FloSeq
Genotype-Phenotype map and the determinants of Maize floral transition

Fabien Chardon (INRA), Christine Dillmann (Univ Paris-Sud), Marie-Laure Martin-Magniette (AgroParisTech), Adrienne Ressayre (INRA), Maud Tenaillon (CNRS)

UMR GQE - Le Moulon
Ferme du Moulon
91190 Gif sur Yvette
www.moulon.inra.fr

Flowering time is one of the most studied trait in crops because it conditions the fulfillment of the cycle from germination to seed production during the favorable growing season. Our project aims at better understanding the determinants of floral transition in maize, in relationship with earliness and climate adaptation. Our plant material is issued from two long-term divergent selection experiments for flowering time, that started from two inbred lines, F252 and MBS. Each DSE resulted in Late and Early genotypes within a homogeneous genetic background. Our objectives are (i) to produce and analyze **RNA-Seq data** obtained from samples of Shoot Apical Meristem during floral transition; (ii) to develop statistical methods to find **genes differentially expressed** between Late and Early genotypes; (iii) to compare results obtained from **the two genetic backgrounds**, F252 and MBS.

State of the art: Selection for flowering time resulted in changes in the developmental sequence

### Low effective population size leads to fast coalescence

Keeping track of the pedigrees:
2 genetic backgrounds,
4 populations,
2 independent families per population.
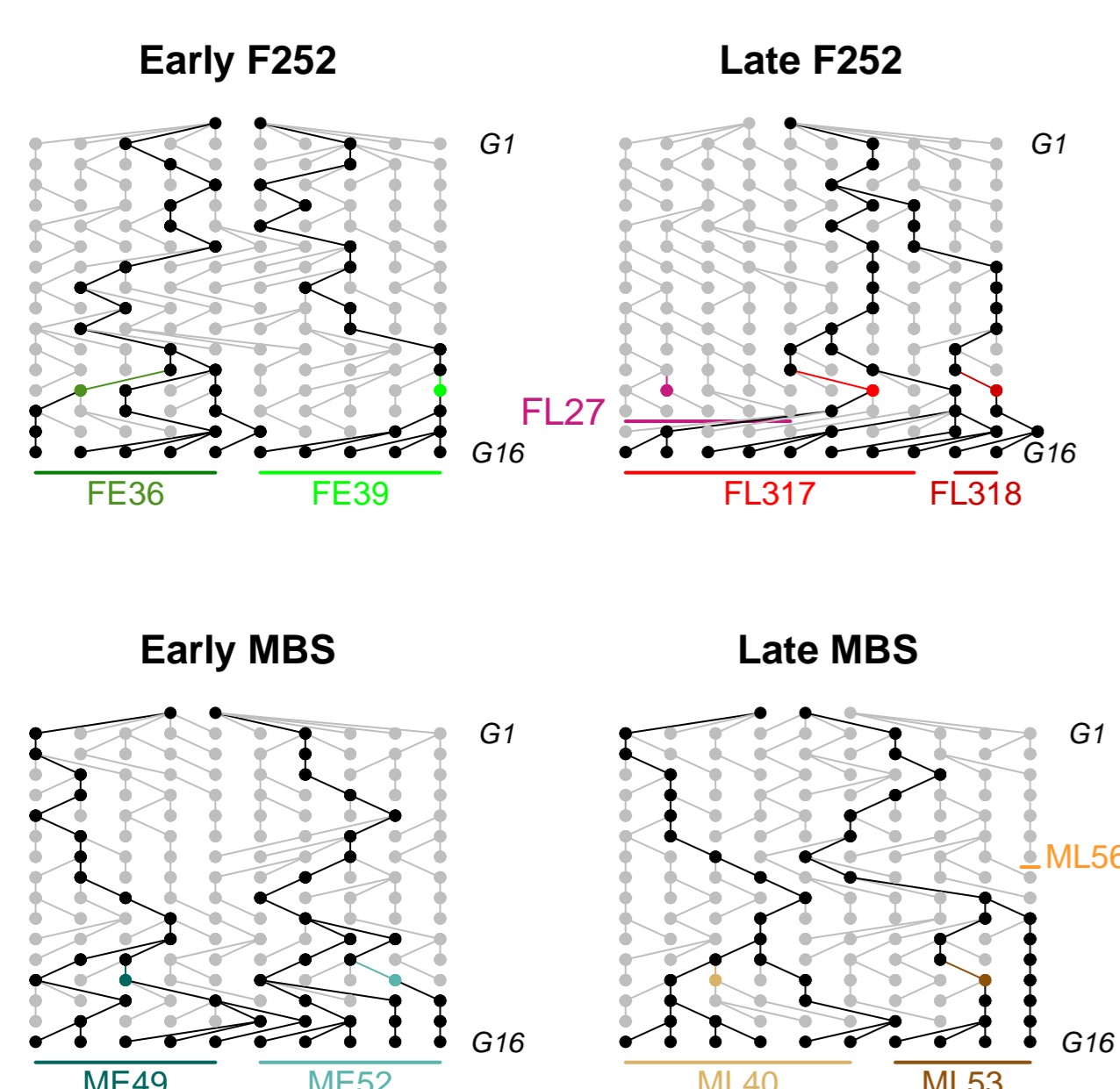
Each DSE (F252 and MBS) was initiated from a single seed lot of a commercial maize inbred. One representative from each family at generation G13 was chosen to fine-study the determinants of the response to selection.

