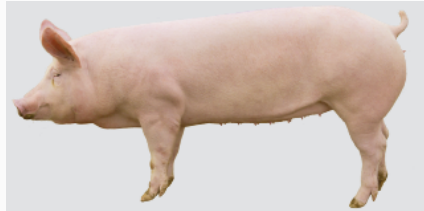
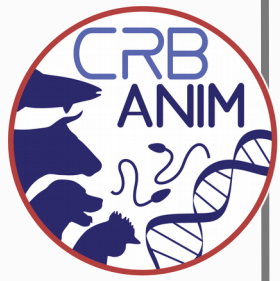
Introduction



- Past positive selection generally detected from present time molecular data.
- Gene bank samples give access to the recent temporal dynamics of allele frequencies.
- Improves detection and annotation of selection events.
- Case study: french sire (LWM) and dam (LWF) Large White lines, comparison of sequences from 1977 and 2012 to 2016.
- New method for time series genomic data, developed within IMAGE H2020 project (PhD Cyriel Paris).

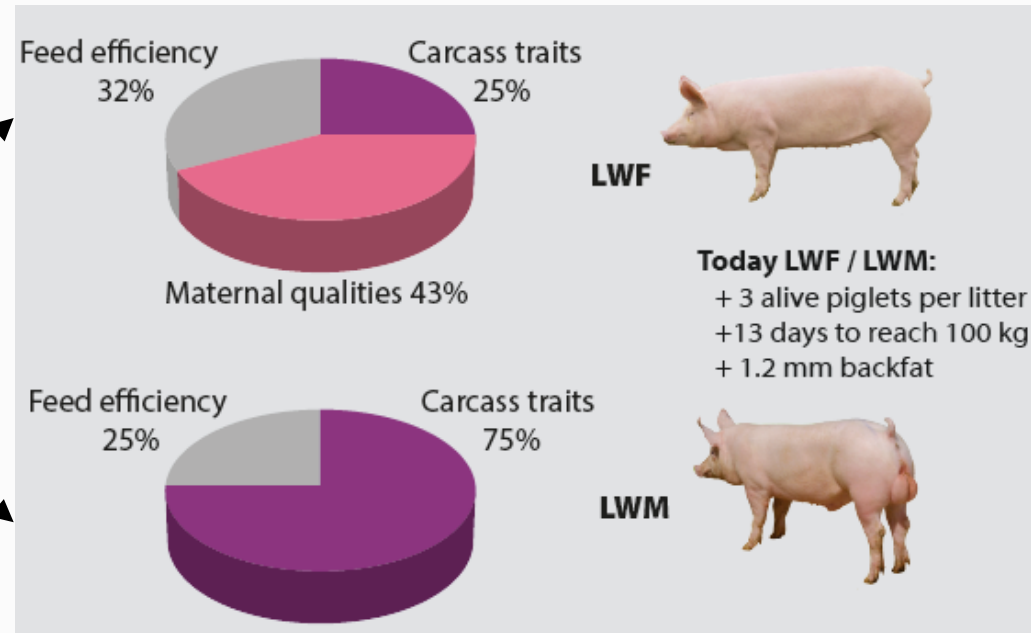


Experimental design



1977: 10 males sequenced at 12X

1995: lines created



2014-2015: 13 males sequenced at 19X

2012 - 2016: 13 males sequenced at 19X



Variant calling



Genome - wide analyzes: 13,408,342

high quality autosomal SNPs

- Found by 3 different callers: samtools, GATK, freeBayes.
- Standard GATK quality filters, bi-allelic.
- Call rate $>90\%$ ($GQ > 20$).