Sustained response to selection indicates genetic changes for the determinants of flowering time

18 generations of Divergent Selection Experiments for flowering time
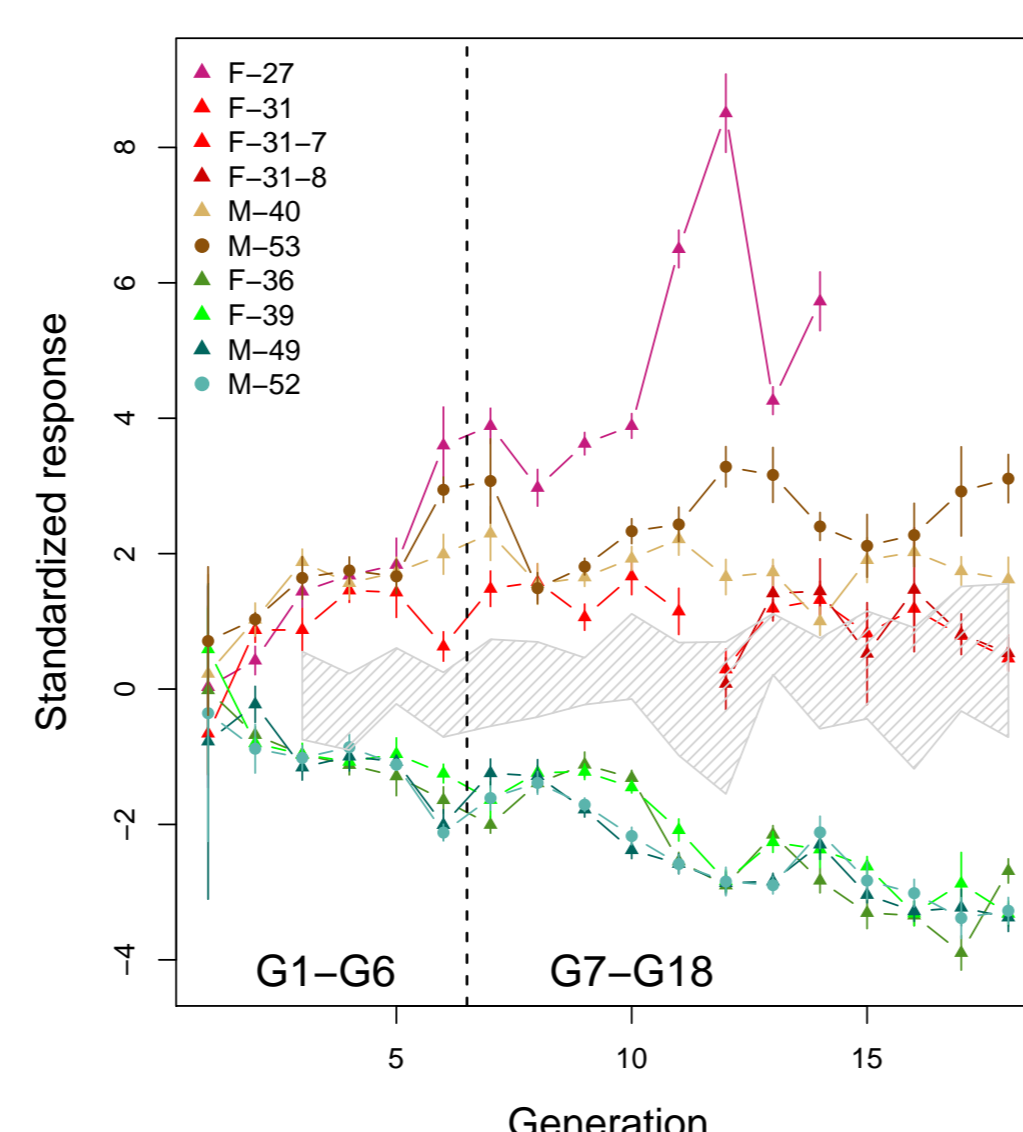
Standardized response is expressed in units of residual deviation for flowering time in the control. Flowering times (expressed in thermal time) were recorded each year on 12 plants per DSE progenitor, and corrected for year and block effects.

Within each Early family, the response to selection is linear. In the Lates families, patterns of response differ between families. Most Late families seem to reach a physiological limit.

⇒ **Seeking for genetic changes**

### Late and Early genotypes differ for both timing and developmental stages for floral transition

Fine analysis of nine representative genotypes

Plants from each genotype were grown in the field in 2012 and 2013. Leaf developmental stage was recorded through time as the number of visible leaves.

At each leaf stage surrounding floral transition, ≈ 30 plants were dissected. Shoot Apical Meristems (SAMs) were pooled according to leaf stage.
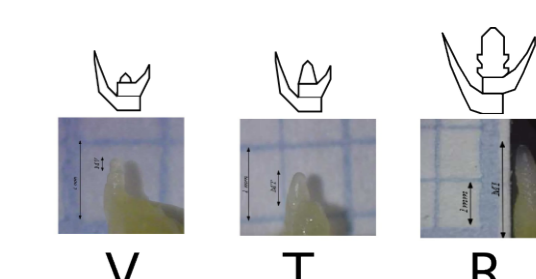
A developmental stage (V=vegetative, T=floral transition, R=reproductive) was assigned *a posteriori* to each pool and corresponded to the average SAM morphological state within the pool.

Within each genetic background (F252 or MBS), Late and Early genotypes differ for the leave developmental stage at which flowering time occurs.

Genotype	Leaf stage						
	L6	L7	L8	L9	L10	L11	L12
FE36		V	T	R			
FL317	V	V	T	R			
FL27			V	T	T	T	
ME52		V	V	T	T		
ML40			V	V	T	T	

Genotype	Leaf stage						
	L6	L7	L8	L9	L10	L11	L12
FE36		V	V	T	R		
FL317		V	V	T	R		
FL27				V	V	T	R
ME52			V	V	T	R	
ML40				V	V	V	T	R

SAM morphological state of SAM pools as a function of leave developmental stage in the 2012 (left) and 2013 (right) samples.
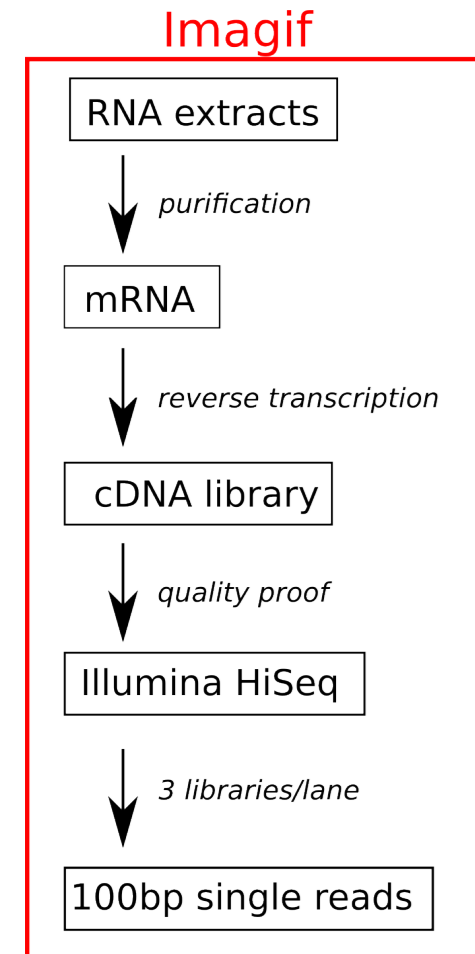
RNA were extracted from each sample and conserved for RNASeq.