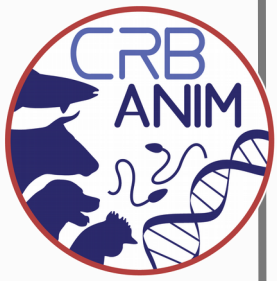
Candidate region analyzes: 28,203,009

autosomal SNPs and indels

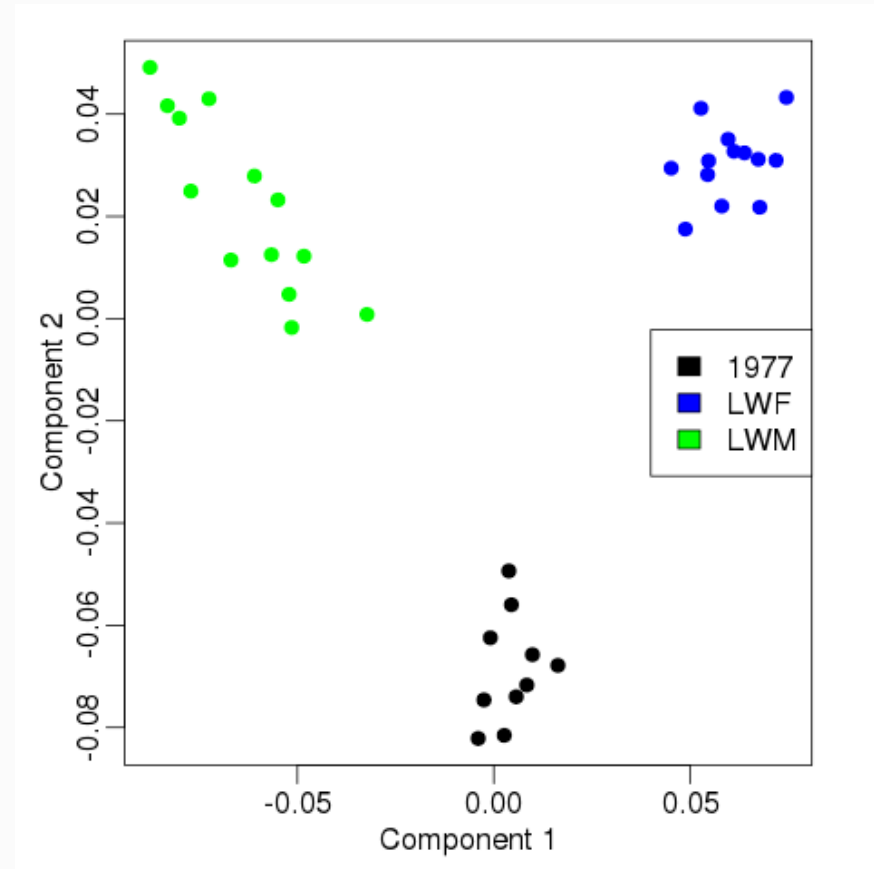
- Found by at least one of the 3 different callers.
- Unfiltered
- No call rate threshold ($GQ > 10$).



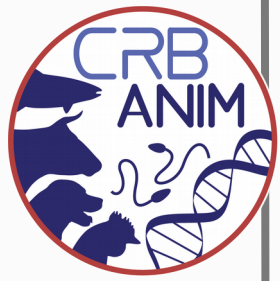
Genetic diversity



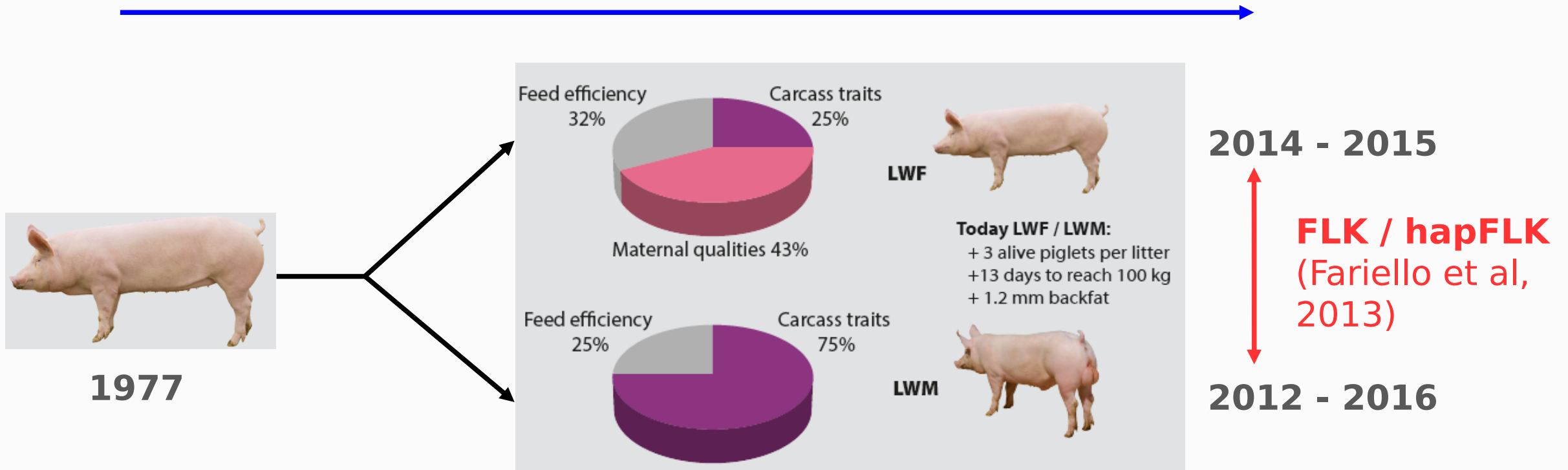
- **Effective size** (1977 – 2014) estimated using R package NB (Hui and Burt, 2015):
80 diploids in LWF, 73 in LWM.
- **Polymorphic sites:**
83 % in 1977 (7 % private),
79 % in LWF (5 % private),
76 % in LWM (4 % private).
- 3 groups **well differentiated**



Methods for detecting selection



df_LWF (Paris et al, 2019)



df_LWM (Paris et al, 2019)



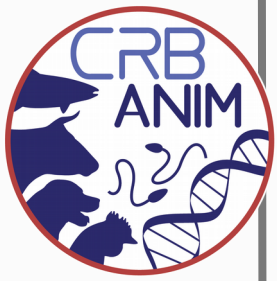
Methods for detecting selection



- For all tests, a **p-value** quantifying the evidence for selection is provided **at each SNP**.
- HapFLK: significant SNPs detected at a False Discovery Rate (FDR) of 15 %.
- FLK: only used to analyse candidate regions in details (little power genome-wide).
- df_LWF/M : significant clusters of low p-values detected with a local score approach (Fariello et al, 2017).