* Preliminary results: 2012 samples from $FE36$ and $FL317$. Samples at the same SAM state were considered as replicates.
* FloSeq project: RNASeq for the full data set. Including 2 genetic backgrounds (F252 and MBS) and biological replicates.
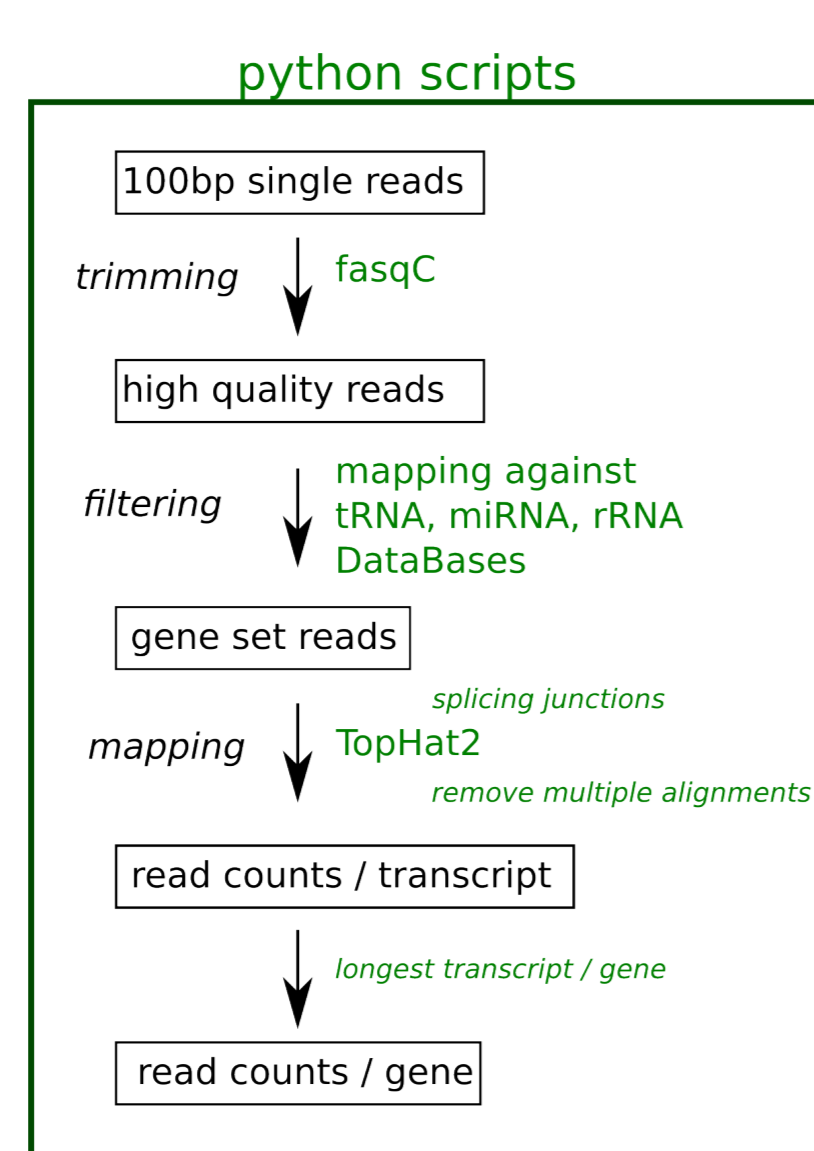
Data treatment

RNASeq

Imagif
RNA extracts → *purification* → mRNA → *reverse transcription* → cDNA library → *quality proof* → Illumina HiSeq → *3 libraries/lane* → 100bp single reads

Repeatability across years or experimentation?

Bioinformatics

python scripts
100bp single reads → *trimming* fasqC → high quality reads → *filtering* mapping against tRNA, miRNA, rRNA DataBases → gene set reads → *mapping* splicing junctions TopHat2 remove multiple alignments → read counts / transcript → longest transcript / gene → read counts / gene

Multiple alignments and gene families? Alternative splicing?