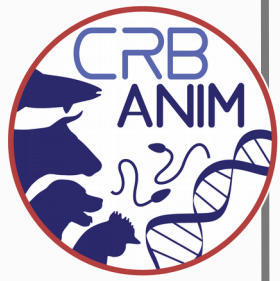
Candidate regions: overview



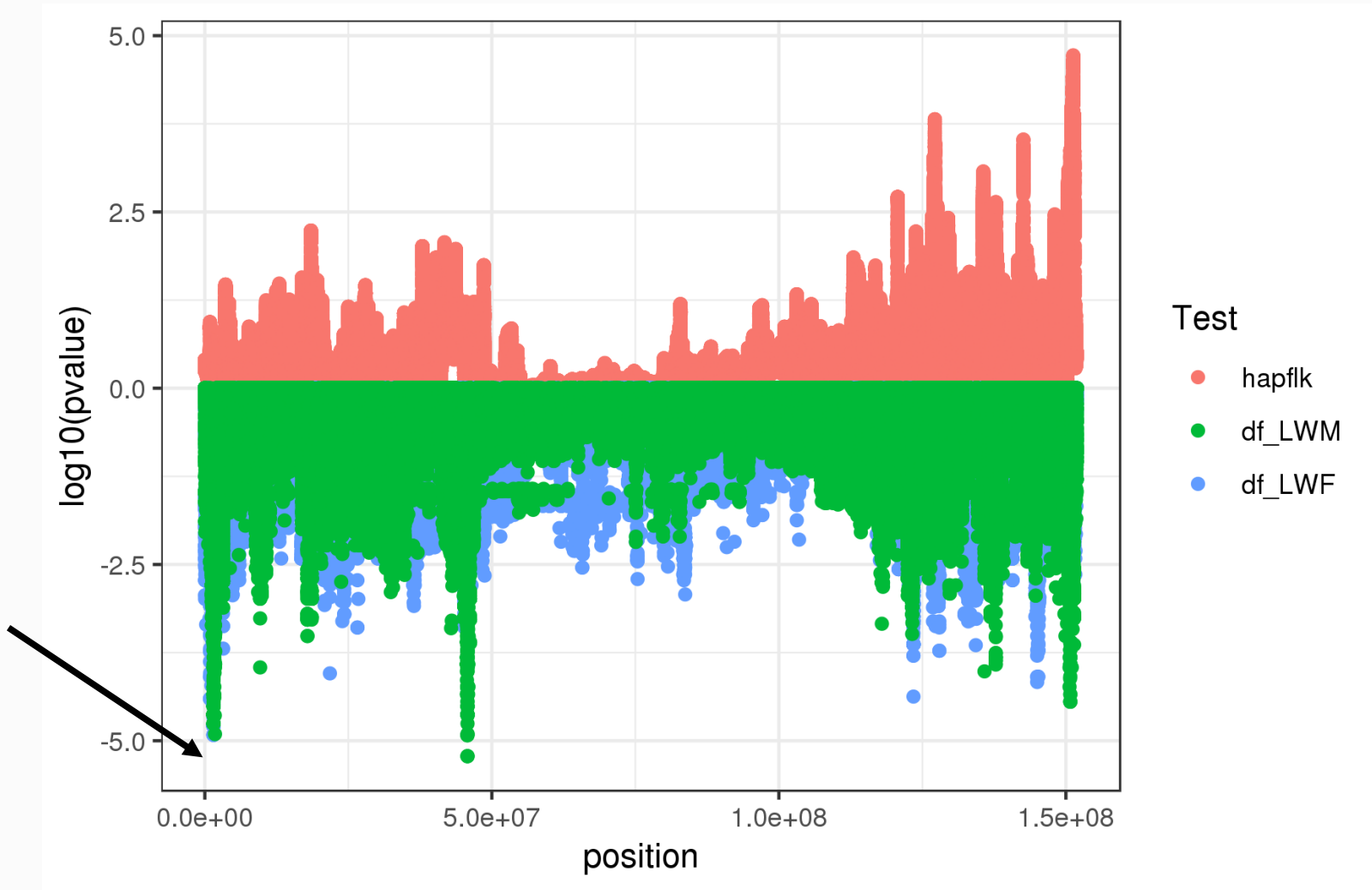
- 30 genomic regions detected by at least one test.
- 7 only in LWF, 4 only in LWM, 17 convergent, 2 divergent.
- Median size 177kb (962bp - 3.2Mb).
- 148 genes in total.



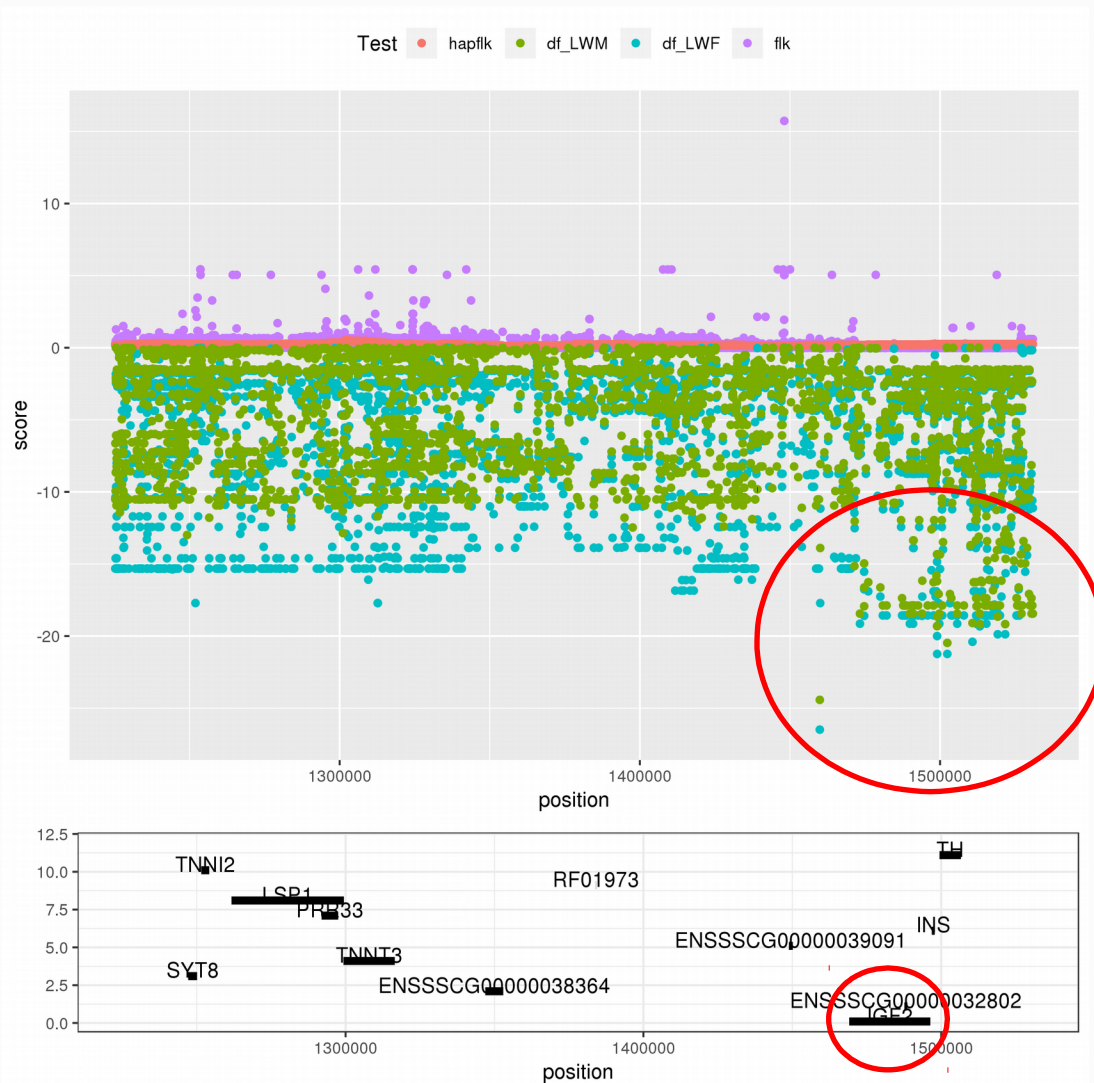
Example 1: IGF2 region (chr 2)



Known mutation affecting muscle growth, fat deposition and heart size (Van Laere et al, 2003)