Statistical analyses

R scripts
read counts / gene → *normalization* DESEQ2 → normalized reads → *transformation* log → Intensity / gene → *dispersion analysis* VOOM → weight / gene / sample → *linear model* LIMMA $y \sim genotype + genotype:stage$ → list of genes differentially expressed → *clustering* predicted intensities Principal Component Analysis cluster = PCA axis → developmental pattern analysis

4 gene clusters

Organ Identity — Flowering Transition — Early *vs* Late — Genotype by stage interactions
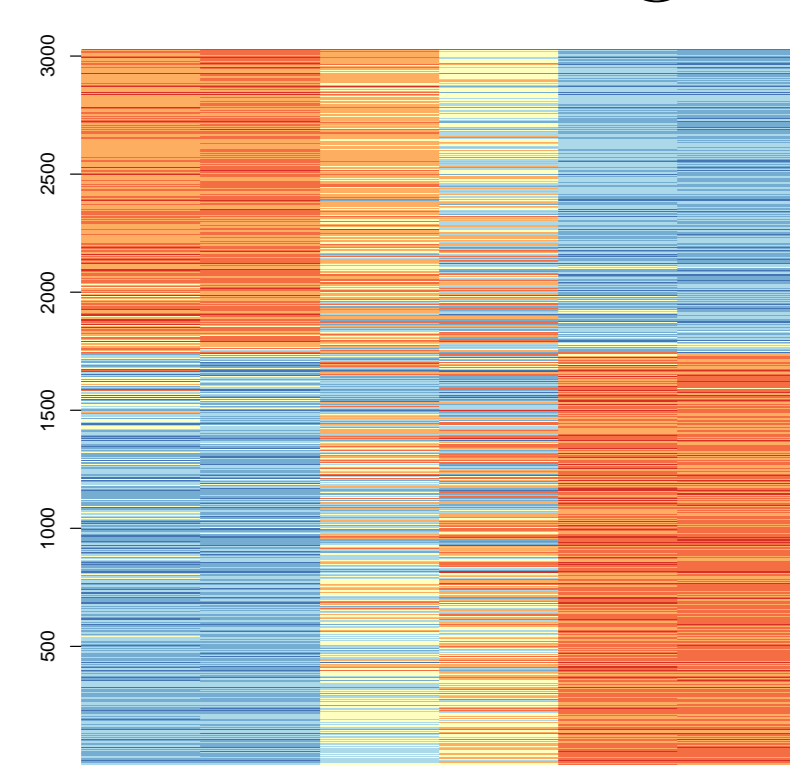
⇒ **Refining statistical analyses?**

The project will contribute to (1) methodological improvements for statistical analysis of RNASeq data; (2) Find new genes involved in maize floral transition; (3) identify adaptive alleles that could be used in breeding programs.
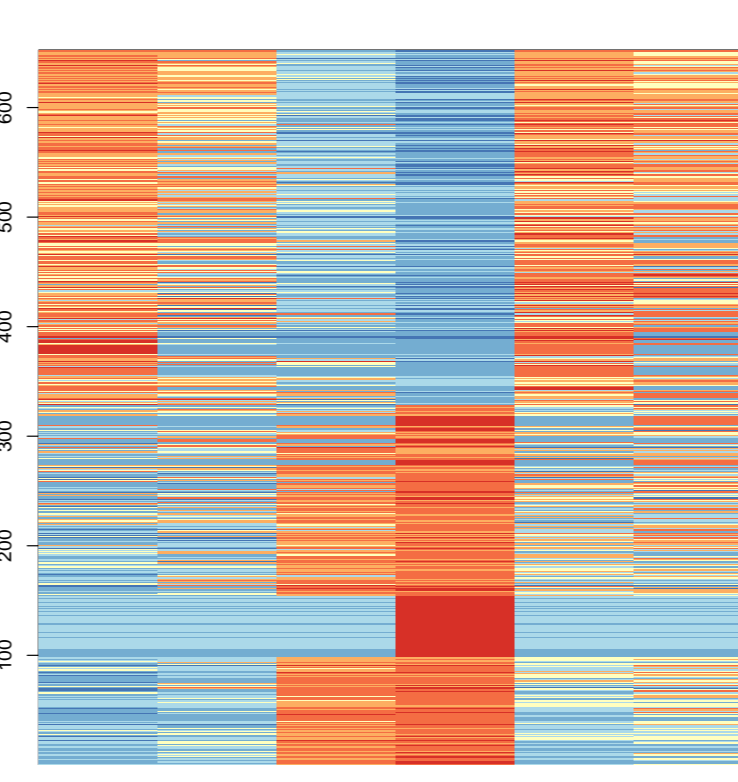
Preliminary results

Genes involved in organ identity

Genes involved in flowering transition

Genes involved in Ealy / Late differences

Heatmap of predicted gene expression for each genotype at each SAM stage. Values are centered around the mean. Blue colors indicate under-expression, Red colors over-expression. Genes grouped by cluster.