Zoom on IGF2 region

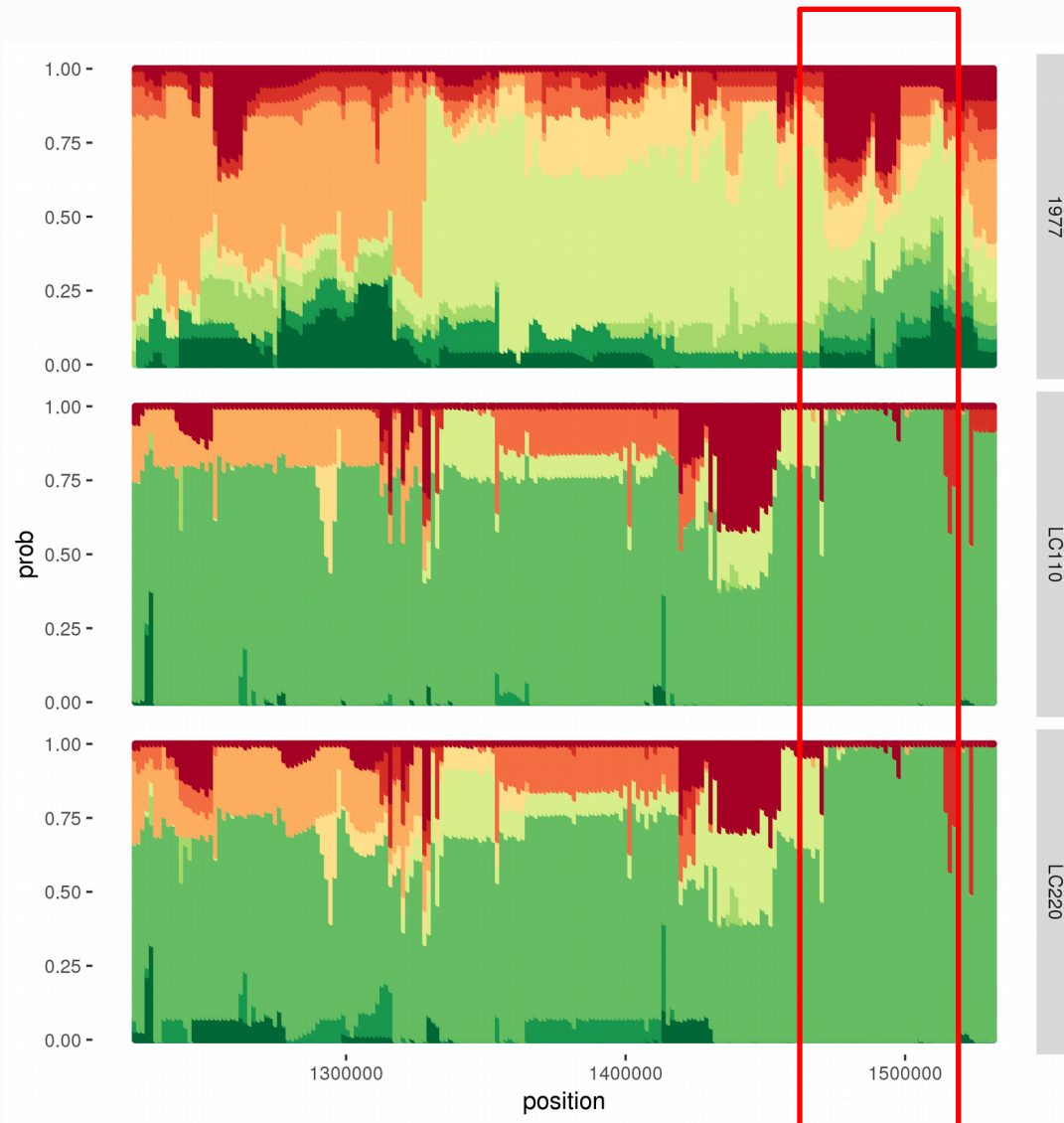
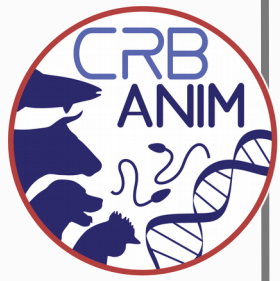


Strong shift of allele frequencies in the two lines, but no signal of differentiation

Highest signal around IGF2



IGF2: Haplotype diversity evolution



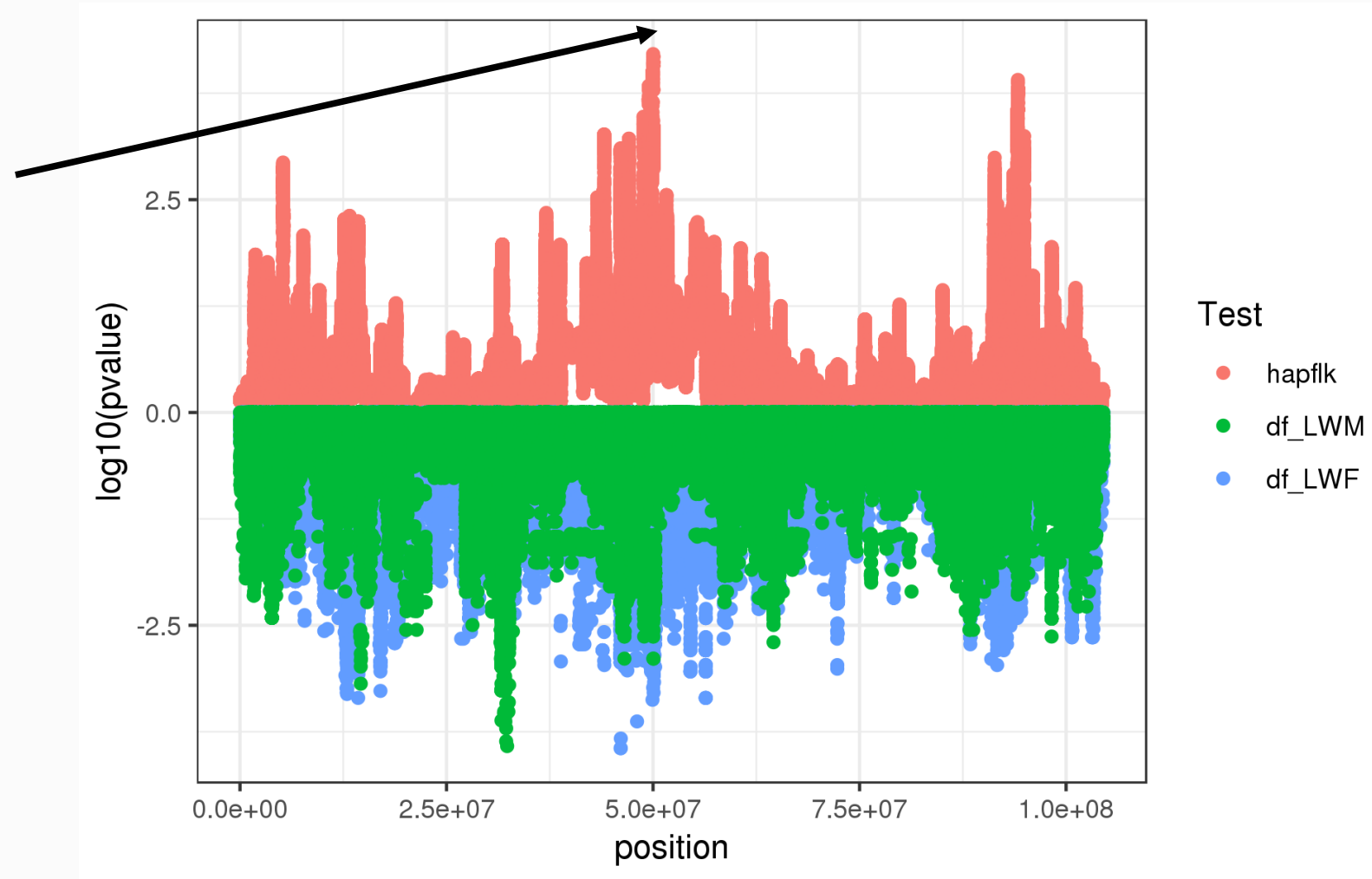
Convergent selection:
same haplotype fixed
in the two lines



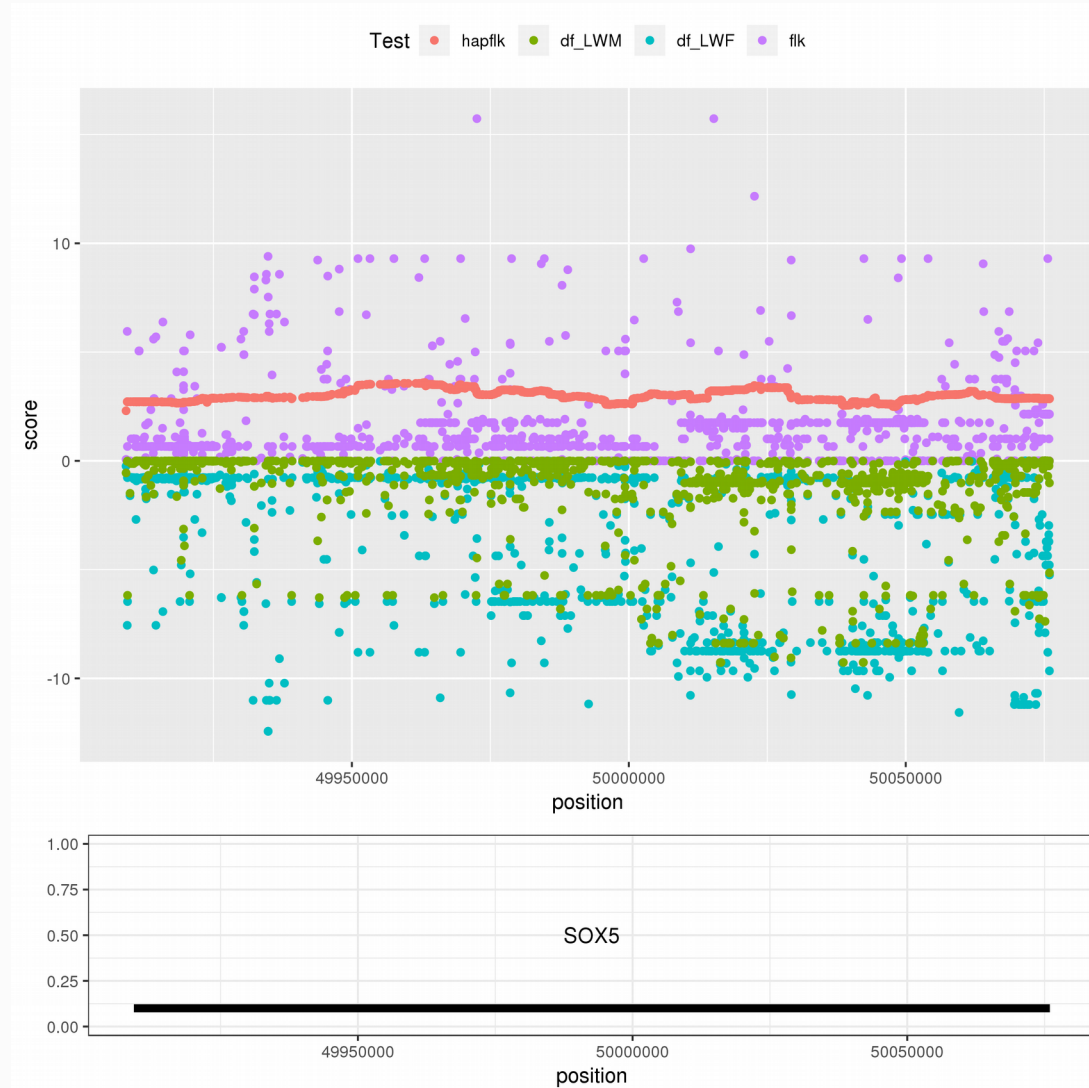
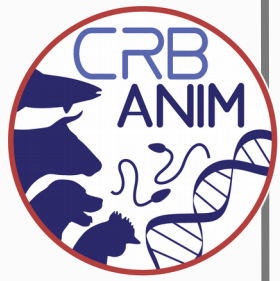
Example 2: SOX5 region



SOX5: embryonic development, chondrogenesis



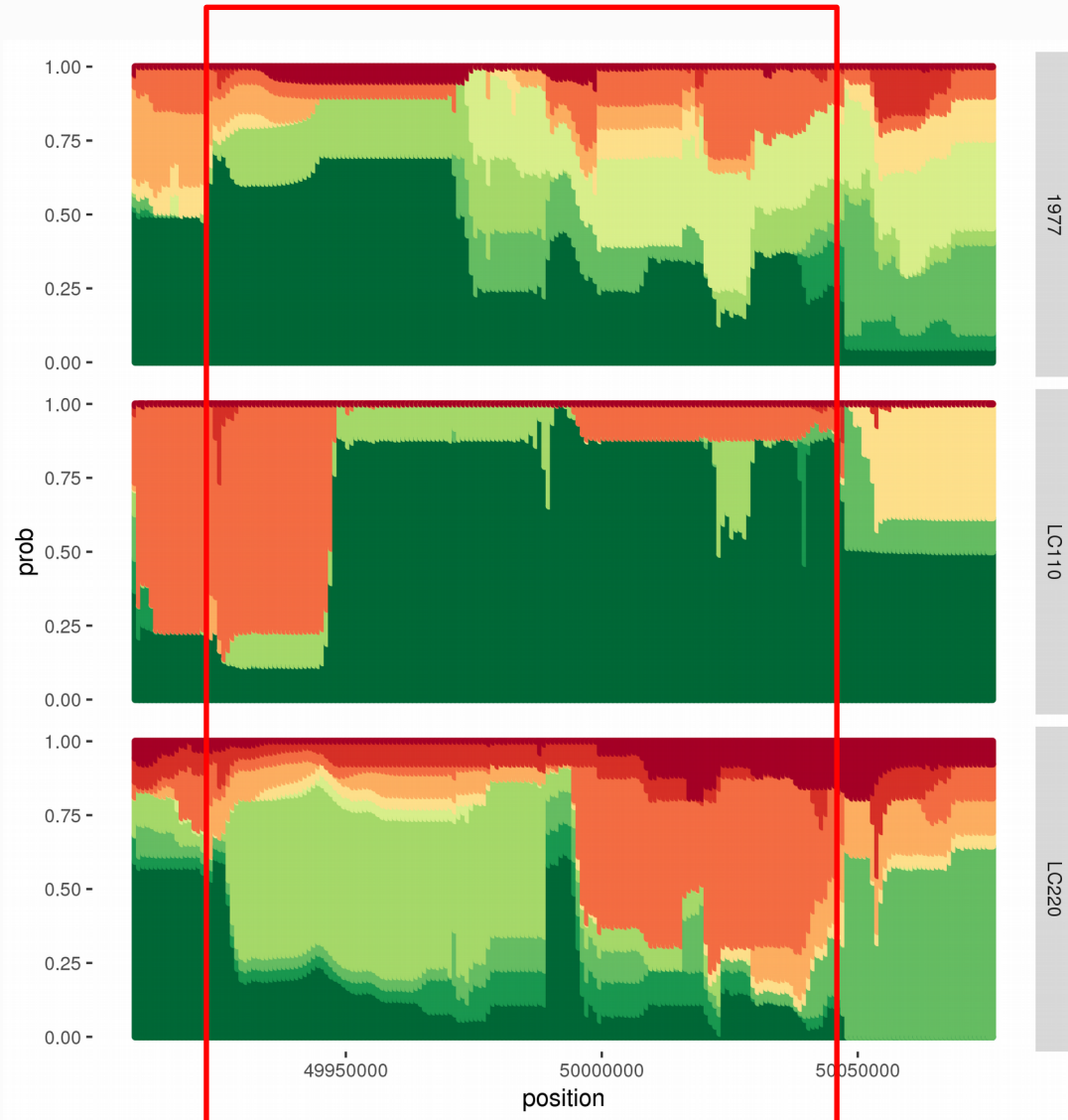
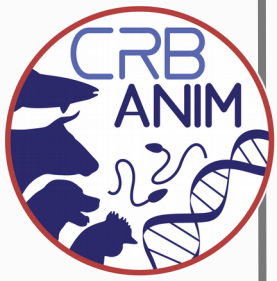
Zoom on SOX5 region



Strong differentiation between lines, several candidate SNPs with high FLK.



SOX5: Haplotype diversity evolution



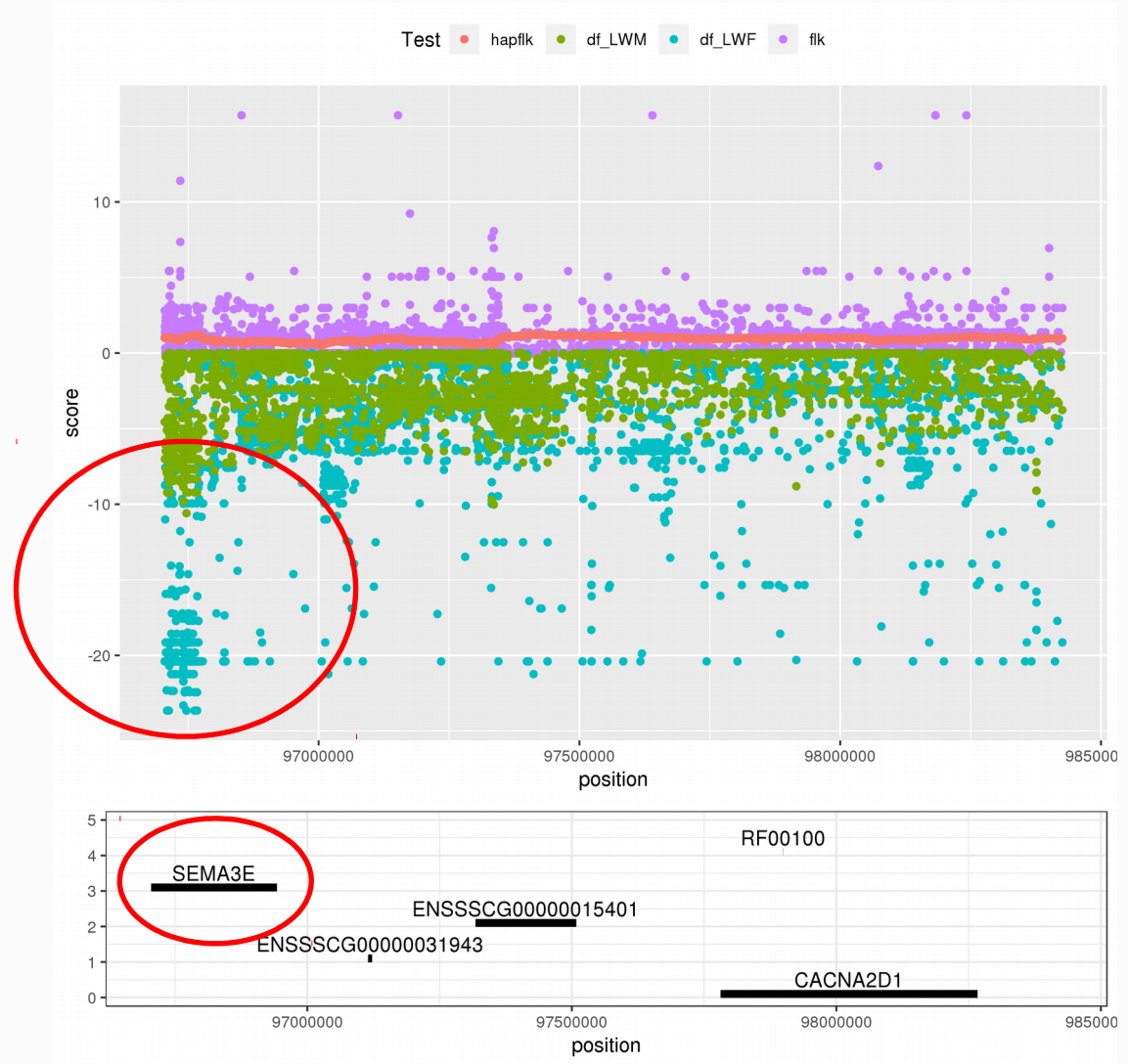
Divergent selection:
distinct haplotypes in
the two lines



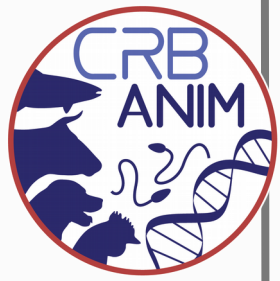
Candidate genes



Several regions with a single gene or a clear local peak on a gene, for instance here SEMA3E (cardiac and vascular development).



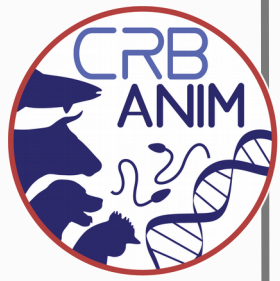
Conclusions & perspectives



- Evidence of population differentiation and slight loss of genetic diversity since 1977 (approx 7 % SNPs).
- 30 candidate regions under positive selection.
- Selection often convergent between lines:
- between 1977 - 1995?
- Ongoing enrichment analysis to highlight functional differences between genes selected in LWF vs LWM.
- Many regions with interesting candidate genes.
- Causal variants can be looked for thanks to whole genome sequencing.



Acknowledgements



- CRB Anim
- IMAGE project, European Union's Horizon 2020 Research and Innovation Programme, grant agreement no. 677353
- Haptitude project, INRA métagrogramme SelGen
- Genotoul bioinformatics platform Toulouse Midi-Pyrénées

